Functional characterization of a bacteriophage-encoded inhibitor of Staphylococcus aureus transcription

Citation

Share Your Story

Functional characterization of a bacteriophage-encoded inhibitor of *Staphylococcus*

***aureus* transcription**

A dissertation presented

by

Cristina Montero Diez

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Microbiology and Molecular Genetics

Harvard University

Cambridge, Massachusetts

June 2013

Functional characterization of a bacteriophage-encoded inhibitor of *Staphylococcus aureus* transcription

Abstract

Transcription initiation—and, specifically, promoter recognition—is a key regulatory step in bacterial gene expression. Promoter recognition is mediated primarily by the σ subunit of RNA polymerase (RNAP), which makes sequence-specific contacts with the -10 and -35 promoter elements in the context of the RNAP holoenzyme. At a subset of promoters with A+T-rich upstream sequences (UP elements), the C-terminal domains of the RNAP α subunit (αCTDs) also contribute to promoter recognition by making sequence-specific contacts with the UP element. Many transcription regulators encoded by bacteria and bacteriophages exert their effects through direct interactions with α or σ. A recently described phage-encoded inhibitor of RNA synthesis in *Staphylococcus aureus* (*Sau*), phage G1 gp67, provides a novel example, binding to *Sau* σ^A in the context of the *Sau* RNAP holoenzyme and exerting promoter-specific effects by inhibiting αCTD function.

Despite the high degree of conservation amongst primary σ factors, gp67 does not bind to *E. coli* (*Eco*) σ^{70}. We use genetic analyses to dissect the *Sau* σ$^A_4$/gp67 interface and identify four amino acid differences between *Sau* σ^A and *Eco* σ^{70} that determine the specificity of gp67 binding. Using this knowledge, we construct an *Eco* σ^{70} variant capable of binding gp67 and show that gp67 associates with *Eco* RNAP holoenzyme reconstituted with the modified σ^{70}, thus enabling us to use *E. coli*-based tools to examine gp67 function. By testing the effect of gp67 on transcription *in vitro* from promoters bearing or lacking a well-characterized UP element, we

uncover a previously unrecognized stimulatory role for sequence-nonspecific contacts of the αCTDs with upstream DNA at a promoter lacking an UP element. We also find that gp67 exerts inhibitory effects on promoter function that, depending on the context, can be αCTD-dependent (as in *Sau*) or αCTD-independent (likely the result of compromising -35 element recognition by σ^{70}). We explore the possibility that the context dependence of gp67 action is the result of fine-structure differences between the *Eco* and *Sau* transcription complexes and propose a *B. subtilis*-based system as an experimentally tractable alternative for *in vivo* and *in vitro* mechanistic studies of gp67 function.

Table of Contents

Acknowledgements

It's hard to believe six years have passed, and yet I find myself here, at the end of the journey. The road's been filled with ups and downs and twists and turns, and I couldn't have navigated it alone… I have no sense of direction and can get lost even *with* a map! I want to thank my advisor, Ann, for being my foolproof GPS unit and unerringly pointing me in the right direction every time I had absolutely no idea where I was or where I was going. Ann has that elusive perfect combination of keen scientific insight & conscientiousness and ability to encourage, inspire, and connect with all members of her laboratory that all students would be fortunate to find in their advisors. Without her support, encouragement, and direction (and her ability to take it in stride when I set off for remote corners of the world in fits of restlessness), I'd likely be mired in quicksand and waving a white flag in surrender instead of here. I consider myself incredibly fortunate to have been able to learn from someone who exemplifies what a scientific investigator and a mentor should be, and I can only aspire to be half the mentor she is.

I am also grateful to all members of the Hochschild Lab, past and present, for making it such an enjoyable place to be. Graduate school is an exercise in fortitude, and actually wanting to show up every morning makes all the difference. Thanks for the laughs, the drama, the scientific input, the coffee/tea breaks, and all the fun in the hot lab. I must especially thank Pádraig Deighan, baymate extraordinaire, who not only sat next to me for six years and answered all my inane questions but also was instrumental in getting a key assay in my dissertation working. He is incredibly knowledgeable and always willing to help and share his expertise, and the friendliest, most approachable person I know. I want to be like him when I grow up.

I wouldn't be here without my friends and classmates, who made sure I didn't spend every waking moment in lab and actually saw the sun and other people every once in a while. To those who embarked on this crazy journey with me because grad school seemed like a better idea than a 'real' job: thanks for the Rock Band marathons, the StarCraft LAN parties, the countless brunches/dinners and movie nights, the willing ears, and—most especially—the remote-controlled flying shark. To my fellow Kravists and scuba divers: thanks for the bruises & injuries and the middle-of-the-night drives to Rockport for crack-of-dawn dives before facing a full day in lab… because learning to kick ass and blowing bubbles (even if I need three layers and heat packs to last more than 15 minutes underwater) is much cheaper than therapy.

Finally, I want to thank my family. Sin vosotros no hubiera llegado hasta aquí—ya sea por lo que me habéis aguantado, los sacrificios que habéis hecho para que tenga una buenísima educación, las oportunidades y experiencias que me habéis brindado con los viajes y las mudanzas, el apoyo incondicional en mis momentos de crisis, o vuestra buena voluntad de dejar lo vuestro de lado cuando necesito ayuda (aunque no sepa que la necesite). Se que soy un poco cascarrabias y poco comunicativa, pero agradezco todo lo que habéis hecho por mí, y os quiero a todos.

List of Figures

List of Tables

Chapter 1:

Bacterial Transcription

Overview

Multi-subunit DNA-dependent RNA polymerases (RNAPs) are the principal enzymes of gene expression across all domains of life and are structurally conserved from bacteria to humans [reviewed in 33], implying a common transcription mechanism in all organisms of which many features are still poorly understood. The relative simplicity of the bacterial transcription machinery, coupled with the availability of a wealth of genetic and biochemical tools, makes it an excellent model system for understanding multisubunit RNAPs and gene expression. Since gene expression is regulated principally at the level of transcription, studies using the experimentally tractable model bacterium *Escherichia coli* (*E. coli*, or *Eco*) can provide global mechanistic insights into RNAP function and regulation.

Bacterial RNA polymerase

Bacterial RNAP is comprised of a single, 380 kilodalton (kDa) core species (referred to as 'core' and abbreviated 'E') composed of five subunits: two copies of α, and one each of β, β', and ω [reviewed in 104]. The high-resolution X-ray crystal structure of core RNAP from *Thermus aquaticus* (*T. aquaticus*, or *Taq*) reveals that the five subunits are arranged in a scaffold that resembles a crab claw; the two largest subunits (β and β') form a cleft that accommodates double-stranded DNA, with the enzyme's catalytic site at the base of the cleft [149]. Eukaryotic RNA Polymerase II—the RNAP responsible for transcription of eukaryotic messenger RNA—has a remarkably similar structural scaffold [25, 33]: although larger (500 kDa versus 380 kDa) and with more subunits (10-12 versus 5), each of the five bacterial RNAP core subunits has a structural homologue in RNAP II, and the crab claw architecture is conserved (**Figure 1.1**). This structural conservation suggests that detailed mechanistic studies conducted on the simpler bacterial RNAPs should facilitate the understanding of the processes that govern gene expression in higher organisms.

A. Yeast RNA Pol II **B.** *Taq* RNAP core
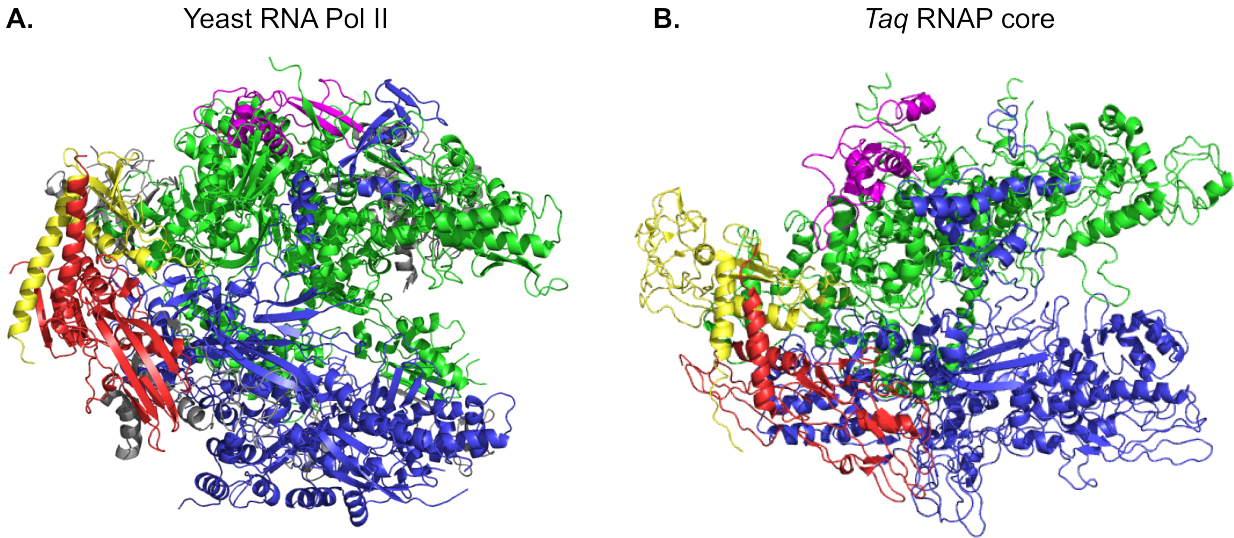
Figure 1.1. Structural conservation between eukaryotic and bacterial multisubunit

RNAPs. The high-resolution X-ray crystal structures for (**A**) *Saccharomyces cerevisiae* (yeast)

RNA Polymerase II (Protein Data Bank [PDB] ID code 1I50; consisting of 10 of the 12 subunits

that make up RNAP II) and (**B**) *Thermus aquaticus* (*Taq*) RNAP core enzyme (PDB ID code

1HQM) are shown. Subunits in *Taq* core RNAP, and the homologous subunits in yeast RNA Pol

II [reviewed in 24], are colored as follows: α_1 and α_2 (Rpb3 and Rpb11 in RNA Pol II) in red and

yellow; β (Rpb2 in RNA Pol II) in blue; β' (Rpb1 in RNA Pol II) in green; and ω (Rpb6 in RNA Pol

II) in magenta. The dark grey ribbons in (**A**) represent subunits in yeast RNA Pol II that have no

homologue in *Taq* core RNAP (Rpb5, Rpb8, Rpb9, Rpb10, and Rpb12). The two enzymes have

a conserved crab claw architecture, with β (Rpb2) and β' (Rpb1) as the 'pincers' that form a cleft

that directs dsDNA to the catalytic site at the base of the cleft.

The structures depicted here were generated by importing structural data from the

corresponding PDB files into PyMOL and subsequently coloring the appropriate polypeptides

(subunits). An extension modeled onto the β' subunit in the *Taq* core RNAP structure is

excluded for simplicity.

The bacterial RNAP core enzyme ($\alpha_2\beta\beta'\omega$) contains all of the catalytic machinery required for RNA polymerization (transcript synthesis) but lacks the sequence-specific DNA binding capability required to recognize promoters and initiate transcription. Specific recognition of various DNA promoter regions is imparted on the core enzyme by its reversible association with a class of initiation factors known as sigma (σ) factors to form the RNAP holoenzyme (subunit composition $\alpha_2\beta\beta'\omega\sigma$; abbreviated 'E•$\sigma$') [reviewed in 57, 104]. The high-resolution X-ray crystal structures of the *T. aquaticus* [106], *Thermus thermophilus* [138], and—most recently—*E. coli* [103, 152] holoenzymes, as well as the *T. aquaticus* holoenzyme/DNA complex [105] have greatly enhanced our understanding of σ/core and holoenzyme/promoter interactions (discussed in subsequent sections of this chapter) and the role these interactions play in the mechanics and regulation of gene expression.

The transcription cycle

Even in the comparatively simple bacterial systems, transcription is a complex, multistep process that can be divided into four general stages: 1) initiation, 2) promoter escape, 3) elongation, and 4) termination **(Figure 1.2)**.

Transcription initiation

Initiation involves the sequence-specific recognition of duplex promoter DNA to form a transcriptionally inactive closed complex (RP_C), followed by a series of conformational changes (isomerizations) that culminate in strand separation (melting) to expose the transcription start site and the formation of a transcriptionally competent open complex (RP_O) at the promoter. The stepwise chemical reactions that constitute this process can be represented by a three-step mechanism characterized by the appropriate equilibrium and rate constants **(Figure 1.3A)**. Each promoter has a unique set of kinetic parameters determined by DNA sequence and
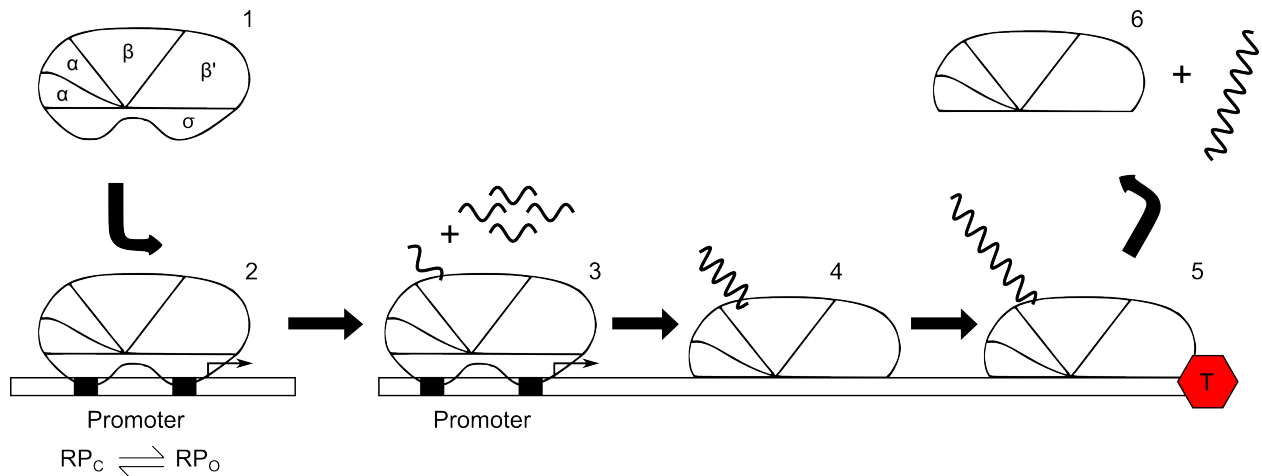
Figure 1.2. The transcription cycle. Upon association of a specific σ factor with core RNAP, the RNAP holoenzyme (1) binds in a sequence-specific manner to duplex promoter DNA to form a transcriptionally inactive closed complex (RP_C) and then isomerizes to form a transcriptionally-competent open complex (RP_O) (2). The RNAP holoenzyme undergoes multiple rounds of abortive transcription, synthesizing and releasing short transcripts whilst remaining bound at the promoter (3). RNAP eventually breaks favorable protein/DNA contacts with the promoter and escapes into productive elongation, during which the growing transcript remains stably associated with RNAP (4). Nucleotide addition continues until the elongating RNAP encounters a termination signal (T) (5), which causes RNAP to dissociate from template DNA and release the mature RNA transcript (6). Free core RNAP can then associate with another σ factor to repeat the cycle. Transcription regulation occurs at all stages of the cycle.
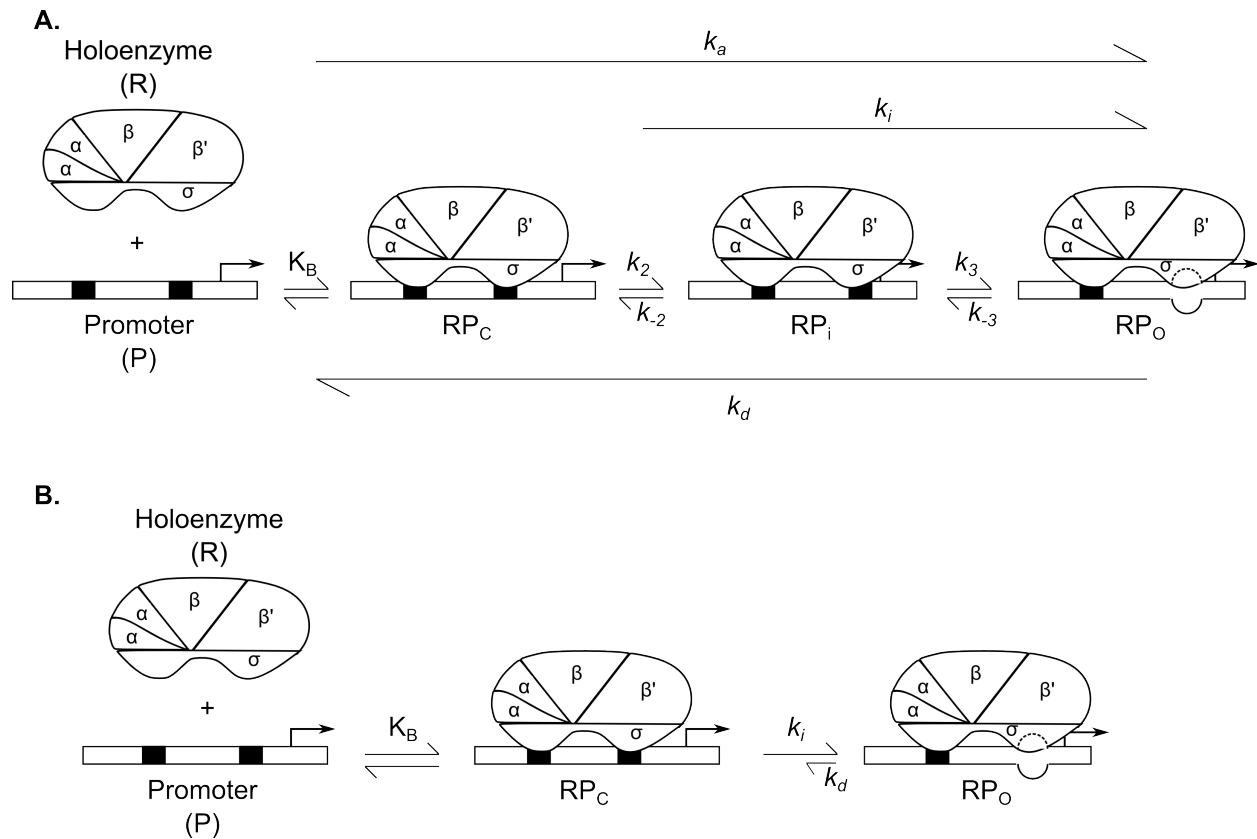
Figure 1.3. Reaction mechanisms to represent transcription initiation. The stepwise

chemical reactions that constitute the complex process of transcription initiation can be

schematized and the separate events represented by equilibrium and rate constants. **(A)**

Generalized three-step reaction mechanism for all promoters. Free RNAP holoenzyme

(represented by R) first associates reversibly with duplex promoter DNA (represented by P) to

form a transcriptionally incompetent closed complex (RP_C). This sequence-specific binding

event is represented by the equilibrium binding constant K_B. Promoter regions contacted by

RNAP holoenzyme in RP_C are upstream of the transcription start site, +1 (represented by the

arrow). RP_C isomerizes reversibly through at least one closed intermediate (RP_i) with an

expanded footprint that spans the transcription start site to form a binary open complex (RP_O),

where the DNA strands surrounding the transcription start site are separated. **(Legend**

continued on next page.)

Figure 1.3 (continued) The kinetic constant that describes the rate of formation of RP_O from free holoenzyme (R) and promoter DNA (P) is described by k_a, the overall association rate constant; k_i is the composite isomerization constant representing all steps from RP_C to RP_O; and k_d is the composite dissociation constant of RP_O to R + P, indicative of the stability of RP_O. (Adapted from [61, 119]). **(B)** Simplified two-step reaction mechanisms for promoters with stable open complexes. At these promoters, the isomerization of the initial RP_C to the final binary RP_O is essentially irreversible, so RP_i can be omitted and k_d is negligible. Most *E. coli* promoters can be represented by this simplified mechanism. (Adapted from [119]).

Promoters can be rate-limited at any step(s) of the mechanism(s), and transcription activators act at the rate-limiting step(s) to increase K_B or k_i (or both) or decrease k_d.

solution conditions that define the overall rate of initiation at that promoter (termed 'promoter strength') [3, 98, 119]. Transcription initiation can be rate-limited at any step in the mechanism and the kinetics modulated by the action of *trans*-acting factors; as a result, much regulation of bacterial—and likely eukaryotic—gene expression occurs at the initiation step (see the **Regulation of transcription initiation** section of this chapter).

Most *E. coli* promoters form stable RP_O's, so the isomerization(s) of the initial RP_C to the final RP_O are essentially irreversible. At these promoters, initiation can be further simplified to a two-step mechanism, where RP_C is in rapid equilibrium with free holoenzyme and duplex promoter DNA (represented by K_B) and the isomerization(s) of RP_C to RP_O are defined by a composite forward reaction rate (k_i), with negligible dissociation of RP_O to free holoenzyme and promoter DNA (represented by k_d) **(Figure 1.3B)** [98, 119]. The rate of initiation at the majority of *E. coli* promoters is, thus, limited by either the initial specific binding of holoenzyme to the promoter or by the conformational changes in the holoenzyme/promoter complex required to melt the promoter and expose the transcription start site. The ribosomal RNA (rRNA) promoters (*rrn* P1 and P2) are a notable exception: these promoters form very unstable, short-lived RP_O's [51, 89], and the dissociation of RP_O (described by k_d) seems to be rate-limiting for transcription [3]. *rrn* promoters are discussed in more detail in the **Promoter architecture** section below.

σ factors are critical during all steps of transcription initiation [reviewed in 57, 104]: they are responsible for promoter recognition by making sequence-specific contacts with duplex promoter DNA in RP_C's [19, 76, reviewed in 104], and they are also directly involved in promoter melting and open complex stability [39, 115]. Due to their diverse—and crucial—roles in the first stage of gene expression, σ factors are often targeted by transcription regulators. One such regulator that targets a primary bacterial σ factor is the focus of this dissertation.

Promoter escape

Once the transcription start site has been exposed by promoter melting, RNAP can initiate transcript synthesis. RNAP holoenzyme initially remains bound at the promoter and directs the reiterative synthesis and release of short (~2-12-nucleotide (nt)) abortive RNA products [72, 120]. This process necessitates the transient unwinding and accommodation of excess downstream DNA in the enzyme's main channel (termed 'scrunching'), which results in strain manifested as a distortion in the DNA backbone [81, 120]. The accumulated strain can be relieved by the release of the abortive RNA product, resetting RNAP for another attempt at transcript synthesis, or by breaking the favorable contacts the holoenzyme makes with promoter DNA so the enzyme can escape into productive elongation (called 'promoter escape'). Like the various steps of initiation, promoter escape can be rate-limiting; furthermore, *trans*-acting factors (e.g. transcript cleavage factors GreA and GreB) have been identified that can regulate transcription at the promoter escape step [reviewed in 72].

Productive elongation and termination

In the elongation phase of transcription, RNAP translocates along DNA and synthesizes the complete RNA transcript, which remains stably associated with the enzyme until the elongating RNAP encounters a termination signal. At this point, RNAP releases the mature transcript and dissociates from DNA (the 'termination' stage of transcription). Upon release of the transcript and dissociation from DNA, core RNAP can associate with another σ factor and repeat the transcription cycle, which enables the rapid reprogramming of bacterial gene expression. The rate of translocation during elongation is not constant: the transcription elongation complex may encounter and respond to nucleic acid sequence-encoded pause sites and arrest sites that affect RNAP processivity and have regulatory functions *in vivo*. Several regulators that affect the elongation and termination properties of bacterial RNAP have been

characterized, including the Gre transcript cleavage factors, the general elongation Nus factors, RfaH, and the operon-specific bacteriophage λ Q and N antiterminators [reviewed in 7, 88].

Sigma (σ) factors

As mentioned previously, the association of σ with core RNAP to form holoenzyme is required for sequence-specific promoter recognition and transcription initiation. Bacterial genomes typically encode multiple σ factors—a single primary (or 'housekeeping') σ factor and several alternative σ factors—with different promoter specificities that endow bacteria with a transcriptional plasticity that enables them to survive in diverse environments and withstand a variety of stresses. Primary σ factors are essential proteins responsible for initiating the transcription of most genes during exponential growth; alternative σ factors are needed to transcribe specific regulons in response to environmental changes, growth transitions, and morphological or developmental changes [reviewed in 58, 114, 141]. Bacterial genomes encode different numbers of σ factors, correlated with the metabolic complexity of the organism—for example, the obligate intracellular bacterium *Mycoplasma genitalium* has a single (primary) σ factor [41], whereas the metabolically diverse bacterium *Streptomyces coelicor* has one primary and 62 alternative σ factors [5]. Of the bacterial species relevant to this dissertation: 1) *E. coli* has seven σ factors: the primary factor, σ^{70} (product of the *rpoD* gene), and six alternative factors; 2) the low G+C Gram-positive bacterium *Staphylococcus aureus* (*S. aureus* or *Sau*) has four characterized σ factors: the primary, σ^{A} (encoded by *sigA*, also known as *plaC*), and three alternative factors; and 3) the related low G+C Gram-positive spore-forming bacterium *Bacillus subtilis* (*B. subtilis* or *Bsu*) has eighteen σ factors: the primary, σ^{A} (encoded by *sigA*), and seventeen alternative σ factors [reviewed in 58, 133].

σ factors are divided into two distinct families based on phylogenetic relatedness: the σ^{70} family found in all bacteria and the σ^{54} family present only in some species. σ^{54} family members

have no structural or sequence homology to σ^{70} and initiate transcription by a distinct

mechanism that requires the action of an ATP-dependent activator that interacts with σ^{54} in the

context of the holoenzyme [reviewed in 49]. This dissertation focuses exclusively on σ^{70} family

members, specifically the primary σ factors of *E. coli* (*Eco* σ^{70}) and *S. aureus* (*Sau* σ^A) (see

Chapters 2 and **3**). The primary σ factor of *B. subtilis* (*Bsu* σ^A) is investigated in the context of a

future experimental avenue to explore (see primarily **Chapter 4**).

σ^{70}-type σ factors are modular proteins with up to four regions of high amino acid

conservation that comprise four flexibly linked domains: $\sigma_{1.1}$, σ_2, σ_3, and σ_4. Domains σ_2–σ_4 have

been crystallized [19, reviewed in 57, 58, 93, 104, 114]. These conserved regions can be further

divided into subregions: region 1.1 (which comprises domain $\sigma_{1.1}$), regions 1.2-2.4 (domain σ_2),

regions 3.0-3.1 (domain σ_3), and regions 4.1-4.2 (domain σ_4); additionally, there is an extended

linker connecting domains σ_3 and σ_4 (the $\sigma_{3.2}$ loop), and—specifically in primary σ factors—a

non-conserved region (σ_{NCR}) between regions 1.2 and 2.1 that varies in sequence and length

between different σ factors [reviewed in 58, 93, 104] **(Figure 1.4)**. Primary σ factors contain all

four conserved regions and the σ_{NCR}, but alternative σ factors lack region 1.1, and many also

lack region 3.0. σ_2 (especially region 2.2) and σ_4 exhibit the highest degree of conservation

across all species, implicating them in the essential σ functions of core RNAP binding

(holoenzyme formation) and sequence-specific promoter recognition [reviewed in 93].

σ^{70}-family σ factors direct transcription from promoters containing two conserved

hexameric sequence elements centered at -10 (consensus sequence TATAAT) and -35

(consensus sequence TTGACA) relative to the transcription start site and separated by 16-18

base pairs (bp) of nonconserved sequence (optimal 17-bp spacer for most *E. coli* promoters)

[57]. Alternative σ factors also direct transcription from -10/-35 promoters, but these elements

have distinct consensus sequences [67]. Conserved regions 2 (σ_2) and 4 (σ_4) make direct

contacts with the -10 and -35 promoter elements, respectively (see **The σ/promoter interface**
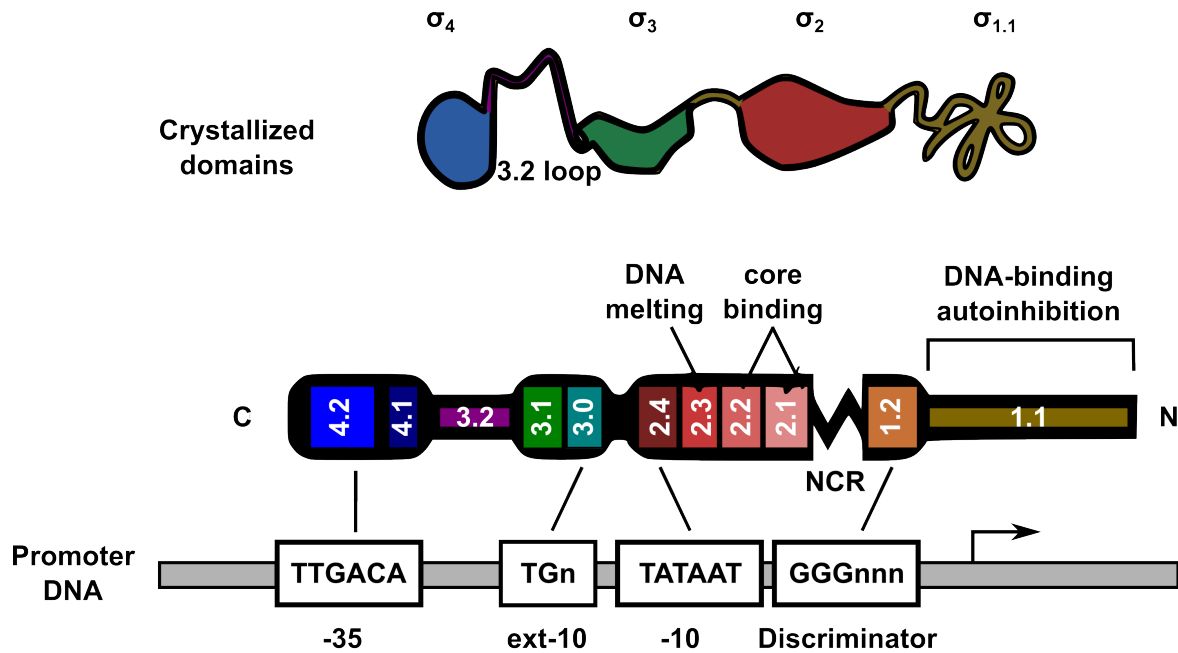
Figure 1.4. Domain structure of primary σ factors. Schematic diagram showing the arrangement of conserved regions across the primary sequence of *Eco* σ^{70}, applicable to all primary σ factors. The subdomains that make base-specific contacts with the -10, extended -10, and -35 promoter elements as well as the discriminator region are indicated below in a schematic representation of a promoter. Other functions of non-DNA-binding subregions are also indicated. The diagram at the top represents the three domains observed by X-ray crystallography connected by flexible loops. NCR ≡ non-conserved region. (Adapted from [104, 113]).

See the text (**The σ/core interface** and **The σ/promoter interface** sections of this chapter) for details about these interactions.

section of this chapter), but these contacts occur only in the context of holoenzyme: free σ^{70} cannot engage in sequence-specific DNA recognition. This is largely attributable to the highly acidic autoinhibitory $\sigma_{1.1}$ domain, which occludes the DNA-recognition domains of σ_2 and σ_4 in free σ^{70}. Deleting this domain allows weak binding of free σ^{70} to promoter DNA [27, 28]. A functionally equivalent rearrangement of $\sigma_{1.1}$ that unmasks the DNA-binding domains of σ^{70} occurs upon holoenzyme formation [17]. σ binding to core RNAP induces significant conformational changes in both partners that result in an extensive σ/core/promoter interface with important functional implications in transcription initiation and elongation (see **The σ/core RNAP interface** and **The σ/promoter interface** sections below).

The σ/core RNAP interface

Genetic and biochemical studies have revealed contacts between all four conserved regions of *Eco* σ^{70} (and, by homology, those of all primary σ factors) and core RNAP [59, 130], suggesting an extensive σ^{70}/core RNAP interface in the σ^{70}-containing holoenzyme (E•σ^{70}). This interface of ~8,500 $Å^2$ has been borne out and refined by the X-ray crystal structures of E•σ^A from *T. aquaticus* [106] and *T. thermophilus* [138] and E•σ^{70} from *E. coli* [103, 152], which reveal four separate contact surfaces between the primary σ and core RNAP in the context of holoenzyme: 1) $\sigma_{1.1}$ contacts the active site channel; 2) σ_2 contacts a coiled-coil motif on β' (also referred to as the β' clamp helices); 3) the $\sigma_{3.2}$ loop binds the RNA exit channel; and 4) σ_4 binds a flexible flap domain on β [reviewed in 104]. Each individual interaction is moderately weak and individual contacts are broken sequentially to release σ from E•σ during the transcription cycle, but the sum of the interactions results in a high-affinity complex with a dissociation constant of ~10^{-9} M [50]. All of these interactions have important functional implications for the properties of RNAP at various stages of the transcription cycle. The majority of the genetic and biochemical studies to investigate the binding of σ to core RNAP have been conducted in *E. coli*; since *Eco*

σ^{70} is central to this dissertation, the σ/core RNAP and σ/promoter interfaces will be described in the context of *Eco* σ^{70} and *Eco* E•σ^{70}.

Contacts of $\sigma_{1.1}$: the active site channel

As mentioned previously, the highly acidic $\sigma^{70}_{1.1}$ autoinhibits promoter recognition by free σ^{70} by occluding the DNA-binding determinants of σ^{70}_2 and σ^{70}_4. Binding of σ^{70} to core RNAP induces a conformational change in σ^{70} that moves $\sigma^{70}_{1.1}$ ~20 Å with respect to σ^{70}_2 and exposes the DNA-binding determinants of σ^{70} [16, 17]. Once this core RNAP binding-induced displacement has occurred, $\sigma_{1.1}$ has been shown by fluorescence resonance energy transfer (FRET) experiments [100] and X-ray crystallography [103] to bind to basic residues in the enzyme's active site channel and block the access of DNA to the catalytic site. RP_O formation necessitates the disruption of the $\sigma^{70}_{1.1}$/active site channel interaction and the movement of $\sigma^{70}_{1.1}$ by ~50 Å [100].

Contacts of σ_2: the β' coiled-coil

σ_2 is the most conserved domain of the σ polypeptide across all species, especially in region 2.2, consistent with the fact that this domain plays critical roles in the formation of holoenzyme and promoter utilization. Deletion analyses of σ^{70}_2 initially identified residues 361-390 (comprising region 2.1) as necessary and sufficient for core RNAP binding [90], strongly suggesting that σ_2 is the primary (and most extensive) determinant for high affinity binding of σ factors to core RNAP. Subsequent mutational analysis identified substitutions in regions 2.1 and 2.2 that severely impair the ability of σ^{70} to bind to core RNAP [130], and crystal structures of σ^{70}_2, *Taq* σ^A_2 and the *Taq* holoenzyme revealed that conserved residues in regions 2.1 and 2.2 form an exposed polar surface that lies at the σ/core interface [19, 96, 106]. The binding site for σ^{70}_2 on core RNAP was originally identified by far-Western and co-immobilization experiments to

be a coiled-coil motif near the N-terminus of the β' subunit (the β' coiled-coil, abbreviated β'cc) comprised of RpoC residues 260-309 [2, 13]. The σ_2/β'cc interaction was subsequently validated by the various X-ray crystal structures of the bacterial holoenzyme [103, 106, 138, 152]. The binding of σ^{70}_2 to the β'cc is essential for holoenzyme formation and induces rearrangements in regions 2.3-2.4 that allow σ^{70}_2 to functionally engage the -10 promoter element and form the open complex [39, 97, 142] (see **The σ/promoter interface** section); in this context, it is also critical for the recognition of promoter-like pause elements during early elongation [121]. There is no apparent steric clash between σ^{70}_2 bound at the β'cc and the nascent transcript of any length, so disruption of the σ^{70}_2/β'cc contact is likely the last step in the release of σ from most elongating RNAPs and occurs after the disruption of other σ/core RNAP contacts have destabilized σ binding to core RNAP (see below).

Contacts of $\sigma_{3.2}$: the RNA exit channel

The $\sigma^{70}_{3.2}$ loop, which connects domains σ_3 and σ_4 of σ^{70}, occupies the RNA exit channel upon holoenzyme formation and constitutes a third contact between σ^{70} and core RNAP. (The RNA exit channel provides a path for the nascent RNA transcript from the catalytic site to the surface of the enzyme as elongation proceeds.) Extension of the nascent RNA chain beyond ~9 nucleotides (nt) causes a steric clash between $\sigma^{70}_{3.2}$ located in the exit channel and the transcript as it enters the channel, so $\sigma^{70}_{3.2}$ must be displaced for productive elongation to occur [8, 106]. The nascent transcript and $\sigma^{70}_{3.2}$ compete for binding to these as yet unidentified exit channel determinants, which affects the stability of transcript binding and modulates abortive initiation and promoter escape. Displacement of the $\sigma^{70}_{3.2}$ loop from the RNA exit channel by the nascent transcript—along with the dislodgment of $\sigma^{70}_{1.1}$ from the active site channel upon RP_O formation—destabilizes the binding of σ^{70} to core RNAP by disrupting two of the four favorable

σ^{70}/core contacts and likely facilitates the release of σ^{70} later in the transcription cycle. Amino

acid substitutions in the $\sigma^{70}_{3.2}$ loop weaken the overall binding of σ^{70} to core RNAP [130].

Contacts of σ_4: the β-flap

The final major contact between σ and core RNAP—and the most relevant to this

dissertation—involves the second-most-highly-conserved domain, σ_4, which contains the

determinants for recognition of the -35 promoter element (see **The σ/promoter interface**

section). Genetic analyses implicated region 4.1 in core RNAP binding [80, 130], and

subsequent evidence indicated that σ^{70}_4 interacts directly with the flexible flap domain in the β

subunit of RNAP (comprising RpoB residues 858-946) [85], consistent with an early study

showing that the binding of σ^{70} to core RNAP protects a protease sensitive site within the flap

region from cleavage [40]. The crystal structure of *Taq* σ^A revealed that residues in region 4.1

make up a concave hydrophobic pocket large enough to accommodate an α-helix [19], and

crystal structures of the RNAP holoenzyme confirmed the binding of σ^{70}_4 to the β-flap and

identified an α-helix at the tip of the β-flap (the β-flap tip helix; RpoB residues 900-909) that fits

into the hydrophobic pocket formed by σ region 4.1 [103, 106, 138, 152]. This interaction was

corroborated by biochemical experiments [48].

The σ^{70}_4/β-flap interaction is not essential for holoenzyme formation, but it is required for

transcription initiation from -10/-35 promoters, the major class of bacterial promoters [85]. This

can be explained by the 16-18 bp spacer separating the -35 hexamer (recognized by σ^{70}_4) from

the -10 hexamer (recognized by σ^{70}_2) at these promoters, which requires a significant increase

in the interdomain distance between σ^{70}_2 and σ^{70}_4 so both elements can be bound

simultaneously during initiation. In free σ^{70}, σ^{70}_2 and σ^{70}_4 are in close proximity (measured by

luminescence resonance energy transfer, LRET), but binding of σ^{70} to core RNAP causes region

4.2 to move by ~15 Å with respect to σ^{70}_2 [17, 85]. This movement puts σ^{70}_4 in a position

compatible with -35 element recognition and is strictly dependent on the σ^{70}_4/β-flap interaction: in E•σ^{70} reconstituted using Δβ-flap core RNAP, the σ^{70}_2 /σ^{70}_4 interdomain distance is similar to that seen in free σ^{70} [85]. (The suboptimal σ^{70}_2/σ^{70}_4 distance in the absence of core RNAP binding also contributes to the lack of sequence-specific DNA recognition by free σ and likely explains why even alternative σ factors that lack $\sigma_{1.1}$ do not bind DNA specifically in solution). The β-flap also undergoes a conformational change upon σ^{70} binding to core RNAP: the flap-tip helix shifts by ~11 Å relative to its position in core RNAP to fit into the hydrophobic pocket of σ^{70}_4 [103, 106, 138, 150, 152]. The flexibility of the β-flap—even when complexed with σ^{70}_4— allows for variability in spacer length at -10/-35 promoters: changing the angle at which the β-flap protrudes from the enzyme changes the σ^{70}_2/ σ^{70}_4 interdomain distance to accommodate the different distances required by a 16-bp versus a 17-bp versus a 18-bp spacer [105, 144].

Binding of σ^{70}_4 to the β-flap during holoenzyme formation positions σ^{70}_4 adjacent to the end of the RNA exit channel [103, 106, 138, 152] and poses a steric barrier to the nascent transcript as it emerges from the channel. Elongation of the transcript beyond this point (addition of the 17th nt) requires the disruption of the σ^{70}_4/β-flap interaction [110]. In combination with the displacements of $\sigma^{70}_{1.1}$ from the active site channel and of $\sigma^{70}_{3.2}$ from the RNA exit channel (described previously) earlier in the transcription cycle, loss of σ^{70}_4/β-flap contacts further destabilizes the binding of σ^{70} to core RNAP and favors the release of σ^{70} from mature elongation complexes. Additionally, the binding of σ_4 to the β-flap and the -35 promoter element makes σ_4 a prime target for a variety of transcription regulators that act at initiation and thus modulate the transcriptional profile of the cell (refer to the **Regulation of transcription initiation** section).

Promoter architecture

The rate of transcription initiation from a promoter defines its strength and, thus, the levels of RNA produced by different operons at any time: strong promoters initiate at a high rate and result in abundant transcript; weaker promoters initiate less frequently and produce fewer transcripts per cell generation. Since bacteria have a single core RNAP species, some differences in promoter activity are the result of competition between different σ factors (with different promoter specificities) for binding to core RNAP to establish the predominant holoenzyme species (E•σ). Even among promoters recognized by the same holoenzyme, however, promoter strength can differ over an enormous range [reviewed in 12, 62, 119], which suggests that promoter strength is a function of the DNA sequence. Early alignments of *E. coli* promoter sequences [60, 64] identified the -10 and -35 hexamers (elements) as the principal defining features of bacterial promoters and their relative strengths, but several additional motifs that modulate strength and regulation have been identified. In this context, promoters can be thought of as being ***modular***: composed of different combinations of sequence elements (motifs, or modules) that together define promoter strength. The motifs are: 1) the -10 element; 2) the -35 element; 3) the extended -10 element; 4) the spacer; 5) the discriminator; 6) the UP element; and 7) transcription regulator binding sites. Promoters differ in the number and combination of these motifs they possess and the contributions of each to promoter strength; additionally, there is sequence variability in the individual motifs amongst different promoters.

The -10 and -35 elements and the spacer

The -10 and -35 promoter elements are the most extensively studied and well-characterized motifs. The -35 element is recognized in a sequence-specific manner by σ_4 bound to the β-flap, and the -10 element—where melting begins during open complex formation—is engaged by σ_2 bound to the β'cc (see **The σ/promoter interface** section for details). Promoters

that depend on both a -35 and a -10 elements (termed '-10/-35 promoters') constitute the major class of bacterial promoters and are transcribed by holoenzymes bearing the primary σ factor or the majority of alternative σ factors. The presence of the spacer—a sequence-variable arrangement of bases that separates the -10 and -35 elements—is an inherent feature of all -10/-35 promoters, which have the architecture depicted in **Figure 1.5A**.

The -10 and -35 elements of promoters recognized by the primary σ factor (e.g. *Eco* σ^{70}, *Sau* σ^A, or *Bsu* σ^A) are hexamers with consensus sequences TATAAT and TTGACA, respectively; the consensus sequences for alternative σ-dependent promoters differ but have the same general architecture [reviewed in 15, 38, 67]. Substitutions in these sequences that alter each hexamer towards the consensus stabilize the binding of holoenzyme at these promoters and improve their strength, whilst substitutions that alter the promoter sequence away from consensus weaken the promoter and result in reduced activity [98]. Naturally occurring strong promoters (e.g. the bacteriophage λ P_R and $P_{R'}$ promoters, or the bacterial rRNA *rrn* promoters) have -10 and -35 sequences that closely match the consensus. The length of the spacer region determines the distance between σ_2 and σ_4 required to contact both elements simultaneously during transcription initiation. For the primary σ-dependent promoters studied, the spacer length is generally 16-18 bp, with 17 bp being the preferred length [57, 60]. As with the -10 and -35 elements, moving away from the optimal spacer length reduces core promoter strength, and most strong promoters have the 17-bp spacer. The *rrn* promoters are notable exceptions: these remarkably strong promoters have near-consensus -10 and -35 elements, but a suboptimal 16-bp spacer (see below for a detailed discussion of *rrn* promoters). Promoters recognized by alternative σ factors differ from those recognized by the primary σ factor in spacer length; for σ^{28}-dependent and σ^{32}-dependent promoters, the spacer appears to be in the range of ~11-15 bp [140, 145].

A. Canonical -10/-35 promoter

B. TGn/-10 ext-10 promoter

C. UP element-dependent promoter

D. *rrnB* operon

rrnB P1promoter

E. Class I activator-dependent promoter
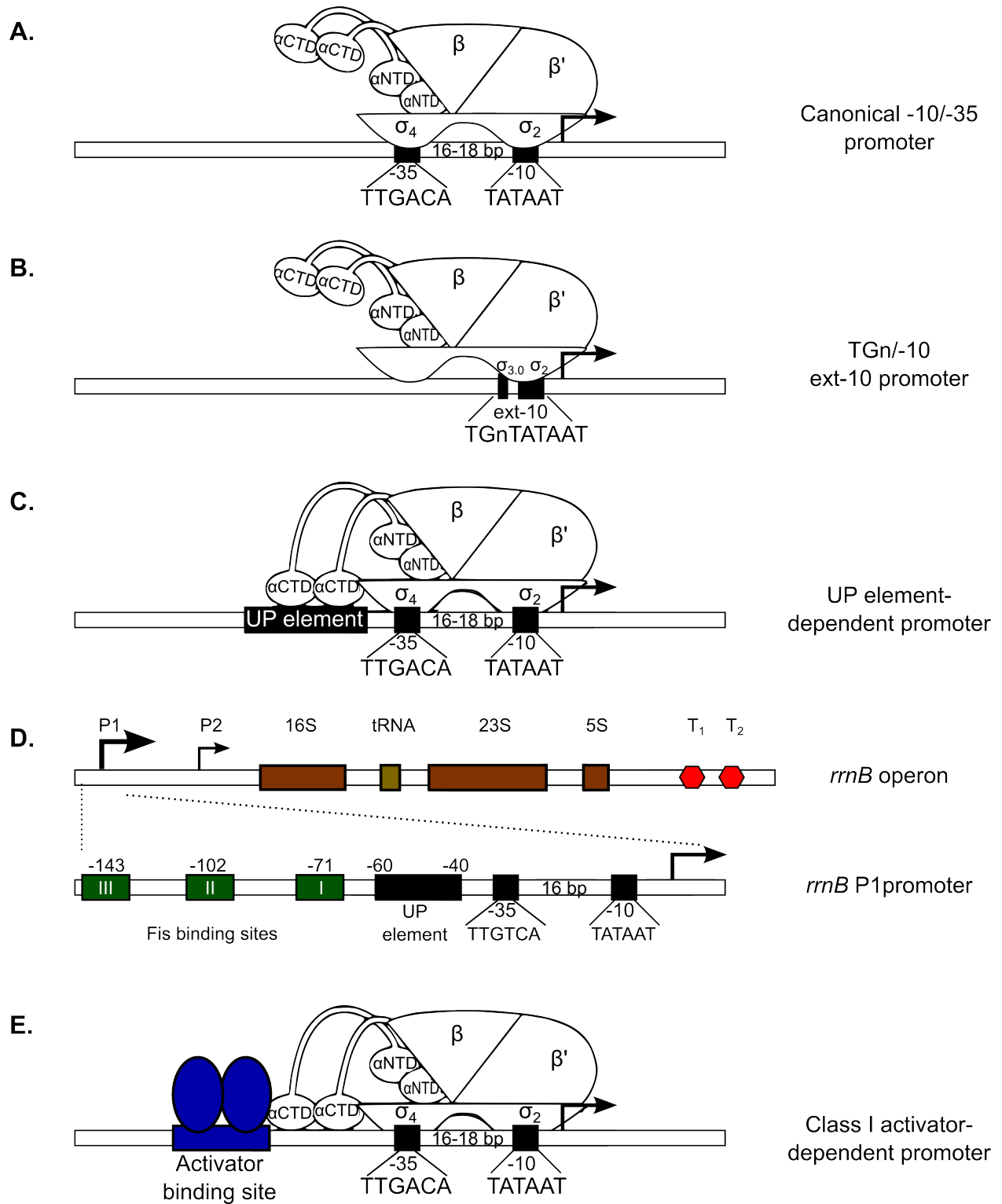
Figure 1.5. Bacterial promoter architectures. Schematic diagram of different classes of promoter architectures found in bacteria. **(Legend continued on next page.)**

Figure 1.5 (continued) See text for details on each type of promoter architecture. **(A)** Canonical -10/-35 promoter. The defining features of this promoter class are the presence of a -10 and a -35 element separated by a 16-18 bp spacer. σ_4 makes base-specific contacts with the -35 element (consensus sequence TTGACA), and σ_2 makes contacts with the -10 element (consensus sequence TATAAT). **(B)** TGn/-10 extended -10 promoter. The defining feature of this promoter class is the presence of a conserved TG dinucleotide 1 bp upstream of the -10 element (the TGn motif). Region 3.0 of σ makes additional contacts with the TGn motif and obviates the need for sequence-specific contacts between σ_4 and a -35 element for efficient transcription initiation. **(C)** UP element-containing promoter. This class of promoters is defined by the presence of a ~20 bp A+T-rich sequence (consensus sequence nnAAAWWTWTTTTnnnAAAnnn, where n ≡ any base and W ≡ A or T) immediately upstream of the -35 element. The UP element contains two subsites that interact sequence-specifically with the αCTDs (one αCTD per subsite) and stabilize the holoenzyme/promoter complex, stimulating transcription. **(D)** The *rrnB* operon and *rrnB* P1 promoter. rRNA (*rrn*) promoters are the most well-characterized UP element-containing promoters in *E. coli*. The seven *rrn* operons each encode a single transcript processed to make 16S, 23S, and 5S rRNAs and at least one tRNA, and transcription initiates from two promoters (*rrn* P1 and P2) spaced ~120 bp apart and terminates at terminators T1 or T2. The *rrnB* operon is shown, with the inset depicting the architecture of the *rrnB* P1 promoter, which consists of a core promoter containing the -10 and -35 elements, an UP element, and three binding sites for the transcription activator Fis upstream of the UP element. The number of Fis binding sites ranges from three to five, depending on the *rrn* operon. (Adapted from [116]). **(Legend continued on next page.)**

Figure 1.5 (continued) (E) Promoter with transcription activator binding site. This promoter class is defined by the presence of a binding site for a particular transcription activator that contacts the RNAP holoenzyme directly when bound at its specific recognition site; such activators function by recruiting holoenzyme to the promoter (increasing K_B) and/or facilitating RP_O formation (increasing k_i). Activator binding sites can be upstream of the core promoter so the activator interacts with the αCTD (class I activation), or partially overlapping the -35 promoter element (class II activation); in the latter case, the activator often interacts with σ_4 bound at the -35 element. An example of a class I promoter is shown.

<u>The extended -10 motif</u>

An additional sequence element important at some *E. coli* σ^{70}-dependent promoters is a conserved TG motif located 1 bp upstream of the -10 hexamer (TGn, where n $\equiv$ any base) [reviewed in 9]. Promoters containing this feature are referred to as 'extended -10' (abbreviated 'ext-10') or 'TGn/-10' promoters. The architecture of a canonical extended -10 promoter is shown in **Figure 1.5B**. Surface-exposed residues on $\sigma^{70}_{3.0}$ make sequence-specific contacts with the TGn motif [4], providing an additional recognition point for holoenzyme that is sufficient to eliminate the need for a -35 element and σ^{70}_4/β-flap contacts in transcription initiation. Initially, extended -10 promoters were believed to make up a small proportion of *E. coli* promoters and to share the common feature of a poor -35 element [9], but more recent work suggests TGn-containing promoters constitute ~20% of *E. coli* promoters, and many of these have good matches to consensus in the -35 region [101]. Two extended -10 promoters with near-consensus -35 elements and TGn motifs but poor -10 elements have been characterized: the *E. coli gapA* P1 promoter (directs transcription of GAPDH, a key enzyme for glucose metabolism) [135] and the bacteriophage T4 P_{minor} promoter [71]. (This class of extended -10 promoter is dubbed '-35/TGn'). These observations suggest that the $\sigma^{70}_{3.0}$/TGn contact is crucial for transcription from promoters with poor matches to <u>either</u> the -35 or -10 consensus hexamers, and that the extended -10 motif can work in conjunction with any combination of the other sequence motifs to modulate basal promoter strength.

Extended -10 motifs are common in primary σ-dependent promoters in Gram-positive bacteria (e.g. *Lactobacillus*, *Streptococcus*, *Lactococcus*, and *Bacillus* species) and often occur in combination with -10 and -35 elements with good matches to the consensus [18, 66, 139]. The extended -10 motif is generally expanded in Gram-positives (consensus sequence TRTGn, where R $\equiv$ purine) and is found in ~45% of *B. subtitlis* promoters, particularly those with A+T-

rich upstream sequences [66, 139], which suggests that Gram-positives readily use the interplay between these different elements to fine-tune promoter activity and gene expression.

The discriminator sequence

The discriminator is the region of the promoter between the -10 element and the transcription start site (+1), originally identified as a G+C-rich region in stable RNA (rRNA and tRNA) promoters believed to play a role in their regulation [136] but now known to be a general feature of primary σ-dependent promoters. $\sigma_{1.2}$ contacts the discriminator region, specifically at the non-template strand base two positions downstream from the -10 hexamer [61, 63]. This interaction affects the rate of dissociation of RP_O (k_d), which is the rate-limiting step in initiation at stable RNA promoters and explains the critical role of the discriminator in *rrn* regulation. A guanine base at this position (on the template strand) results in the strongest interaction with $\sigma_{1.2}$ and thus the most stable RP_O's; a cytosine base severely weakens the interaction and results in very unstable RP_O's [61, 63]. The majority of *E. coli* promoters form long-lived RP_O's and are not rate-limited at k_d, so most promoters disfavor a C at this position but otherwise show no base preference [63]. The stable RNA promoters, where RP_O instability is a key feature of regulation, have a very strong preference for the C that results in weakened $\sigma_{1.2}$ binding [60].

The UP element

The separable promoter motif that is most relevant to this dissertation is the upstream (UP) element, a ~20-bp A+T rich region located immediately upstream of the -35 element [123]. UP elements interact directly with the C-terminal domain of the α subunit of RNAP (αCTD) [6, 123] (see **The α subunit of RNAP** section of this chapter). This protein/DNA interaction recruits holoenzyme to UP element-containing promoters (increases K_B in initiation) and facilitates the formation of RP_C, but it may also increase the rate of later isomerization steps in the pathway to

RP_O formation [89, 118], which accounts for the remarkable strength of most UP element-containing promoters. The optimized (consensus) UP element sequence—which results in higher activation than known naturally occurring UP elements—was defined by *in vitro* selection to be nnAAAWWTWTTTTnnnAAAnnn (where n ≡ any base and W ≡ A or T) [35]. As with the other described promoter motifs, the strength of UP element-mediated transcription stimulation correlates with closeness to the consensus sequence, and naturally occurring UP elements vary widely (from <2-fold to ~90-fold) in their stimulatory effects [122, 123]. The UP element actually consists of a proximal and a distal subsite, with 'TTn' marking the transition between the two subsites [36]. The αCTD from each of the two α subunits binds in a sequence-specific manner to one of the subsites. The presence of either subsite upstream of the core promoter stimulates transcription (the effects of the proximal subsite are stronger, likely due to stabilizing protein/protein contacts between the αCTD bound at this site and σ_4 bound immediately downstream at the -35 element), but the maximal effect is obtained with the full UP element [36]. **Figure 1.5C** shows a promoter with a full UP element.

UP elements are moderately widespread in *E. coli*, with 2-3% of promoters having matches to consensus at ≥11 of 15 crucial positions [36, 53] and a greater number having close matches to the consensus in one subsite. UP element-containing promoters in *E. coli* direct the expression of a variety of genes and are not restricted to genes transcribed by E•σ^{70}: for example, *fliC* (transcribed by E•σ^{28}) has been shown to be UP element-dependent [42], and the promoter for σ^{28}-dependent *hemL* has a good match to the full consensus UP element but has not been assessed experimentally [36]. Based on sequence analyses, A+T-rich upstream sequences appear more prevalent in Gram-positive bacteria (e.g. *Clostridium* and *Bacillus* species), often in combination with extended -10 elements, but few have been tested directly [54, 66]. The residues of the αCTD involved in UP element recognition are almost universally

conserved, which suggests that αCTD/DNA interactions play a crucial role in the expression and regulation of at least a subset of genes.

rRNA promoters: structure and regulation

The most well characterized UP element-containing promoters in *E. coli* are the E•σ[70]-dependent rRNA promoters, particularly *rrnB* P1. Ribosome biogenesis presents a huge metabolic expense, so the expression of ribosome components is highly regulated at multiple levels to ensure the right number of ribosomes is synthesized to meet the cell's needs at a particular growth rate without squandering metabolic resources. At high growth rates (e.g. exponential growth in rich medium), cells have ~70,000 ribosomes and transcription from the seven *rrn* operons in *E. coli* can account for up to 70% of all cellular transcription [11, 82, 116]. Under less favorable nutritional conditions (e.g. stationary phase or amino acid starvation), the growth rate slows considerably and the rate of ribosome biosynthesis decreases concomitantly [46]. rRNA transcription is the rate-limiting step in ribosome biosynthesis in bacteria and is itself limited by the rate of transcription initiation at the *rrn* promoters, so much of the regulation of ribosome biogenesis is a function of the nature of the *rrn* promoters.

The seven *rrn* operons are transcribed by E•σ[70] from two promoters, *rrn* P1 and *rrn* P2, spaced ~120 bp apart; the P1 promoters are stronger during rapid growth and have been more thoroughly investigated, but both promoters are regulated by many of the same mechanisms [116]. The operons each contain three rRNA genes (16S, 23S, and 5S) and at least one tRNA gene transcribed as a single RNA. *rrn* promoters are remarkably strong promoters: the core promoters all contain consensus -10 elements and near-consensus -35 elements with a single deviation (TTG<u>T</u>CA versus the TTGACA consensus), but these features alone cannot account for their extraordinarily high activity. Sequences upstream of the *rrn* P1 core promoter were shown to increase transcription by as much as 300-fold [52, 89, 118, 123]. This upstream region

contains two sequence motifs responsible for *rrn* P1 promoter strength: a full UP element immediately upstream of the -35 element (positions -40 to -60), and binding sites for the transcription activator Fis upstream of -61. **Figure 1.5D** shows the architecture of the *E. coli rrnB* operon and the *rrnB* P1 promoter. The UP element is the major contributor to *rrnB* P1 promoter strength: the two subsites provide good αCTD binding sites and are optimally positioned relative to the core promoter, resulting in ≥30-fold stimulation *in vivo* and *in vitro* in the presence of only the basal transcription machinery [118, 123]. The αCTD bound at the proximal UP subsite also contacts σ^{70}_4 bound at the -35 element, further stabilizing the *rrnB* P1/holoenzyme complex and stimulating transcription [23, 123, 125]. Fis activates transcription three- to eight-fold by helping to recruit holoenzyme to *rrnB* P1 through a protein/protein contact with the αCTD and by facilitating later steps in transcription initiation [reviewed in 116].

Despite their extraordinary strength relative to other *E. coli* promoters, *rrn* promoters are optimized for efficient regulation rather than maximal strength: the -35 element is a 5/6 match to consensus, the spacer is 16 bp instead of the optimal 17 bp, they lack an extended -10 element (present in most of the *B. subtilis rrn* P1 promoters [66]), and the discriminator region is G+C-rich and has the disfavored C (template strand) at the -7 position contacted by $\sigma^{70}_{1.2}$ (see **The discriminator sequence** subsection, above). Substitutions in the -35 element to full consensus and insertions in the spacer to 17 bp greatly increase the activity of the core promoter [43], suggesting that *rrn* promoters sacrifice some core promoter strength in favor of modulating their activity quickly in response to changes in the cell's metabolic requirements. *rrn* promoters form extremely short-lived RP_O's (facilitated in part by a weak $\sigma_{1.2}$/discriminator interaction) and thus require a significantly higher concentration of the initiating nucleotide (iNTP) to stabilize the open complex and escape into productive elongation [44, 51, 89]. RP_O instability is crucial to the regulation of *rrn* promoters by small molecule effectors whose concentrations change quickly during growth transitions and nutritional shifts: the alarmone guanosine tetraphosphate (ppGpp)

and iNTP concentration. ppGpp, the levels of which increase during nutrient starvation, binds to RNAP and decreases RP_O half-life at all promoters; at *rrn* promoters—which already form short-lived complexes—this likely promotes RP_O collapse before NTP addition can occur and dramatically reduces rRNA synthesis [reviewed in 116]. In stationary phase, a state of minimal protein synthesis, NTP levels are low and thus the iNTP cannot readily bind to and stabilize *rrn* RP_O's, reducing rRNA synthesis and ribosome biogenesis [44]. The transcription factor DksA, which binds to RNAP and also reduces RP_O stability at all promoters, potentiates the effects of ppGpp and iNTP concentration on *rrn* promoters [reviewed in 116]. *B. subtilis rrn* promoters share some features with *E. coli rrn* promoters, but the nature of the core promoter—near-consensus extended -10 and -35 elements and the optimal 17-bp spacer—seems to contribute more to their strength than upstream sequences. *Bsu rrnB* P1 UP elements stimulated transcription only ~3-fold *in vitro* (compared to ~30-fold for *Eco rrnB* P1), and there appears to be no Fis-like activator [84]. Although rRNA transcription is also regulated by ppGpp and NTP concentration in *B. subtilis*, the mechanisms appear to be different [84].

Transcription factor binding sites

The promoter motifs that define various promoter architectures discussed so far involve the basal transcription machinery: the promoter and RNAP holoenzyme, with no additional proteins. The *E. coli* genome encodes ≥300 factors predicted to bind at promoters and up-regulate (activators) or down-regulate (repressors) transcription [94, 117]; this suggests that promoter strength is also modulated by *trans*-acting factors at many promoters. Typically, these regulators bind to specific DNA recognition sites near to or overlapping the RNAP holoenzyme binding site(s) and affect transcription initiation either indirectly or through direct contact with RNAP (usually with α or σ) [14, 70, 74]. Since these regulators are sequence-specific DNA binding proteins, they affect only promoters that have the corresponding binding site(s). The

presence or absence of these sites adds another layer of complexity to promoter architecture that contributes to the wide range of promoter activities in bacteria. **Figure 1.5E** shows an example of a promoter with a transcription factor binding site. Regulation of transcription initiation by *trans*-acting factors is discussed in more detail in **The regulation of transcription initiation** section of this chapter.

The σ/promoter interface

σ factors direct RNAP to promoter sequences during transcription initiation. Four regions of the primary σ, discussed here in the context of *Eco* σ^{70} based on the close homology with the available *Taq* σ^A structures, have been shown to play critical roles in promoter recognition: $\sigma_{1.2}$, σ_2, $\sigma_{3.0}$, and σ_4.

Promoter contacts of σ_4: the -35 element

Eco σ^{70}_4 contains a helix-turn-helix (HTH) DNA-binding motif spanning region 4.2. Genetic [47, 131] and structural [19, 105] data reveal that residues within the second helix of the HTH motif—termed the recognition helix—mediate base-specific contacts with duplex DNA at the promoter -35 element when σ^{70}_4 is positioned appropriately by its interaction with the β-flap in holoenzyme. A number of residues in *Eco* σ^{70}_4 make hydrogen bonds with bases in the -35 element duplex: residue R584 contacts the C:G bp at the fifth position (counting from the 5' end) of the -35 element; residue R585 contacts the G:C bp at the third position; and residue Q589 contacts the T:A bp at the first position. Several other residues within the recognition helix make sequence-specific van der Waals contacts with bases in the -35 element, and residues in *Eco* σ^{70}_4 both within and outside of the recognition helix make sequence-nonspecific contacts with the phosphate backbone both in and upstream of the -35 promoter element [19]. σ_4/-35 element interactions are established during promoter recognition and persist through the isomerization(s)

leading to RP$_O$ formation and into the first stages of transcript synthesis (abortive initiation), until RNAP breaks all contacts with the promoter and escapes into productive elongation [99, 128]. Many transcription activators that target σ make protein/protein contacts with σ$_4$ that stabilize the binding of σ$_4$ to the promoter -35 element (see **The regulation of transcription initiation**).

Promoter contacts of σ$_2$: the -10 element

The interaction between σ$_2$ and the -10 element is more complex, since σ$_2$ makes sequence-specific contacts with single-stranded DNA and is directly involved in promoter melting. Promoter melting begins at position -11 of the -10 element and proceeds to position ~+3 in the final RP$_O$, exposing the transcription start site. Various lines of evidence support a model for strand separation in which the initiating event is the extrusion of the consensus -11A base from the double helix and its capture within a hydrophobic pocket in σ$_2$ [39, 68]. The interactions of the flipped -11A base with the σ$_2$ residues that constitute the pocket (F419, R423, F425, and Y430, *Eco* numbers based on the *Taq* σ$^A_{2-3}$/ssDNA structure) stabilize the melted state and enable the interactions of σ$_2$ with the DNA backbone over positions -10 to -8 that drive the downstream unwinding of the DNA helix [39]. The highly conserved -7T base at the downstream end of the -10 element is flipped into a hydrophilic pocket comprised of residues from σ conserved regions 1.2, 2.1, and 2.3 [39]. Recognition of the -10 element by σ is facilitated by its interaction with the β'cc, which is thought to function as an allosteric effector in the context of the holoenzyme to appropriately position σ$_2$ to engage the -10 element [142]. Substitutions in either σ$^{70}_2$ or the β'cc that weaken this interaction significantly decrease nontemplate strand binding, promoter melting, and open complex stability [142, 143].

30

σ_2 and σ_4 mediate the major contacts between σ and the promoter in the context of the holoenzyme, but many σ factors also engage in direct contacts with auxiliary promoter elements. $\sigma^{70}_{1.2}$ makes sequence-specific contacts with DNA in the discriminator region, specifically with the base on the nontemplate strand two positions downstream of the -10 promoter element [61]. This interaction plays an important role in RP_O stability and is believed to occur with single-stranded nontemplate DNA in the context of the open complex to prevent strand collapse. The $\sigma^{70}_{1.2}$/discriminator interaction is strongest when there is a G at the corresponding nontemplate position, resulting in a stable RP_O, and weakest when the base is a C, as in the *rrn* promoters [61]. $\sigma^{70}_{1.2}$ consists of two α helices oriented ~90° with respect to each other and connected by a short linker [106, 138]; the first helix (residues 93-108 of *Eco* $\sigma^{70}_{1.2}$) is positioned close to the discriminator in structural models of RP_O, and substitutions at two residues in this helix (Y101 and R103) significantly decrease RP_O stability [61]. *Taq* $\sigma^A_{1.2}$ makes a similar contact with a GGGA motif immediately downstream of the -10 element at *Taq* promoters [37]. Some promoters recognized by the primary σ factor have the extended -10 architecture, TGnTATAAT. At these promoters, residues in $\sigma_{3.0}$ interact with the TGn motif; specifically, genetic analyses suggest that the residues corresponding to *Eco* σ^{70} E458 and H455 in region 3.0 contact the G:C bp at position -14 in the TGn motif [4, 10, 127].

The α subunit of RNAP

As described during the discussion on UP elements, the α subunit of RNAP (product of the *rpoA* gene) can also play a direct—and crucial at some promoters—role in transcription initiation. It consists of two independently folded domains, the N-terminal domain (αNTD) and C-terminal domain (αCTD), connected by a flexible linker that can stretch up to 44 Å and allow the αCTD to assume a variety of positions relative to the αNTD [6]. The αNTD contains the

determinants for dimerization and for the interaction with other core RNAP subunits [83] and provides a platform for the assembly of the active core enzyme as follows: $2\alpha \rightarrow \alpha_2 \rightarrow \alpha_2\beta \rightarrow \alpha_2\beta\beta' \rightarrow \alpha_2\beta\beta'\omega$ [75]. The αCTD is a DNA-binding domain that participates in promoter recognition by making direct sequence-specific contacts with UP elements [123] and can engage in sequence-nonspecific interactions with the phosphate backbone in upstream DNA [124]. Depending on the nature of the UP element (a full UP element with two subsites, a proximal site only, or a distal site only), either one or both of the αCTDs makes sequence-specific contacts with DNA [36]. Genetic screens for α mutants deficient in UP element recognition followed by biochemical characterization identified the DNA-binding patch of the αCTD, which consists of seven residues: L262, R265 (most critical), N268, C269, G296, K298, and S299 (*Eco* numbers) [45, 102]. These residues are nearly universally conserved in bacteria, suggesting that UP elements with sequences similar to the *E. coli* consensus are widespread in the bacterial kingdom and important in gene expression. At promoters that do not depend on UP elements or activators, the αCTDs are dispensable for transcription: α subunits lacking the CTD (αΔCTD) can assemble into enzymatically-active RNAP that is competent for transcription. This dispensability is the basis for the *in vivo* bacterial two-hybrid system to study protein-protein interactions [30, 32] and one-hybrid system to study σ_4/DNA interactions [31, 108, 109].

In addition to protein/DNA contacts, the αCTD can engage in several protein/protein contacts that influence transcription initiation. First, the αCTD bound at the proximal UP element subsite can interact with σ_4 bound at the -35 promoter element to stabilize the holoenzyme/ promoter complex and facilitate transcription initiation by stabilizing RP_C or an intermediate on the pathway to RP_O [23, 125]. Second, many transcription factors with binding sites upstream of (and that do not overlap with) the core promoter interact with the αCTD to effectively recruit holoenzyme to the promoter and stabilize RP_C, activating transcription [reviewed in 14, 70, 126]. A classic example of this αCTD-mediated transcription activation is the cyclic-AMP receptor

32

protein (CRP) working at the *lac* promoter (P*lac*): CRP binds at a specific site centered at -61.5

and stabilizes nonspecific αCTD/DNA contacts in the region between the CRP binding site and

the promoter -35 element [14] (see **Figure 1.5E**). The αCTD is also the target of at least one

non-classical regulator that does not bind DNA: immediately after infecting *E. coli*,

bacteriophage T4 ADP-ribosylates the *E. coli* RNAP αCTDs at residue R265 (the most critical

DNA-binding residue), specifically inhibiting transcription from UP element-dependent promoters

[reviewed in 34]. This is an attractive propagative strategy for the phage, since the primary UP

element-dependent promoters in *E. coli* are the *rrn* promoters that account for ~70% of all

transcription during rapid growth. Selectively inhibiting the utilization of these host promoters

frees up a significant amount of RNAP that can be redirected to phage promoters. A variation of

this UP element/αCTD-mediated regulatory paradigm is investigated for the *S. aureus*

bacteriophage G1-encoded regulator, gene product 67 (gp67) in **Chapters 2** and **3** of this

dissertation.

Regulation of transcription initiation

Transcription initiation is a key regulated step in the coordinated expression of genes in

response to environmental conditions. The rate of initiation depends primarily on intrinsic

promoter strength, the interaction of classical activators and repressors with specific DNA

binding sites to recruit or obstruct the relevant holoenzyme, the action of small ligands or protein

effectors that alter the kinetics of the various steps in initiation, and the interaction of particular σ

factors with core RNAP to form specific holoenzyme species. Each step on the pathway from

free RNAP and promoter DNA to the final transcriptionally competent open complex presents an

opportunity for regulation. The central importance of σ factors in holoenzyme formation and

transcription initiation make them logical, and crucial, regulatory targets.

Activators and repressors that target σ

A single core RNAP species transcribes all genes in bacteria, so the global pattern of gene expression under any condition is largely determined by the identity and activity of the predominant transcribing holoenzyme species. The activity of a particular holoenzyme species can be modulated by the action of classical activators and repressors that bind to specific recognition sites upstream of or overlapping the core promoter. Repressors antagonize holoenzyme function and result in reduced transcription from target promoters; activators augment holoenzyme activity at target promoters and increase transcription, generally through direct protein/protein contacts with RNAP [reviewed in 12]. Activators that function through an interaction with the αCTD have been discussed previously, but at least 10 activators have been identified that target σ^{70}_4 and promote the initial binding of holoenzyme to duplex promoter DNA (increase K_B) and/or the isomerization(s) to RP_O (increase k_i) [reviewed in 29, 57]. Activators that target σ_4 generally have binding sites immediately adjacent to or partially overlapping the -35 promoter element. A classical example of this class of regulator is the bacteriophage λ CI protein (λCI). λCI stimulates transcription of its own gene from the P_{RM} promoter by binding to a site centered at -42 and interacting with σ^{70}_4 bound at the -35 element to promote RP_O formation [reviewed in 70]. Transcription factors are synthesized in response to specific environmental or developmental cues that signal the need for the expression of particular genes and ensure that the appropriate holoenzyme preferentially initiates transcription at these promoters and/or is actively occluded from the promoters of genes that encode unneeded products.

Modulating σ factor competition for core RNAP

The identity of the predominant holoenzyme species is determined by competition between different σ factors for a limiting amount of core RNAP [56, 79]. The levels, activity, and affinity for core RNAP of the individual σ factors influence this competition. Core RNAP binding

affinity is an inherent property of each σ factor: the primary σ factor generally has the highest affinity, and the alternative σ factors can vary dramatically in their (usually weaker) binding affinities [95]. σ factor levels are determined by the interplay of *de novo* synthesis regulated at the transcriptional or—less frequently—translational level and degradation by intracellular proteolysis. For example, the *E. coli* alternative σ factor σ^{38}, which directs gene expression during stationary phase and in response to various stresses, is transcribed at only basal levels during exponential growth. Its expression is significantly upregulated during the transition to stationary phase or under general stress, when $E•\sigma^{38}$ becomes the predominant transcribing holoenzyme [77]. The RssB response regulator then rapidly targets σ^{38} for degradation by the ClpXP protease upon outgrowth or removal of the stress so $E•\sigma^{70}$ again predominates and housekeeping genes are preferentially transcribed [151].

De novo synthesis followed by rapid turnover in response to environmental or developmental cues is an energetically expensive strategy to regulate the abundance of an active σ factor and its cognate holoenzyme. An alternative approach is to reversibly control σ factor activity post-translationally using a class of regulators called 'anti-σ factors' that are σ factor-specific. This strategy would allow the cell to maintain pools of multiple σ factors at all times without compromising the activity of the current predominant holoenzyme species, and to enact rapid shifts in the transcription program without the energetic and time costs of *de novo* synthesis. Bacteriophages, which often depend on the host's transcription machinery for expression of their own genes, can also exploit this strategy: by regulating host σ factor activity, they can redirect host holoenzyme(s) to phage promoters and/or promote the formation of holoenzyme(s) containing phage-encoded σ factors.

Anti-σ factors are defined as transcription regulators that function by specifically (and directly) binding to their cognate σ factor, resulting in the specific inhibition of transcription from promoters dependent on that σ factor for recognition [20, 65, 69, 73]. Initially, anti-σ factors were

believed to function by simple sequestration: binding of the anti-σ factor to the σ factor prevents the association of σ with core RNAP by occluding core-binding determinants or inducing conformational changes in σ incompatible with holoenzyme formation. While many of the characterized anti-σ factors do function primarily by this mechanism, some anti-σ factors have been shown to have effects in the context of the holoenzyme. Many bacterially-encoded anti-σ factors target alternative σ factors. A classical example is FlgM, a regulator of σ^{28}-dependent late (class III) flagellar genes in Gram-negative and Gram-positive bacteria. Early (class I) and middle (class II) flagellar genes are transcribed by E•σ^{70}, so early in the cascade FlgM binds to and sequesters σ^{28} (the product of a class II gene) in the cytoplasm and prevents its association with core RNAP by masking binding determinants on σ^{28}_2 and σ^{28}_4; upon completion of the hook basal body (HBB) structure, FlgM is secreted by the HBB and σ^{28} is freed and associates with core RNAP (forming E•σ^{28}) to express the late flagellar genes [reviewed in 20, 65, 73]. Cytoplasmic FlgM also inhibits σ^{28}-dependent transcription by forming a transient ternary complex with E•σ^{28} and promoting its dissociation [21, 22].

Alternative σ factors have lower binding affinities for core RNAP and are generally less abundant than the primary σ factor, whose levels remain fairly constant throughout the growth cycle (at least in *E. coli* [56]); thus, subverting the primary σ may favor the formation of alternative holoenzymes. A primary σ-specific anti-σ factor would reversibly inactivate at least a portion of the pool of the primary σ under conditions where expression of other regulons is needed. One such bacterially-encoded anti-σ has been identified in *E. coli*: the regulator of sigma D (Rsd) protein, a stationary phase-specific anti-σ^{70} [78]. Rsd forms a stoichiometric complex with *Eco* σ^{70} in stationary phase by simultaneously contacting σ^{70}_2 and σ^{70}_4 at residues normally involved in binding the β'cc and β-flap for holoenzyme assembly [146].

Anti-σ factors that target the primary σ have also been identified among bacteriophage-encoded regulators of transcription. Most bacteriophages do not encode their own RNAP and

must co-opt the host's transcription machinery. By encoding anti-σ factors specific to the host's

primary σ factor, bacteriophages can shift the transcription program in the host to favor

expression of phage genes upon infection. The bacteriophage T4-encoded anti-σ^{70} factor AsiA

is one well-characterized example. However, unlike Rsd, it interacts with *Eco* σ^{70}_4 in the context

of holoenzyme [129], occluding σ^{70}_4 determinants necessary for the σ^{70}_4/β-flap interaction [55,

132] and stabilizing a σ^{70}_4 conformation incompatible with -35 element binding [87]. In this

context, it inhibits E•σ^{70}-dependent transcription from -10/-35 promoters in the phage genome

(early genes) and host genome, but has no effect on promoters lacking a canonical -35 element

(e.g. extended -10 promoters) [129, 148]. AsiA has a second, essential, role as a co-activator of

T4 middle gene transcription in conjunction with the phage-encoded activator MotA. Middle

promoters lack a -35 consensus sequence and instead have strong -10 or extended -10

elements and a binding site for MotA centered at -30 (the 'Mot box'). As a stable component of

E•σ^{70}, AsiA interacts with MotA bound at the Mot box to redirect the AsiA/E•σ^{70} complex to

middle promoters [147]. AsiA also directly contacts the β-flap in the context of E•σ^{70}, replacing

σ^{70}_4/β-flap contacts [147, 148]. This may stabilize the AsiA/E•σ^{70} ternary complex and contribute

to both inhibition of transcription from -10/-35 promoters and co-activation of phage middle

promoters. The multifunctional operation of an anti-σ factor as a stable component of the

bacterial holoenzyme may prove to be a general regulatory paradigm and is investigated with

the bacteriophage G1-encoded anti-*Sau* σ^A gp67 in **Chapters 2** and **3**.

Bacteriophages and *S. aureus* phage G1 gp67

Bacteriophages (phages)—viruses that specifically infect bacteria—are the most

abundant and diverse entities in the biosphere, with an estimated global population of 10^{30} [1].

Phages are widely distributed in every ecosystem where potential host bacteria can be found

and have important roles in bacterial ecology and evolution and in the virulence of bacterial

pathogens. Phage classification consists of one order (*Caudovirales*, the tailed phages) comprised of three large families (*Myoviridae*, *Siphoviridae*, and *Podoviridae*), and ten smaller families, with 31 recognized genera. Families are chiefly defined by the nature of the page nucleic acid—single- or double-stranded DNA or single- or double-stranded RNA—and overall virion morphology. Tailed phages constitute the largest and most diverse phage group, with >96% of known phages belonging to this group—including some of the most studied: the coliphages T4 (*Myoviridae*), λ (*Siphoviridae*), and T7 (*Podoviridae*) [1]. These phages have linear dsDNA genomes, with genes for related functions clustering together in modules that are expressed sequentially during infection (early, middle, and late genes). Tailed phages can be virulent or temperate. In the case of virulent phages, the developmental program proceeds quickly, culminating with lysis of the host cell and the release of mature phage virion progeny, which initiate new rounds of infection. Temperate phages can either proceed with the lytic cycle upon infection or enter into a latent state (lysogeny) that can be stably maintained for many cell generations; however, in response to particular environmental cues, the quiescent phage (known as a prophage) can be activated such that lytic development ensues [1].

Like eukaryotic viruses, phages are obligate intracellular parasites and are dependent on the host's transcription and/or translation machinery for propagation. Most of the regulation of phage development occurs at the level of transcription, and phage regulatory mechanisms that act at any of the stages of transcription have been found. Detailed mechanistic studies of these phage processes have greatly facilitated our understanding of bacterial RNAP and bacterial transcription and its regulation. A number of phages (e.g. coliphage T7) encode their own single-subunit RNAPs, which are structurally different from bacterial (and eukaryotic) multisubunit RNAPs, but most (if not all) must make use of the host transcription machinery at some point(s) during infection. As such, phages have developed a wide array of mechanisms to modify and appropriate host RNAP to serve viral needs, typically through direct covalent

modifications of RNAP or through specific RNAP-binding proteins [reviewed in 107]. A well-characterized example of this versatile subversion of host transcription is presented by the large, virulent coliphage T4, which has been discussed briefly in other sections of this chapter. Shortly after infection, two T4 enzymes ADP-ribosylate residue R265 of the αCTDs; this inhibits transcription from UP element-dependent promoters in the host (e.g. *rrn*), which frees up E•σ^{70} to transcribe early phage genes [reviewed in 107]. The anti-σ^{70} factor AsiA (as discussed above), an early gene product, then associates with and remodels σ^{70}_4 in the context of E•σ^{70}, inhibiting all host transcription from -10/-35 promoters and redirecting AsiA-modified E•σ^{70} to phage middle promoters in conjunction with the activator MotA (another early gene product). The middle genes include gene 55, which encodes a highly divergent σ factor (gp55) needed for the recognition of T4 late promoters. AsiA binding to σ^{70} likely destabilizes σ^{70}/core RNAP interactions and thus may enable gp55 to effectively compete to form E•gp55 and direct the expression of genes involved in virion maturation and host cell lysis [reviewed in 107].

S. aureus and other pathogenic bacteria are an important cause of severe infection and mortality, and there is a troubling rise in widespread antibiotic resistance in these pathogens [111]. This has led to significant effort being put into the development of novel antimicrobials that target essential bacterial cellular processes. With next-generation sequencing, >400 dsDNA phage genomes have been sequenced, and metagenomic analyses suggest that only a small percentage of the global phage metagenome has been sampled and that many phage open reading frames (ORFs) have not been functionally characterized [86]. This untapped proteomic diversity, along with the narrow host range of most phages, has catalyzed the resurgence of phages as potential therapeutics [134] and sources of validated drug discovery targets. The idea is to mine the genomes of phages of pathogenic bacteria to identify phage proteins that target essential bacterial processes—like transcription and RNAP—and, that by understanding the mechanism of action of these phage-encoded inhibitors, small molecules that mimic the

behavior of the phage can be developed. To this end, Liu *et al.* [91] and Kwan *et al.* [86] sequenced the genomes of 27 phages of *S. aureus* and conducted a functional genomics screen to systematically identify phage-encoded inhibitors of *S. aureus* growth. One such growth-inhibitory protein, **gp67** (encoded by *orf67*) was identified in the virulent phage G1, and is the subject of this dissertation.

Phage G1 is a member of the *Myoviridae* family and of the genus Twort-like, a group of twelve *S. aureus* polyvalent phages with large (≥130 kilobases) genomes that are obligatorily lytic and infect most clinical isolates of staphylococci [92]. Phages of this genus have very high DNA homologies (88-99%) and encode many of the same proteins, including homologues of G1 gp67; phage Twort is the exception, with only ~50% genomic and protein identity to the other Twort-like phages, but it encodes an obvious gp67 homologue also present in *Listeria* and *E. faecalis* phages [92, 112]. Twort-like phages do not encode their own RNAP and thus must utilize the host holoenzyme (*Sau* E•σ^A) to express at least early and middle phage genes, and analysis of the phage G1 genome sequence revealed many putative strong -10/-35 promoters likely recognized by E•σ^A, including the promoter for *orf67* (J. Osmundson, pers. comm; [92]). The genomes of Twort-like phages also encode a putative σ factor (gp56 in phage G1, a 220 residue protein) [92, 137], which suggests a paradigm similar to that of the related coliphage T4: Twort-like phages appropriate *Sau* E•σ^A to shut down host transcription and drive early/middle phage gene expression, which then facilitates the formation of holoenzyme containing a phage-encoded σ factor to express late genes and complete the phage life cycle. (The putative σ factor has not been isolated or tested for transcription *in vitro*, however). Gp67 is an attractive candidate to mediate this early appropriation and shutdown of host transcription, since its production significantly decreased RNA production in *Sau* [91], and it was subsequently shown to specifically bind *Sau* σ^A_4 in the context of E•σ^A and inhibit σ^A-dependent transcription [26].

Osmundson *et al.* ([112]; see **Appendix 2**) used biochemical and crystallographic

techniques to investigate the mechanism of action of the 23 kDa anti-σ^A factor gp67. They

demonstrated the inability of gp67 to form a ternary complex with a hybrid holoenzyme

consisting of *E. coli* core RNAP and *Sau* σ^A and therefore sought to analyze the mechanism of

gp67 action in a native *S. aureus* system. In particular, they used RNA-seq to examine the

genome-wide effects of gp67 production on transcription *in vivo*, and they developed a *S.*

aureus-based *in vitro* transcription system to study the effects of gp67 on specific promoters.

The data suggest that, in *S. aureus*, gp67 is a promoter-specific transcription inhibitor that

functions by becoming a stable part of E•σ^A through a strong interaction with σ^A_4 and exerts its

effects not through σ^A, but by interfering with the binding of the αCTDs to UP elements. In this

manner, gp67 can form a stable ternary complex with E•σ^A at all σ^A-dependent promoters but

inhibits transcription only from promoters that depend on the UP element for efficient initiation

(e.g. rRNA promoters). Consistent with this picture, the high-resolution X-ray crystal structure of

the *Sau* σ^A_4/gp67 complex reveals that gp67 binding does not induce conformational changes in

σ^A_4 or occlude the core- or DNA-binding binding determinants of σ^A_4, as all previously

characterized anti-σ factors do.

Summary

Multisubunit RNAPs display striking sequence and structural conservation across all

domains of life, suggesting that detailed mechanistic studies using experimentally tractable

bacterial model systems can provide global insight into gene expression. In bacteria,

transcription initiation—the first step in gene expression—is the gateway to a complex,

multilayered regulatory network that ensures the coordinated expression of specific sets of

genes under particular conditions and enacts rapid changes in the cell's transcriptional profile in

response to environmental and developmental cues. σ factors, which make extensive contacts

with the core RNAP enzyme to form the initiation-competent holoenzyme, are intimately

involved in every step of initiation, from holoenzyme formation to sequence-specific promoter

recognition to open complex formation, and thus make excellent versatile targets for regulation.

Anti-σ factors are transcription regulators that interact with their cognate σ factor and influence

RNAP activity. Prior to this work, only two anti-σ factors specific for a primary σ had been

identified: the coliphage T4-encoded AsiA protein and the *E. coli*-encoded stationary phase-

specific Rsd protein. These factors both target *Eco* σ^{70} but they inhibit transcription by different

mechanisms, suggesting that a broader range of primary σ-specific anti-σ factors may exist.

Virulent bacteriophages are logical sources for novel anti-σ factors and antimicrobials, since

many phages have evolved varied and effective mechanisms for subverting essential cellular

processes (transcription chief among them) in the host to ensure viral propagation. The large

number of sequenced phage genomes and limited functional characterization of phage ORFs

present a largely unexplored proteomic landscape potentially rife with transcription regulators

that can contribute to our enhanced understanding of transcription and its regulation, and may

provide prophylactic or therapeutic alternatives to antibiotics in combating infectious disease.

Gp67 is a relatively uncharacterized phage G1-encoded inhibitor of *S. aureus* growth

that targets the primary σ factor *(Sau* σ^A) and inhibits σ^A-dependent transcription but lacks

structural or sequence similarity to other anti-σ factors. The work presented in this dissertation

to ascertain the mechanism of action of gp67 using *E. coli* tools was undertaken in conjunction

with Joseph Osmundson and Seth Darst at The Rockefeller University upon the identification of

gp67 [26, 91] and the determination of a preliminary *Sau* σ^A_4/gp67 X-ray co-crystal structure in

the Darst laboratory (2008) to circumvent some of the technical challenges inherent in working

with *S. aureus*. **Chapter 2** presents the detailed genetic analyses conducted to thoroughly map

the *Sau* σ^A_4/gp67 interface and identify the key determinants responsible for the specificity of

gp67 for *Sau* σ^A, which provided *in vivo* genetic validation of the *Sau* σ^A_4/gp67 co-crystal

structure and generated a crucial control protein for the *in vivo* and *in vitro* experiments conducted by Osmundson *et al.* using the *S. aureus* system ([112]; **Appendix 2**). The experiments presented in **Chapter 2** also enabled the development of an *E. coli*-based system in which gp67 is active (discussed in Chapter 3); in addition, they made it possible to interrogate the specificity of gp67 binding in the context of the primary σ factors of other bacteria, with the aim of developing an alternative system to study gp67 using the low G+C Gram-positive model bacterium *B. subtilis* (discussed in Chapter 4).

Chapter 3 details the functional *in vivo* and *in vitro* analyses undertaken in *E. coli* to dissect the mechanism of gp67-mediated transcription inhibition using a system developed with the genetic data from Chapter 2. The apparent context dependence of gp67 as a transcription inhibitor in a heterologous system is investigated, and a structure-based model for gp67-mediated inhibition of UP element utilization ([112]; **Appendix 2**) is presented and directly tested using *E. coli* tools. Finally, **Chapter 4** summarizes the work conducted during the dissertation research to enhance our mechanistic understanding of transcription regulation by the anti-σ factor G1 gp67 and outlines further experiments to be conducted but that lie outside the scope of this dissertation. The development of a gp67-responsive *B. subtilis* system is discussed as an alternative to the *S. aureus* system that is more experimentally tractable but also less heterologous than the *E. coli*-based system of Chapter 3. Further experiments to interrogate the apparent context dependence of gp67-mediated transcription inhibition in the heterologous *E. coli* system are also presented.

References

1.   *Bacteriophages: Biology and Applications*, ed. Kutter, E. and Sulakvelidze, A. 2005, Boca Raton: CRC Press.

2.   **Arthur, T. M. and Burgess, R. R.** 1998. Localization of a sigma70 binding site on the N terminus of the *Escherichia coli* RNA polymerase beta' subunit. J Biol Chem **273**(47): 31381-7.

3.   **Barker, M. M*., et al.*** 2001. Mechanism of regulation of transcription initiation by ppGpp. I. Effects of ppGpp on transcription initiation in vivo and in vitro. J Mol Biol **305**(4): 673-88.

4.   **Barne, K. A*., et al.*** 1997. Region 2.5 of the *Escherichia coli* RNA polymerase sigma 70 subunit is responsible for the recognition of the'extended– 10' motif at promoters. EMBO J **16**(13): 4034-40.

5.   **Bentley, S. D*., et al.*** 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature **417**: 141-7.

6.   **Blatter, E. E*., et al.*** 1994. Domain organization of RNA polymerase alpha subunit: C-terminal 85 amino acids constitute a domain capable of dimerization and DNA binding. Cell **78**(5): 889-96.

7.   **Borukhov, S., Lee, J., and Laptenko, O.** 2005. Bacterial transcription elongation factors: new insights into molecular mechanism of action. Mol Microbiol **55**: 1315-24.

8.   **Borukhov, S. and Nudler, E.** 2003. RNA polymerase holoenzyme: structure, function and biological implications. Curr Opin Microbiol **6**(2): 93-100.

9.   **Bown, J. A*., et al.*, *Extended -10 promoters*, in *Mechanisms of Transcription - Nucleic Acids and Molecular Biology Vol. 11*, Eckstein, F. and Lilley, D. M. J., Editors. 1997, Springer. p. 41-52.

10.   **Bown, J. A*., et al.*** 1999. Organization of open complexes at *Escherichia coli* promoters. Location of promoter DNA sites close to region 2.5 of the sigma70 subunit of RNA polymerase. J Biol Chem **274**(4): 2263-70.

11.   **Bremer, H.** and **Dennis, P. P.**, *Modulation of chemical composition and other parameters of the cell by growth rate*, in *Escherichia coli and Salmonella: Cellular and*

Molecular Biology (2nd edn), Neidhardt, F. C., et al., Editors. 1996, American Society for Microbiology: Washington, DC.

12. **Browning, D. F. and Busby, S. J.** 2004. The regulation of bacterial transcription initiation. Nat Rev Microbiol **2**(1): 57-65.

13. **Burgess, R. R., Arthur, T. M., and Pietz, B. C.** 1998. Interaction of *Escherichia coli* sigma 70 with core RNA polymerase. Cold Spring Harb Symp Quant Biol **63**: 277-87.

14. **Busby, S. and Ebright, R. H.** 1994. Promoter structure, promoter recognition, and transcription activation in prokaryotes. Cell **79**(5): 743-6.

15. **Busby, S. J.** 2009. More pieces in the promoter jigsaw: recognition of -10 regions by alternative sigma factors. Mol Microbiol **72**(4): 809-11.

16. **Callaci, S., Heyduk, E., and Heyduk, T.** 1998. Conformational changes of *Escherichia coli* RNA polymerase sigma70 factor induced by binding to the core enzyme. J Biol Chem **273**(49): 32995-3001.

17. **Callaci, S., Heyduk, E., and Heyduk, T.** 1999. Core RNA polymerase from *E. coli* induces a major change in the domain arrangement of the sigma 70 subunit. Mol Cell **3**: 229-38.

18. **Camacho, A. and Salas, M.** 1999. Effect of mutations in the "extended -10" motif of three *Bacillus subtilis* sigmaA-RNA polymerase-dependent promoters. J Mol Biol **286**(3): 683-93.

19. **Campbell, E. A.*, et al.*** 2002. Structure of the bacterial RNA polymerase promoter specificity sigma subunit. Mol Cell **9**(3): 527-39.

20. **Campbell, E. A., Westblade, L. F., and Darst, S. A.** 2008. Regulation of bacterial RNA polymerase sigma factor activity: a structural perspective. Curr Opin Microbiol **11**(2): 121-7.

21. **Chadsey, M. S. and Hughes, K. T.** 2001. A multipartite interaction between *Salmonella* transcription factor sigma28 and its anti-sigma factor FlgM: implications for sigma28 holoenzyme destabilization through stepwise binding. J Mol Biol **306**(5): 915-29.

22. **Chadsey, M. S., Karlinsey, J. E., and Hughes, K. T.** 1998. The flagellar anti-sigma factor FlgM actively dissociates *Salmonella typhimurium* sigma28 RNA polymerase holoenzyme. Genes Dev **12**(19): 3123-36.

23. **Chen, H., Tang, H., and Ebright, R. H.** 2003. Functional interaction between RNA polymerase alpha subunit C-terminal domain and sigma70 in UP-element- and activator-dependent transcription. Mol Cell **11**(6): 1621-33.

24. **Cramer, P.** 2002. Multisubunit RNA polymerases. Curr Opin Struct Biol **12**(1): 89-97.

25. **Cramer, P.*, et al.*** 2000. Architecture of RNA polymerase II and implications for the transcription mechanism. Science **288**: 640-9.

26. **Dehbi, M.*, et al.*** 2009. Inhibition of transcription in *Staphylococcus aureus* by a primary sigma factor-binding polypeptide from phage G1. J Bacteriol **191**(12): 3763-71.

27. **Dombroski, A. J., Walter, W. A., and Gross, C. A.** 1993. Amino-terminal amino acids modulate sigma-factor DNA-binding activity. Genes Dev **7**: 2446-55.

28. **Dombroski, A. J.*, et al.*** 1992. Polypeptides containing highly conserved regions of transcription initiation factor sigma 70 exhibit specificity of binding to promoter DNA. Cell **70**(3): 501-12.

29. **Dove, S. L., Darst, S. A., and Hochschild, A.** 2003. Region 4 of sigma as a target for transcription regulation. Mol Microbiol **48**(4): 863-74.

30. **Dove, S. L. and Hochschild, A.** 2004. A bacterial two-hybrid system based on transcription activation. Methods Mol Biol **261**: 231-46.

31. **Dove, S. L., Huang, F. W., and Hochschild, A.** 2000. Mechanism for a transcriptional activator that works at the isomerization step. Proc Natl Acad Sci U S A **97**(24): 13215-20.

32. **Dove, S. L., Joung, J. K., and Hochschild, A.** 1997. Activation of prokaryotic transcription through arbitrary protein-protein contacts. Nature **386**: 627-30.

33. **Ebright, R. H.** 2000. RNA polymerase: Structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. J Mol Biol **304**: 687-98.

34. **Ebright, R. H. and Busby, S.** 1995. The *Escherichia coli* RNA polymerase alpha subunit: structure and function. Curr Opin Genet Dev **5**(2): 197-203.

35. **Estrem, S. T.*, et al.*** 1998. Identification of an UP element consensus sequence for bacterial promoters. Proc Natl Acad Sci U S A **95**(17): 9761-6.

36. **Estrem, S. T.*, et al.*** 1999. Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase alpha subunit. Genes Dev **13**(16): 2134-47.

37. **Feklistov, A.*, et al.*** 2006. A basal promoter element recognized by free RNA polymerase sigma subunit determines promoter recognition by RNA polymerase holoenzyme. Mol Cell **23**(1): 97-107.

38. **Feklistov, A. and Darst, S. A.** 2009. Promoter recognition by bacterial alternative sigma factors: the price of high selectivity? Genes Dev **23**(20): 2371-5.

39. **Feklistov, A. and Darst, S. A.** 2011. Structural basis for promoter-10 element recognition by the bacterial RNA polymerase sigma subunit. Cell **147**(6): 1257-69.

40. **Fisher, R. and Blumenthal, T.** 1980. Analysis of RNA polymerase by trypsin cleavage. Evidence for a specific association between subunits sigma and beta involved in the closed to open complex transition. J Biol Chem **255**(22): 11056-62.

41. **Fraser, C. M.*, et al.*** 1995. The minimal gene complement of *Mycoplasma genitalium*. Science **270**: 397-403.

42. **Fredrick, K.*, et al.*** 1995. Promoter architecture in the flagellar regulon of *Bacillus subtilis*: high-level expression of flagellin by the sigma D RNA polymerase requires an upstream promoter element. Proc Natl Acad Sci U S A **92**: 2582-6.

43. **Gaal, T.*, et al.*** 1989. Saturation mutagenesis of an *Escherichia coli* rRNA promoter and initial characterization of promoter variants. J Bacteriol **171**(9): 4852-61.

44. **Gaal, T.*, et al.*** 1997. Transcription regulation by initiating NTP concentration: rRNA synthesis in bacteria. Science **278**: 2092-7.

45. **Gaal, T.*, et al.*** 1996. DNA-binding determinants of the alpha subunit of RNA polymerase: novel DNA-binding domain architecture. Genes Dev **10**(1): 16-26.

46.   **Gaal, T.**, **Ross, W.**, and **Gourse, R. L.**, *Ribosomal RNA promoter-RNA polymerase interactions and rRNA transcription in Escherichia coli*, in *Mechanisms of Transcription - Nucleic Acids and Molecular Biology Vol. 11*, Eckstein, F. and Lilley, D. M. J., Editors. 1997, Springer. p. 87-100.

47.   **Gardella, T., Moyle, H., and Susskind, M. M.** 1989. A mutant *Escherichia coli* sigma 70 subunit of RNA polymerase with altered promoter specificity. J Mol Biol **206**(4): 579-90.

48.   **Geszvain, K.*, et al.** 2004. A hydrophobic patch on the flap-tip helix of *E.coli* RNA polymerase mediates sigma 70 region 4 function. J Mol Biol **343**(3): 569-87.

49.   **Ghosh, T., Bose, D., and Zhang, X.** 2010. Mechanisms for activating bacterial RNA polymerase. FEMS Microbiol Rev **34**(5): 611-27.

50.   **Gill, S. C., Weitzel, S. E., and von Hippel, P. H.** 1991. *Escherichia coli* sigma 70 and NusA proteins. I. Binding interactions with core RNA polymerase in solution and within the transcription complex. J Mol Biol **220**(2): 307-24.

51.   **Gourse, R. L.** 1988. Visualization and quantitative analysis of complex formation between *E. coli* RNA polymerase and an rRNA promoter *in vitro*. Nucleic Acids Res **16**(20): 9789-809.

52.   **Gourse, R. L., de Boer, H. A., and Nomura, M.** 1986. DNA determinants of rRNA synthesis in *E. coli*: growth rate dependent regulation, feedback inhibition, upstream activation, antitermination. Cell **44**(1): 197-205.

53.   **Gourse, R. L.*, et al.** 1998. Strength and regulation without transcription factors: Lessons from bacterial rRNA promoters. Cold Spring Harb Symp Quant Biol **63**: 131-9.

54.   **Graves, M. C. and Rabinowitz, J. C.** 1986. *In vivo* and *in vitro* transcription of the *Clostridium pasteurianum* ferredoxin gene. Evidence for "extended" promoter elements in gram-positive organisms. J Biol Chem **261**(24): 11409-15.

55.   **Gregory, B. D.*, et al.** 2004. A regulator that inhibits transcription by targeting an intersubunit interaction of the RNA polymerase holoenzyme. Proc Natl Acad Sci U S A **101**(13): 4554-9.

56. **Grigorova, I. L*., et al.*** 2006. Insights into transcriptional regulation and sigma competition from an equilibrium model of RNA polymerase binding to DNA. Proc Natl Acad Sci U S A **103**(14): 5332-7.

57. **Gross, C. A*., et al.*** 1998. The functional and regulatory roles of sigma factors in transcription. Cold Spring Harb Symp Quant Biol **63**: 141-55.

58. **Gruber, T. M. and Gross, C. A.** 2003. Multiple sigma subunits and the partitioning of bacterial transcription space. Annu Rev Microbiol **57**: 441-66.

59. **Gruber, T. M*., et al.*** 2001. Binding of the initiation factor sigma 70 to core RNA polymerase is a multistep process. Mol Cell **8**(1): 21-31.

60. **Harley, C. B. and Reynolds, R. P.** 1987. Analysis of *E. coli* promoter sequences. Nucleic Acids Res **15**(5): 2343-61.

61. **Haugen, S. P*., et al.*** 2006. rRNA promoter regulation by nonoptimal binding of sigma region 1.2: an additional recognition element for RNA polymerase. Cell **125**(6): 1069-82.

62. **Haugen, S. P., Ross, W., and Gourse, R. L.** 2008. Advances in bacterial promoter recognition and its control by factors that do not bind DNA. Nat Rev Microbiol **6**(7): 507-19.

63. **Haugen, S. P*., et al.*** 2008. Fine structure of the promoter-sigma region 1.2 interaction. Proc Natl Acad Sci U S A **105**(9): 3292-7.

64. **Hawley, D. K. and McClure, W. R.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res **11**(8): 2237-55.

65. **Helmann, J. D.** 1999. Anti-sigma factors. Curr Opin Microbiol **2**(2): 135-41.

66. **Helmann, J. D.** 1995. Compilation and analysis of *Bacillus subtilis* sigma A-dependent promoter sequences: evidence for extended contact between RNA polymerase and upstream promoter DNA. Nucleic Acids Res **23**(13): 2351-60.

67. **Helmann, J. D. and Chamberlin, M. J.** 1988. Structure and function of bacterial sigma factors. Annu Rev Biochem **57**: 839-72.

68.   **Heyduk, E.*, et al.* 2006. A consensus adenine at position -11 of the nontemplate strand of bacterial promoter is important for nucleation of promoter melting. J Biol Chem **281**(18): 12362-9.

69.   **Hinton, D. M.*, et al.* 2005. Transcriptional takeover by sigma appropriation: remodelling of the sigma 70 subunit of *Escherichia coli* RNA plymerase by the bacteriophage T4 activator MotA and co-activator AsiA. Microbiology **151**: 1729-40.

70.   **Hochschild, A. and Dove, S. L.** 1998. Protein-protein contacts that activate and repress prokaryotic transcription. Cell **92**(5): 597-600.

71.   **Hook-Barnard, I., Johnson, X. B., and Hinton, D. M.** 2006. Escherichia coli RNA polymerase recognition of a sigma70-dependent promoter requiring a -35 DNA element and an extended -10 TGn motif. J Bacteriol **188**(24): 8352-9.

72.   **Hsu, L. M.** 2002. Promoter clearance and escape in prokaryotes. Biochim Biophys Acta **1577**: 191-207.

73.   **Hughes, K. T. and Mathee, K.** 1998. The anti-sigma factors. Annu Rev Microbiol **52**: 231-86.

74.   **Ishihama, A.** 1992. Role of the RNA polymerase alpha subunit in transcription activation. Mol Microbiol **6**(22): 3283-8.

75.   **Ishihama, A.** 1981. Subunit of assembly of *Escherichia coli* RNA polymerase. Adv Biophys **14**: 1-35.

76.   **Jain, D.*, et al.* 2004. Structure of a ternary transcription activation complex. Mol Cell **13**(1): 45-53.

77.   **Jishage, M. and Ishihama, A.** 1995. Regulation of RNA polymerase sigma subunit synthesis in *Escherichia coli*: intracellular levels of sigma 70 and sigma 38. J Bacteriol **177**(23): 6832-5.

78.   **Jishage, M. and Ishihama, A.** 1998. A stationary phase protein in *Escherichia coli* with binding activity to the major sigma subunit of RNA polymerase. Proc Natl Acad Sci U S A **95**(9): 4953-8.

79.   **Jishage, M.***, et al.* 1996. Regulation of RNA polymerase sigma subunit synthesis in *Escherichia coli*: intracellular levels of four species of sigma subunit under various growth conditions. J Bacteriol **178**(18): 5447-51.

80.   **Joo, D. M.***, et al.* 1998. Multiple regions on the *Escherichia coli* heat shock transcription factor sigma32 determine core RNA polymerase binding specificity. J Bacteriol **180**(5): 1095-102.

81.   **Kapanidis, A. N.***, et al.* 2006. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. Science **314**: 1144-7.

82.   **Keener, J.** and **Nomura, M.**, *Regulation of ribosome biosynthesis*, in *Escherichia coli and Salmonella: Cellular and Molecular Biology (2nd edn)*, Neidhardt, F. C., et al., Editors. 1996, American Society for Microbiology: Washington, DC.

83.   **Kimura, M., Fujita, N., and Ishihama, A.** 1994. Functional map of the alpha subunit of *Escherichia coli* RNA polymerase. Deletion analysis of the amino-terminal assembly domain. J Mol Biol **242**(2): 107-15.

84.   **Krasny, L. and Gourse, R. L.** 2004. An alternative strategy for bacterial ribosome synthesis: *Bacillus subtilis* rRNA transcription regulation. EMBO J **23**(22): 4473-83.

85.   **Kuznedelov, K.***, et al.* 2002. A role for interaction of the RNA polymerase flap domain with the sigma subunit in promoter recognition. Science **295**: 855-7.

86.   **Kwan, T.***, et al.* 2005. The complete genomes and proteomes of 27 *Staphylococcus aureus* bacteriophages. Proc Natl Acad Sci U S A **102**(14): 5174-9.

87.   **Lambert, L. J.***, et al.* 2004. T4 AsiA blocks DNA recognition by remodeling sigma70 region 4. EMBO J **23**(15): 2952-62.

88.   **Landick, R.** 2006. The regulatory roles and mechanism of transcriptional pausing. Biochem Soc Trans **34**: 1062-6.

89.   **Leirmo, S. and Gourse, R. L.** 1991. Factor-independent activation of *Escherichia coli* rRNA transcription. I. Kinetic analysis of the roles of the upstream activator region and supercoiling on transcription of the rrnB P1 promoter in vitro. J Mol Biol **220**(3): 555-68.

90.   **Lesley, S. A. and Burgess, R. R.** 1989. Characterization of the *Escherichia coli* transcription factor sigma 70: localization of a region involved in the interaction with core RNA polymerase. Biochemistry **28**(19): 7728-34.

91.   **Liu, J*., et al.*** 2004. Antimicrobial drug discovery through bacteriophage genomics. Nat Biotechnol **22**(2): 185-91.

92.   **Lobocka, M*., et al.*** 2012. Genomics of staphylococcal Twort-like phages--potential therapeutics of the post-antibiotic era. Adv Virus Res **83**: 143-216.

93.   **Lonetto, M., Gribskov, M., and Gross, C. A.** 1992. The sigma 70 family: sequence conservation and evolutionary relationships. J Bacteriol **174**(12): 3843-9.

94.   **Madan Babu, M. and Teichmann, S. A.** 2003. Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. Nucleic Acids Res **31**(4): 1234-44.

95.   **Maeda, H., Fujita, N., and Ishihama, A.** 2000. Competition among seven *Escherichia coli* sigma subunits: relative binding affinities to the core RNA polymerase. Nucleic Acids Res **28**(18): 3497-503.

96.   **Malhotra, A., Severinova, E., and Darst, S. A.** 1996. Crystal structure of a sigma 70 subunit fragment from *E. coli* RNA polymerase. Cell **87**(1): 127-36.

97.   **Marr, M. T. and Roberts, J. W.** 1997. Promoter recognition as measured by binding of polymerase to nontemplate strand oligonucleotide. Science **276**: 1258-60.

98.   **McClure, W. R.** 1985. Mechanism and control of transcription initiation in prokaryotes. Annu Rev Biochem **54**: 171-204.

99.   **Mecsas, J., Cowing, D. W., and Gross, C. A.** 1991. Development of RNA polymerase-promoter contacts during open complex formation. J Mol Biol **220**(3): 585-97.

100. **Mekler, V*., et al.*** 2002. Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. Cell **108**(5): 599-614.

101. **Mitchell, J. E*., et al.*** 2003. Identification and analysis of 'extended -10' promoters in *Escherichia coli*. Nucleic Acids Res **31**(16): 4689-95.

102. **Murakami, K., Fujita, N., and Ishihama, A.** 1996. Transcription factor recognition surface on the RNA polymerase alpha subunit is involved in contact with the DNA enhancer element. EMBO J **15**(16): 4358-67.

103. **Murakami, K. S.** 2013. X-ray crystal structure of Escherichia coli RNA polymerase sigma70 holoenzyme. J Biol Chem **288**(13): 9126-34.

104. **Murakami, K. S. and Darst, S. A.** 2003. Bacterial RNA polymerases: the wholo story. Curr Opin Struct Biol **13**: 31-9.

105. **Murakami, K. S.*, et al.*** 2002. Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. Science **296**: 1285-90.

106. **Murakami, K. S., Masuda, S., and Darst, S. A.** 2002. Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 A resolution. Science **296**: 1280-4.

107. **Nechaev, S. and Severinov, K.** 2003. Bacteriophage-induced modifications of host RNA polymerase. Annu Rev Microbiol **57**: 301-22.

108. **Nickels, B. E.** 2009. Genetic assays to define and characterize protein-protein interactions involved in gene regulation. Methods **47**: 53-62.

109. **Nickels, B. E.*, et al.*** 2002. Protein-protein and protein-DNA interactions of sigma70 region 4 involved in transcription activation by lambda cI. J Mol Biol **324**: 17-34.

110. **Nickels, B. E.*, et al.*** 2005. The interaction between sigma70 and the beta-flap of *Escherichia coli* RNA polymerase inhibits extension of nascent RNA during early elongation. Proc Natl Acad Sci U S A **102**(12): 4488-93.

111. **Nordmann, P.*, et al.*** 2007. Superbugs in the coming new decade; multidrug resistance and prospects for treatment of *Staphylococcus aureus, Enterococcus* spp. and *Pseudomonas aeruginosa* in 2010. Curr Opin Microbiol **10**(5): 436-40.

112. **Osmundson, J.*, et al.*** 2012. Promoter-specific transcription inhibition in *Staphylococcus aureus* by a phage protein. Cell **151**(5): 1005-16.

113. **Osterberg, S., del Peso-Santos, T., and Shingler, V.** 2011. Regulation of alternative sigma factor use. Annu Rev Microbiol **65**: 37-55.

114.   **Paget, M. S. and Helmann, J. D.** 2003. The sigma70 family of sigma factors. Genome Biol **4**(1): 203.

115.   **Panaghie, G.*, et al.*** 2000. Aromatic amino acids in region 2.3 of *Escherichia coli* sigma 70 participate collectively in the formation of an RNA polymerase-promoter open complex. J Mol Biol **299**(5): 1217-30.

116.   **Paul, B. J.*, et al.*** 2004. rRNA transcription in *Escherichia coli*. Annu Rev Genet **38**: 749-70.

117.   **Perez-Rueda, E. and Collado-Vides, J.** 2000. The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. Nucleic Acids Res **28**(8): 1838-47.

118.   **Rao, L.*, et al.*** 1994. Factor independent activation of *rrnB* P1. An "extended" promoter with an upstream element that dramatically increases promoter strength. J Mol Biol **235**(5): 1421-35.

119.   **Record, M. T., Jr.*, et al.***, *Escherichia coli RNA polymerase (Eσ70), promoters, and the kinetics of the steps of transcription initiation* in *Escherichia coli and Salmonella: Cellular and Molecular Biology (2nd edn)*, Neidhardt, F. C., et al., Editors. 1996, American Society for Microbiology: Washington, DC. p. 792-820.

120.   **Revyakin, A.*, et al.*** 2006. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. Science **314**: 1139-43.

121.   **Ring, B. Z., Yarnell, W. S., and Roberts, J. W.** 1996. Function of *E. coli* RNA polymerase sigma factor sigma 70 in promoter-proximal pausing. Cell **86**(3): 485-93.

122.   **Ross, W.*, et al.*** 1998. *Escherichia coli* promoters with UP elements of different strengths: modular structure of bacterial promoters. J Bacteriol **180**(20): 5375-83.

123.   **Ross, W.*, et al.*** 1993. A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. Science **262**: 1407-13.

124.   **Ross, W. and Gourse, R. L.** 2005. Sequence-independent upstream DNA-alphaCTD interactions strongly stimulate *Escherichia coli* RNA polymerase-*lacUV5* promoter association. Proc Natl Acad Sci U S A **102**(2): 291-6.

125. **Ross, W.*, et al.* 2003. An intersubunit contact stimulating transcription initiation by *E coli* RNA polymerase: interaction of the alpha C-terminal domain and sigma region 4. Genes Dev **17**(10): 1293-307.

126. **Russo, F. D. and Silhavy, T. J.** 1992. Alpha: the Cinderella subunit of RNA polymerase. J Biol Chem **267**(21): 14515-8.

127. **Sanderson, A.*, et al.* 2003. Substitutions in the *Escherichia coli* RNA polymerase sigma70 factor that affect recognition of extended -10 elements at promoters. FEBS Lett **544**(1-3): 199-205.

128. **Schickor, P.*, et al.* 1990. Topography of intermediates in transcription initiation of *E.coli*. EMBO J **9**(7): 2215-20.

129. **Severinova, E., Severinov, K., and Darst, S. A.** 1998. Inhibition of *Escherichia coli* RNA polymerase by bacteriophage T4 AsiA. J Mol Biol **279**(1): 9-18.

130. **Sharp, M. M.*, et al.* 1999. The interface of sigma with core RNA polymerase is extensive, conserved, and functionally specialized. Genes Dev **13**(22): 3015-26.

131. **Siegele, D. A.*, et al.* 1989. Altered promoter recognition by mutant forms of the sigma 70 subunit of *Escherichia coli* RNA polymerase. J Mol Biol **206**(4): 591-603.

132. **Simeonov, M. F.*, et al.* 2003. Characterization of the interactions between the bacteriophage T4 AsiA protein and RNA polymerase. Biochemistry **42**(25): 7717-26.

133. **Somerville, G. A. and Proctor, R. A.** 2009. At the crossroads of bacterial metabolism and virulence factor synthesis in Staphylococci. Microbiol Mol Biol Rev **73**: 233-48.

134. **Stone, R.** 2002. Bacteriophage therapy. Stalin's forgotten cure. Science **298**: 728-31.

135. **Thouvenot, B., Charpentier, B., and Branlant, C.** 2004. The strong efficiency of the *Escherichia coli gapA* P1 promoter depends on a complex combination of functional determinants. Biochem J **383**(Pt 2): 371-82.

136. **Travers, A. A.** 1980. Promoter sequence for stringent control of bacterial ribonucleic acid synthesis. J Bacteriol **141**(2): 973-6.

137.   **Vandersteegen, K.*, et al.*** 2011. Microbiological and molecular assessment of bacteriophage ISP for the control of *Staphylococcus aureus*. PLoS One **6**(9): e24418.

138.   **Vassylyev, D. G.*, et al.*** 2002. Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. Nature **417**: 712-9.

139.   **Voskuil, M. I. and Chambliss, G. H.** 1998. The -16 region of *Bacillus subtilis* and other gram-positive bacterial promoters. Nucleic Acids Res **26**(15): 3584-90.

140.   **Wang, Y. and deHaseth, P. L.** 2003. Sigma 32-dependent promoter activity in vivo: sequence determinants of the groE promoter. J Bacteriol **185**(19): 5800-6.

141.   **Wosten, M. M.** 1998. Eubacterial sigma-factors. FEMS Microbiol Rev **22**(3): 127-50.

142.   **Young, B. A.*, et al.*** 2001. A coiled-coil from the RNA polymerase beta' subunit allosterically induces selective nontemplate strand binding by sigma(70). Cell **105**(7): 935-44.

143.   **Young, B. A., Gruber, T. M., and Gross, C. A.** 2004. Minimal machinery of RNA polymerase holoenzyme sufficient for promoter melting. Science **303**: 1382-4.

144.   **Young, B. A., Gruber, T. M., and Gross, C. A.** 2002. Views of transcription initiation. Cell **109**(4): 417-20.

145.   **Yu, H. H.*, et al.*** 2006. Mutational analysis of the promoter recognized by *Chlamydia* and *Escherichia coli* sigma(28) RNA polymerase. J Bacteriol **188**(15): 5524-31.

146.   **Yuan, A. H.*, et al.*** 2008. Rsd family proteins make simultaneous interactions with regions 2 and 4 of the primary sigma factor. Mol Microbiol **70**(5): 1136-51.

147.   **Yuan, A. H. and Hochschild, A.** 2009. Direct activator/co-activator interaction is essential for bacteriophage T4 middle gene expression. Mol Microbiol **74**(4): 1018-30.

148.   **Yuan, A. H., Nickels, B. E., and Hochschild, A.** 2009. The bacteriophage T4 AsiA protein contacts the beta-flap domain of RNA polymerase. Proc Natl Acad Sci U S A **106**(16): 6597-602.

149.   **Zhang, G.*, et al.*** 1999. Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 A resolution. Cell **98**(6): 811-24.

150.   **Zhang, G. and Darst, S. A.** 1998. Structure of the *Escherichia coli* RNA polymerase alpha subunit amino-terminal domain. Science **281**: 262-6.

151.   **Zhou, Y.*, et al.*** 2001. The RssB response regulator directly targets sigma(S) for degradation by ClpXP. Genes Dev **15**(5): 627-37.

152.   **Zuo, Y., Wang, Y., and Steitz, T. A.** 2013. The Mechanism of *E. coli* RNA Polymerase Regulation by ppGpp Is Suggested by the Structure of their Complex. Mol Cell **50**(3): 430-6.

Chapter 2:

Genetic dissection of the *Sau* σ$^A_4$ / gp67 interface

Attributions

The work to characterize the *S. aureus* bacteriophage G1-encoded anti-σ^A factor gp67 (previously referred to as ORF67 in the literature) was undertaken in conjunction with Seth A. Darst and Joseph Osmundson, a doctoral student in the Darst laboratory, at The Rockefeller University. J. Osmundson solved the high-resolution X-ray crystal structure of the *Sau* σ^A_4/gp67 complex (see **Appendix 2** or [37]) and provided the structural data and interpretations reported in this chapter. I undertook a structure-independent, unbiased genetic approach to characterize the *Sau* σ^A_4/gp67 interface with the goal of generating mutants in both binding partners that would provide insight into the specificity of gp67 for *Sau* σ^A_4, provide genetic support for the biological relevance of the crystal structure, generate tools to directly test some of the predictions about the mechanism of action of gp67 based on structural modeling, and set up a system with which to study gp67 function *in vivo* and *in vitro* in the experimentally more tractable model bacterium, *E. coli*. The genetic analyses reported in this chapter generated a crucial control for the *in vivo* and *in vitro* experiments reported in [37] using a native *S. aureus* system: a *Sau* σ^A mutant that is specifically deficient for gp67 binding but is otherwise functional. I performed all of the experiments reported in this chapter and wrote the text of the chapter, with editorial assistance from A. Hochschild.

Introduction

All transcription in bacteria is performed by the 380 kDa core RNAP (subunit composition $\alpha_2\beta\beta'\omega$), which is catalytically active for RNA synthesis but cannot engage in sequence-specific DNA binding [31]. Promoter-specific DNA recognition is conferred on core RNAP by the reversible association of the σ initiation factor to form the RNAP holoenzyme (subunit composition $\alpha_2\beta\beta'\omega\sigma$) [15]. Bacteria generally encode multiple σ factors with distinct promoter specificities: a single essential primary σ factor that directs the expression of most genes during exponential growth, and several alternative σ factors that transcribe specific regulons in response to environmental or developmental cues [16, 38, 48]. The multiplicity of and interplay between σ factors endow bacteria with a transcriptional plasticity that enables them to inhabit diverse environments and withstand a variety of stresses. The genome of the Gram-negative model bacterium *E. coli* encodes seven σ factors (*Eco* σ^{70} is the primary σ factor), the low-G+C Gram-positive pathogen *S. aureus* has four (*Sau* σ^A is the primary σ factor), and the related spore-forming low-G+C Gram-positive model bacterium *B. subtilis* has eighteen (*Bsu* σ^A is the primary σ factor) [16, 44].

σ^{70}-family σ factors are modular proteins with up to four regions of conserved sequence that comprise four flexibly linked domains: $\sigma_{1.1}$, σ_2, σ_3, and σ_4 [15, 27, 31]. σ_2 and σ_4 exhibit the highest degree of conservation across all species and make extensive contacts with core RNAP and promoter DNA in the context of holoenzyme. σ^{70}-family σ factors direct transcription from promoters containing two conserved hexameric sequence elements centered at -10 (consensus sequence TATAAT) and -35 (consensus sequence TTGACA) relative to the transcription start site and separated by 16-18 bp of nonconserved sequence, with an optimal spacing of 17 bp [15, 31]. σ_2 is the primary and most extensive determinant for the high affinity binding of σ factors to core RNAP: σ_2 binds a coiled-coil motif in β' (the β'cc), an interaction that is essential for holoenzyme formation and that induces conformational changes in σ_2 that enable it to

functionally engage the -10 promoter element during transcription initiation [15, 31]. In the context of holoenzyme, σ_2 is involved in nucleating promoter melting starting at base position -11 and extending to $\sim$+3 in the open complex [9, 10, 39], and making sequence-specific contacts with the non-template strand of the melted promoter [10, 29]. σ_4 contacts the flexible flap domain of β (the β-flap), an interaction that is not essential for holoenzyme formation but that appropriately positions σ_4 to engage the -35 promoter element at -10/-35 promoters—the major class of promoters in *E. coli*—when σ_2 is simultaneously bound at the -10 element [23, 31].

Transcription initiation is a key regulatory step in coordinated gene expression in bacteria. Each step on the pathway from free holoenzyme and promoter DNA to the transcriptionally competent open complex presents an opportunity for regulation. σ factors are critical during all steps of transcription initiation, since they are required for holoenzyme formation, make sequence-specific contacts with duplex promoter DNA in closed complexes, and are directly involved in promoter melting; thus, σ factors are attractive regulatory targets. Unlike in eukaryotes, a single core RNAP transcribes all genes in bacteria, so a cell's transcriptional profile under any condition is determined by the predominant transcribing holoenzyme species. Transcription activators and repressors augment or antagonize holoenzyme function at specific promoters to increase or decrease transcription from these promoters, engaging in protein/protein contacts with the αCTDs or σ_4 to facilitate the initial recognition of duplex promoter DNA and/or the isomerization(s) to the open complex [1, 15, 40]. Additionally, the competition between different σ factors for a limiting amount of core RNAP determines the nature of the predominant holoenzyme species [14, 22]. σ factor competition is influenced by the abundance of specific active σ factors and their binding affinities for core RNAP. Core RNAP binding affinity is an immutable property of each σ factor and overwhelmingly favors the binding of the primary σ factor [28], so bacteria and their infecting phages have evolved a multitude of approaches to regulate the abundance of active σ factors.

One approach used by both bacteria and phages is to reversibly control σ factor activity post-translationally using anti-σ factors, allowing the cell to maintain a pool of multiple inactive σ factors that do not interfere with the activity of the current predominant holoenzyme but that can enact immediate changes in the cell's transcriptional profile in response to external cues.

Anti-σ factors bind directly and tightly to their cognate σ factor and specifically inhibit transcription from promoters dependent on that σ factor for recognition [3, 17, 20]. Many anti-σ factors inhibit transcription by sequestering their cognate σ factor, occluding core RNAP-binding determinants or inducing conformational changes in σ incompatible with holoenzyme formation. (For example, the *E. coli*-encoded stationary-phase anti-σ^{70} factor Rsd acts primarily in this manner, binding to σ^{70}_2 and σ^{70}_4 simultaneously and preventing these regions from contacting the β'cc and β-flap on RNAP to form the σ^{70}-containing holoenzyme, E•σ^{70} [49]). Some anti-σ factors, however, have been shown to exert their inhibitory functions in the context of the holoenzyme, either by promoting the dissociation of a specific holoenzyme species (e.g. the bacterially-encoded anti-σ^{28}, FlgM [4, 5]), or by forming a stable ternary complex with the bacterial holoenzyme and inhibiting transcription from specific classes of promoters recognized by that holoenzyme. The coliphage T4-encoded anti-σ^{70} AsiA is a well-characterized example of this latter class of anti-σ factors: AsiA binds to *Eco* σ^{70}_4 in the context of E•σ^{70} and induces a conformational change in σ^{70}_4 that renders it unable to functionally engage the -35 element, specifically inhibiting transcription from E•σ^{70}-dependent -10/-35 promoters [25, 42, 51]. The AsiA/E•σ^{70} ternary complex is also redirected to T4 middle promoters [19] through a protein/protein interaction between AsiA and the phage activator MotA [50]. The AsiA-mediated inhibition of transcription from the major class of *E. coli* promoters may also facilitate the formation of an alternative holoenzyme containing the T4-encoded σ factor gp55 (E•gp55) that transcribes T4 late genes [33].

Since most phages utilize the host transcription machinery at various stages during infection, phages have developed a wide array of mechanisms to modify and appropriate host RNAP to favor the expression of viral genes. Anti-σ factors specific to the host's primary σ factor are logical candidates for directing the shutdown of host metabolism and the transcriptional reprogramming induced by virulent phage infection. Prior to this work, only two anti-σ factors specific to a primary σ had been characterized: *E. coli*-encoded Rsd and T4-encoded AsiA. Both target *Eco* σ^{70} but function through entirely different mechanisms, suggesting that a broader range of primary σ-specific anti-σ factors may exist. The largely unexplored landscape of sequenced phage genomes is a potential source of novel anti-σ factors that can contribute to an enhanced understanding of transcription and its regulation and may provide alternatives to antibiotics in combating infectious disease.

A high-throughput functional genomics screen to identify phage-encoded inhibitors of *S. aureus* growth identified gp67 (formerly ORF67), a gene product of the virulent G1 phage, as one such inhibitor [24, 26]. Subsequent biochemical characterization established that gp67: 1) significantly decreases RNA production in *S. aureus* , 2) interacts directly with *Sau* σ^A via a strong interaction with *Sau* σ^A_4 but not with *Eco* σ^{70} or *Eco* σ^{70}_4 despite the high degree of amino acid conservation in σ_4 of the two primary σ factors (**Figure 2.1**), and 3) forms a stable ternary complex with the native *S. aureus* σ^A-containing holoenzyme (E$_{Sau}$•σ^A) [6, 26]. These data revealed that gp67 is an anti-σ factor specific to the primary σ factor of *S. aureus*, and raised the question of whether it functions by a mechanism similar to that of AsiA despite the lack of sequence or structural similarity between the two proteins or presents a novel regulatory paradigm. To interrogate the mechanism of action of gp67, J. Osmundson and S. A. Darst determined the high-resolution X-ray crystal structure of the *Sau* σ^A_4/gp67 complex (**Appendix 2**; [37]). We sought to use unbiased genetic analyses to identify the region(s) and individual residues of *Sau* σ^A_4 and gp67 responsible for the specificity of this interaction, reasoning that

```
        528            538             548            558           568            578
         |              |               |              |             |              |
Eco σ⁷⁰₄  LELPLDSATTESLRAATHDVLAGLTAREAKVLRMRFGIDMNTDYTLEEVGKQFDV
Sau σᴬ₄   -------------KEQLEDVLDTLTDREENVLRLRFGLDDGRTRTLEEVGKVFGV
                               |              |             |              |
                              304            314           324            334

        588            598             608
         |              |               |
Eco σ⁷⁰₄  TRERIRQIEAKALRKLRHPSRSEVLRSFLDD
Sau σᴬ₄   TRERIRQIEAKALRKLRHPSRSKRLKDFMD-
             |              |               |
            344            354             364
```
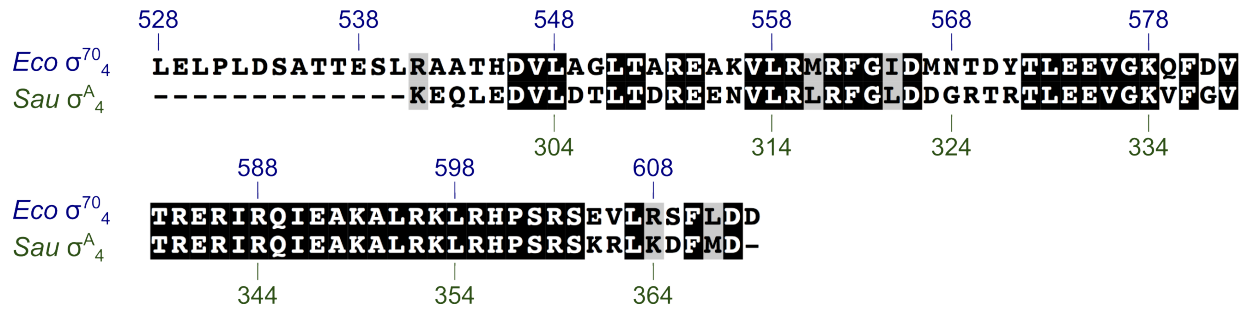
Figure 2.1. Amino acid conservation in region 4 of the primary σ factors of *E. coli* and *S. aureus*. Pairwise alignment of the protein sequences of σ₄ of *Eco* σ⁷⁰ (residues 528-613; labeled in blue) and *Sau* σᴬ (residues 297-368; labeled in green). Identical amino acids at equivalent positions are indicated by a black background; similar amino acids are indicated by a grey background. Gaps in the alignment are indicated by dashes. Over the length of the alignment, *Eco* σ⁷⁰₄ and *Sau* σᴬ₄ are ~67% identical (48 out of 72 residues) and ~74% similar (53 out of 72 residues), indicative of a high degree of sequence conservation. The residues used to define region 4 for *Eco* σ⁷⁰ were based on previous work conducted in the laboratory [35]. *Sau* σᴬ₄ was defined by the fragment used by J. Osmundson for crystallographic experiments.

these residues would be part of the *Sau* σ$^A_4$/gp67 interface and that understanding this interface

would provide insight into the mechanism by which gp67 inhibits *Sau* σ^A-dependent transcription.

(For example, if the *Sau* σ$^A_4$ residues that are part of the interface with gp67 include β-flap-

binding determinants, gp67 could function as a specific inhibitor of *Sau* σ$^A_4$-dependent -10/-35

promoters by preventing the *Sau* σ$^A_4$/β-flap contacts required to appropriately position *Sau* σ$^A_4$

to engage the -35 element, as is the case with AsiA [13, 43]). We further reasoned that the

genetic dissection of the *Sau* σ$^A_4$/gp67 interface would provide *in vivo* validation for the structure

reported by Osmundson *et al.* [37] and might enable us to develop an *E. coli*-based system

where gp67 is functional that would facilitate *in vivo* and *in vitro* analyses of gp67-mediated

transcription inhibition using an array of *E. coli*-derived tools. Here we report that *Sau* σ^A

residues 309-335 define the specificity-determining region for gp67 binding and we identify the

side chains of four key residues within this region that contribute significantly to gp67 binding

specificity, in agreement with the *Sau* σ$^A_4$/gp67 co-crystal structure. We further report that we

can make substitutions in this specificity-determining region in *Eco* σ$^{70}_4$ that enable gp67 binding

and we extend our findings to the primary σ factors of other bacteria. We also describe our

efforts to identify the *Sau* σ$^A_4$-binding determinants of gp67 and the use of genetic assays to test

two structural predictions about the effect of gp67 on core RNAP- and DNA-binding by *Sau* σ$^A_4$

in the context of holoenzyme.

Results

The region comprising *Sau* σ^A residues 309-335 determines the specificity of gp67 binding

      As shown in **Figure 2.1**, there is a high degree of amino acid similarity within region 4 of

Sau σ^A and *Eco* σ^{70}. Despite this similarity, gp67 binds *Sau* σ$^A_4$ but not *Eco* σ$^{70}_4$ [6, 37], and we

wanted to investigate what accounts for this selectivity. In particular, we hoped to identify

specific amino acid differences between *Sau* σ$^A_4$ and *Eco* σ$^{70}_4$ that could account for the binding

selectivity of gp67. To assess whether we could use genetic tools to address this question, we first determined whether we could detect a strong interaction between Sau σ^A_4 and gp67 (and the corresponding lack of interaction between Eco σ^{70}_4 and gp67) in $vivo$ using the bacterial two-hybrid system [7, 8] depicted in **Figure 2.2**. As expected, we were able to detect a strong interaction between an α-Sau σ^A_4 fusion protein and a λCI-gp67 fusion protein using the two-hybrid system, but failed to see an interaction between the α-Eco σ^{70}_4 and λCI-gp67 fusion proteins. [As noted in the legend of **Figure 2.2**, the Eco σ^{70}_4 moiety of all α-Eco σ^{70}_4 fusion proteins used in this dissertation bears substitution D581G, which stabilizes the folded structure of the fusion protein; Sau σ^A_4 bears a glycine at the equivalent position]. We then used the two-hybrid system to identify the specificity-determining region for gp67 binding by constructing a series of $E.$ $coli$ and $S.$ $aureus$ chimeric α-σ_4 fusion proteins and testing their abilities to interact with the λCI-gp67 fusion protein. We replaced different segments of Eco σ^{70}_4 with the corresponding segment of Sau σ^A_4, looking for the smallest swapped region that would enable an otherwise wild-type α-Eco σ^{70}_4 fusion protein to interact with the λCI-gp67 fusion protein. We also constructed the reciprocal chimeras, replacing the equivalent segments of Sau σ^A_4 with the corresponding Eco σ^{70}_4 region to verify that the swap(s) abolished the ability of an otherwise wild-type α-Sau σ^A_4 fusion protein to interact with the λCI-gp67 fusion protein in the two-hybrid assay. The complete panel of chimeric α-σ_4 fusion proteins used in these experiments is depicted in **Figure 2.3**.
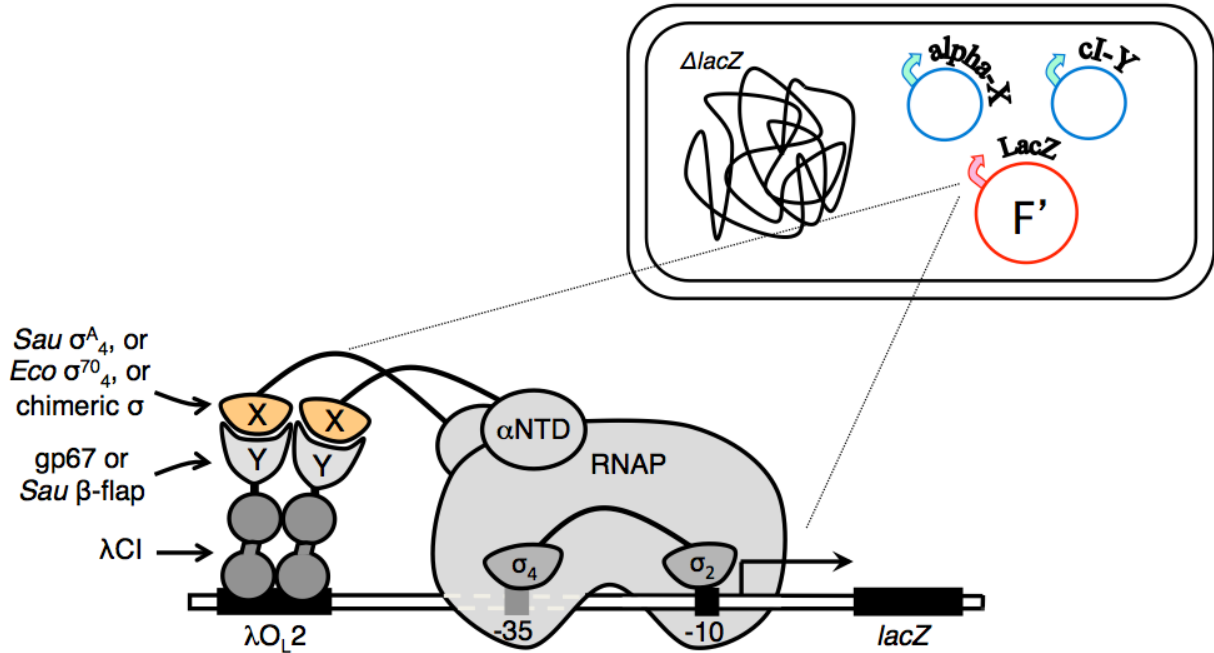
Figure 2.2. Transcription-based bacterial two-hybrid assay. The assay is based on the premise that contact between a protein domain X fused to the N-terminal domain of the α subunit of RNAP (αNTD) and a partner protein domain Y fused to the DNA-bound bacteriophage λ CI protein (λCI) activates transcription from the test promoter $P_{lac}O_L2$-62, which bears λ operator O_L2 centered 62 bp upstream of the start site of the core *lac* promoter. In strain FW102, the test promoter—which directs transcription of a *lacZ* reporter—is located on an F' episome. *lacZ* expression is quantified by β-galactosidase assays, and the level of activation observed is reflective of the strength of the interaction between protein domains X and Y. In this dissertation, protein domain X represents *Sau* σ^A_4, *Eco* σ^{70}_4, or any one of a panel of *Eco* σ^{70}_4/ *Sau* σ^A_4 chimeras schematically depicted in **Figure 2.3**. In all cases, the *Eco* σ^{70}_4 moiety bears substitution D581G, which stabilizes the folded structure of *Eco* σ^{70}_4 and facilitates the detection of two-hybrid interactions [23, 35]. *Sau* σ^A_4 bears a glycine at the corresponding position (*Sau* σ^A G337). Protein domain Y represents gp67 or the *Sau* β-flap (*Sau* RpoB residues 789-917); the two-hybrid interaction with the *Sau* β-flap is used to assess the functional integrity and stable production of the α-σ_4 fusion proteins.
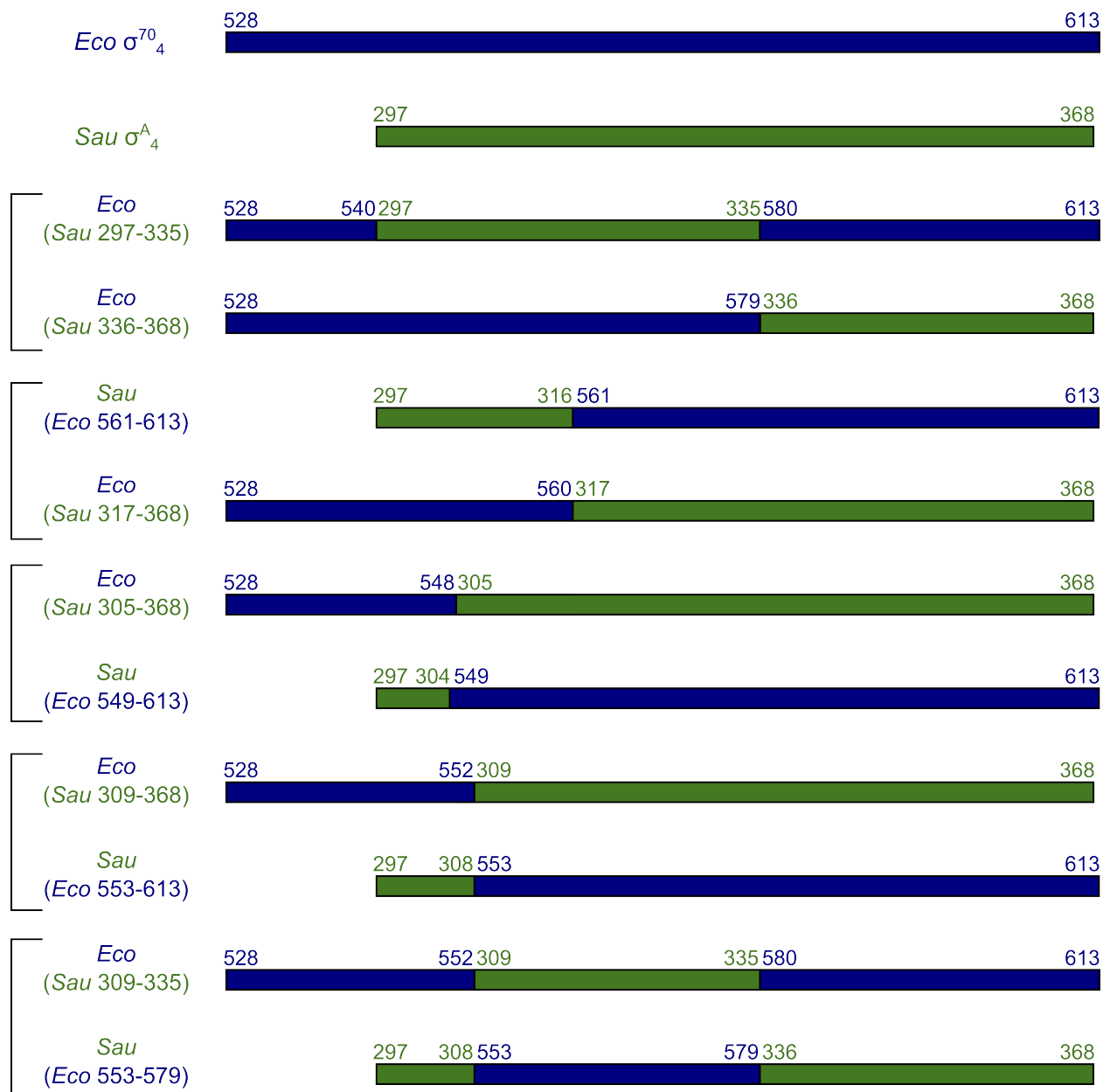
Figure 2.3. Panel of *Eco* σ$^{70}_4$/*Sau* σ$^A_4$ chimeric proteins used to identify the specificity-determining region for gp67 binding. *Eco* σ$^{70}_4$ residues are represented by a blue bar and *Sau* σ$^A_4$ residues by a green bar in this schematic. We created pairs of reciprocal chimeric α-*Eco* σ$^{70}_4$/*Sau* σ$^A_4$ fusion proteins by replacing a specific segment of *Eco* σ$^{70}_4$ with the corresponding residues from *Sau* σ$^A_4$. **(Legend continued on next page.)**

Figure 2.3 (continued) Conversely, we replaced the equivalent region in *Sau* σ^A_4 with the corresponding residues from *Eco* σ^{70}_4. The pairs of reciprocal chimeras are delineated by a bracket on the left of the schematic. Each of these chimeric α-*Eco* σ^{70}_4/*Sau* σ^A_4 fusion proteins, along with native α-*Eco* σ^{70}_4 and α-*Sau* σ^A_4, was assayed for the ability to interact with a λCI-gp67 fusion protein in the bacterial two-hybrid system. The functional integrity of the α-σ_4 fusion proteins was evaluated by testing for a two-hybrid interaction with a λCI-*Sau* β-flap fusion protein and a one-hybrid protein/DNA interaction with an ectopically positioned -35 promoter element. See the text, **Figure 2.4**, and **Figure 2.5** for details.

To assess the structural integrity of the chimeric α-σ$_4$ fusion proteins and verify that any loss of interaction with the λCI-gp67 fusion protein was strictly due to the absence of the appropriate residues in the specificity-determining region for gp67 binding and not to protein misfolding, we took advantage of the interactions of σ$_4$ with core RNAP and promoter DNA. As in the holoenzyme, the σ$_4$ moiety of an α-σ$_4$ fusion protein interacts with the β-flap moiety of a λCI-*Sau* β-flap fusion protein in the context of the two-hybrid system to activate transcription. We can also detect the binding of the σ$_4$ moiety of an α-σ$_4$ fusion protein to an ectopically positioned -35 promoter element *in vivo* using a bacterial one-hybrid system [35], depicted in **Figure 2.4**. Using these secondary assays, we determined that all of the chimeras outlined in **Figure 2.3** except α-*Sau*(*Eco* 561-613) are properly folded and thus informative in defining the gp67 specificity-determining region; α-*Sau*(*Eco* 561-613) is excluded from the data presented.

The N-terminal half of *Sau* σ$^A_4$ has the greatest number of amino acid differences relative to *Eco* σ$^{70}_4$, so we focused on *Sau* σ^A residues 297-335 as the likely source of gp67 binding specificity. As seen in **Figure 2.5A**, an α-*Eco* σ$^{70}_4$ fusion protein bearing *Sau* σ^A residues 297-335 at the corresponding positions (α-*Eco*(*Sau* 297-335)) can interact with the λCI-gp67 fusion protein. Conversely, an α-*Sau* σ$^A_4$ fusion protein bearing the corresponding *Eco* residues at positions 297-335 (α-*Eco*(*Sau* 336-368)) fails to interact with the λCI-gp67 fusion protein. We further narrowed down this region by shifting the N-terminal endpoint of the *Sau* σ$^A_4$ segment included in our chimeric α-σ$_4$ fusion proteins to the start of the next block of nonconserved residues in this region: from *Sau* σ^A residue 297 to residue 305 (see α-*Eco*(*Sau* 305-368) in **Figure 2.5C**) and from residue 305 to residue 309 (see α-*Eco*(*Sau* 309-368) in **Figure 2.5C**). This analysis allowed us to define the minimal region that could switch the specificity of gp67 binding: the presence of *Sau* σ^A residues **309-335** allows *Eco* σ$^{70}_4$ to bind gp67, and replacing these residues in *Sau* σ$^A_4$ with the corresponding *Eco* residues abolishes gp67 binding (see α-*Eco*(*Sau* 309-335) and α-*Sau*(*Eco* 553-579) in **Figure 2.5A**).

70
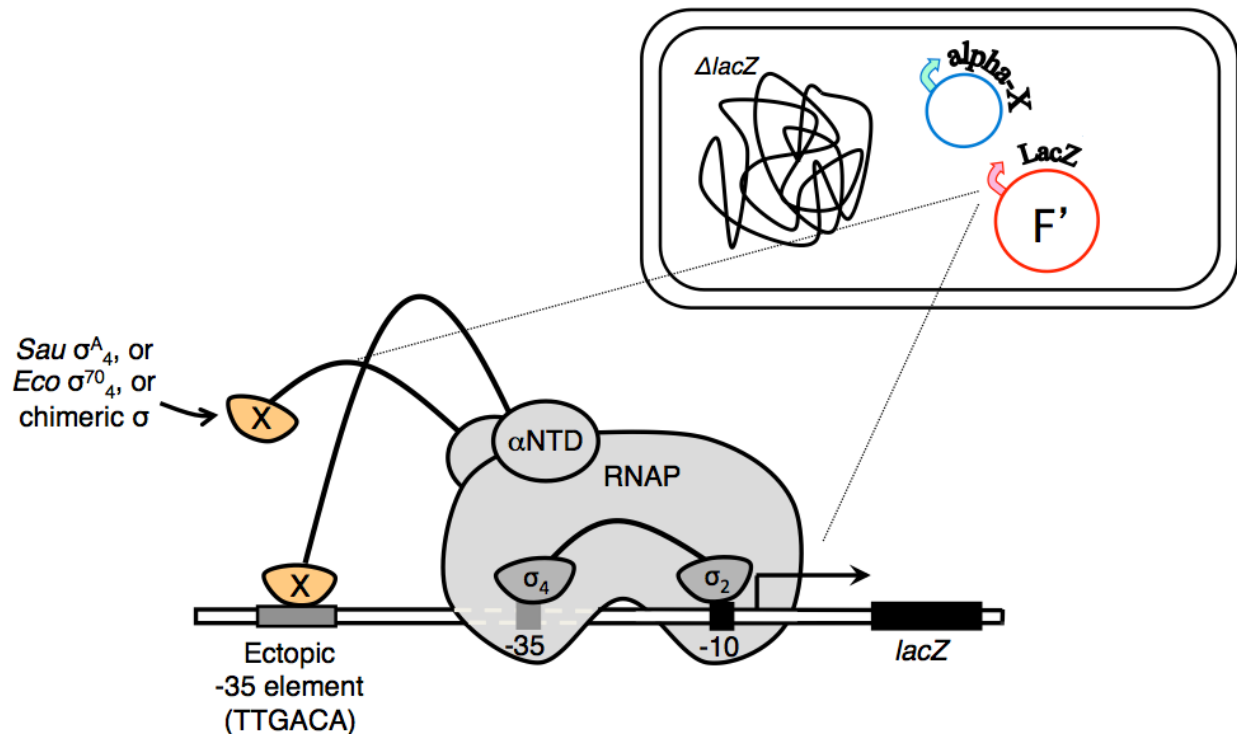
Figure 2.4. Transcription-based bacterial one-hybrid system to detect σ_4/-35 element interactions *in vivo*. The assay is based on the premise that contact between a protein domain X fused to the αNTD and its DNA-binding site located upstream of the test promoter, which directs transcription of a *lacZ* reporter, activates transcription from the test promoter. In this formulation, the test promoter is P$_{lac}$Cons-35C, which bears a consensus -35 element (TTGACA) upstream of the core *lac* promoter that serves as a binding site for the σ_4 moiety of the α-σ_4 fusion protein. In strain FW102, the test promoter is located on an F' episome. *lacZ* expression is quantified by β-galactosidase assays, and the level of activation observed is reflective of the strength of the interaction between protein domain X and the ectopically positioned -35 element. In this dissertation, protein domain X represents *Sau* σ^A_4, *Eco* σ^{70}_4, or any one of a panel of *Eco* σ^{70}_4/ *Sau* σ^A_4 chimeras outlined in **Figure 2.3**. In all cases, the *Eco* σ^{70}_4 moiety bears the stabilizing substitution D581G [23, 35].
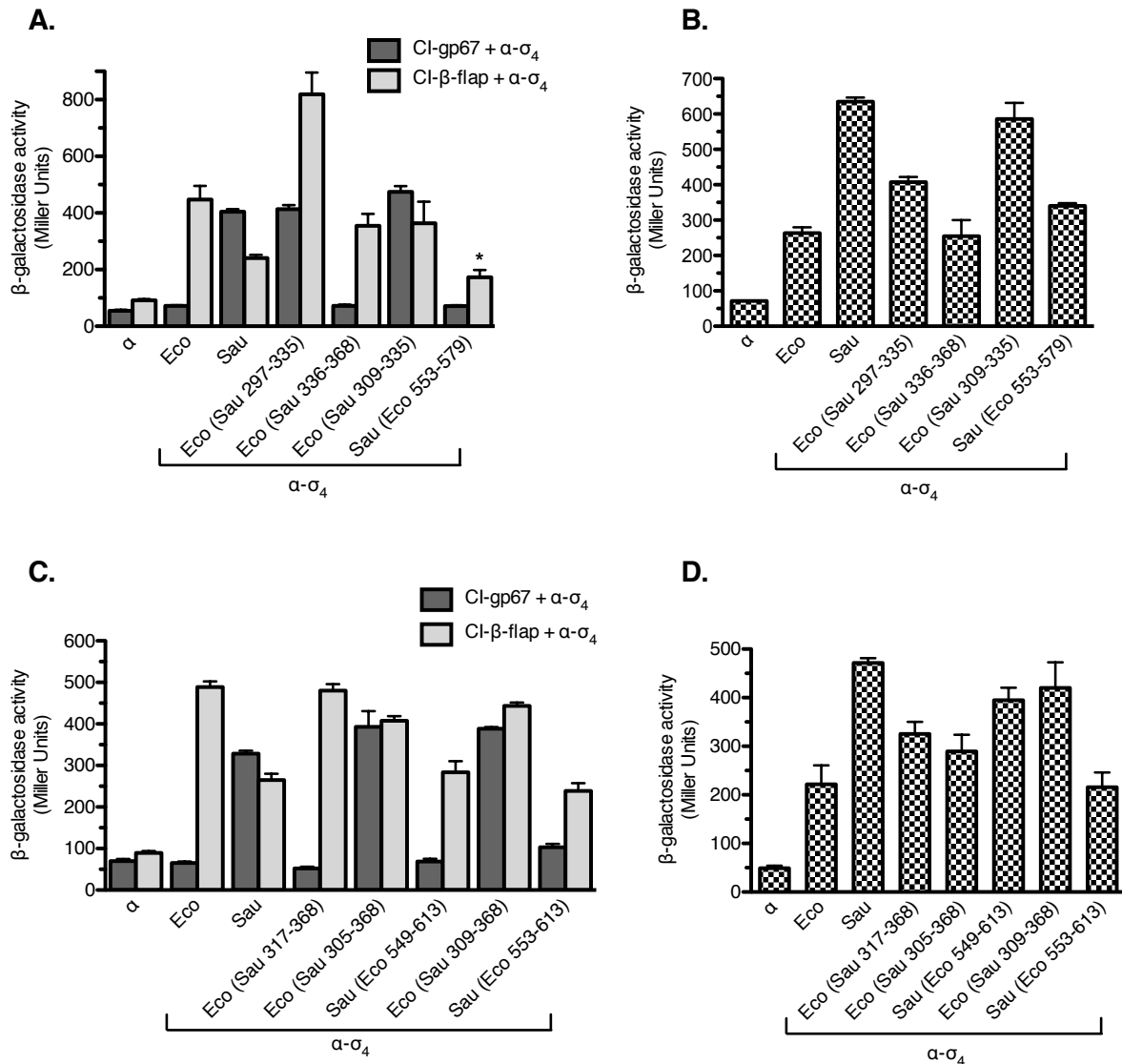
Figure 2.5. *Sau* σ^A **residues 309-335 define the specificity-determining region for gp67 binding.** β-galactosidase activity is reported in Miller Units. Each bar represents the average Miller Units of three biological replicates assayed in the same experiment; error bars are ± standard deviation (SD). **(A)** Two-hybrid interactions of a subset of chimeric α-σ_4 fusion proteins that identify the divergent N-terminus of *Sau* σ^A_4 (residues 297-335) as containing determinants important for the specificity of gp67 binding and further narrow down the specificity-determining region to residues 309-335. Dark grey bars represent the interaction with the λCI-gp67 fusion protein. **(Legend continued on next page.)**

Figure 2.5 (continued) Light grey bars represent the interaction with the λCI-*Sau* β-flap fusion protein to assess the structural integrity of the α-σ_4 chimeras. The results indicate that *Eco* σ^{70}_4 bearing *Sau* residues 309-335 at the corresponding positions (*Eco*(*Sau* 309-335)) is fully competent to interact with gp67 and, conversely, that *Sau* σ^A_4 bearing the *Eco* residues in this region (*Sau*(*Eco* 553-579)) no longer interacts detectably with gp67. The asterisk (*) indicates that the observed lack of gp67 binding by the α-*Sau*(*Eco* 553-579) fusion protein is not the result of misfolding despite the weaker interaction with the λCI-*Sau* β-flap fusion protein, since α-*Sau*(*Eco* 553-579) can functionally engage the -35 element (binding nearly as well as α-*Eco*(*Sau* 297-335)). **(B)** One-hybrid interactions of the chimeric α-σ_4 fusion proteins with an ectopically positioned -35 element, a second test for proper folding and functionality of the fusion proteins. The results indicate that all of the α-σ_4 fusion proteins are properly folded and functional. **(C)** Two-hybrid interactions of additional chimeric α-σ_4 fusion proteins used to identify *Sau* σ^A residues 309-335 as the minimal specificity-determining region. Dark grey bars represent the interaction with the λCI-gp67 fusion protein and light grey bars the interaction with the λCI-*Sau* β-flap fusion protein to assess the structural integrity of the α-σ_4 chimeras. All α-σ_4 fusion proteins that have the *Sau* σ^A residues at the positions corresponding to 309-335 can bind gp67, and all the α-σ_4 fusion proteins can bind the *Sau* β-flap and are thus properly folded. **(D)** One-hybrid interactions of the additional chimeric α-σ_4 fusion proteins with an ectopically positioned -35 element to further assess proper folding and functionality of the fusion proteins. The results indicate that all of the α-σ_4 fusion proteins are functional.

<u>The side chains of *Sau* σ^A residues 309, 312, 313, and 335 contribute to gp67 binding specificity</u>

There are eleven amino acid differences between *Sau* σ^A_4 and *Eco* σ^{70}_4 in the identified specificity-determining region; two of these involve similar residues in the two proteins (*Sau* σ^A L316 / *Eco* σ^{70} M560 and *Sau* σ^A L321 / *Eco* σ^{70} I565), but the remaining nine differences comprise residues with different chemical properties and became the focus of subsequent analyses. The divergent residues cluster into three locations in the primary sequence of the specificity-determining region: *Sau* σ^A residues 309, 312, and 313 at the beginning; residues 323-327 in the middle; and residue 335 at the end. We were interested in determining which residues within the identified specificity-determining region contributed significantly to gp67 binding specificity, so we conducted two complementary mutational studies of this region. In the first study, we made sequential substitutions in *Sau* σ^A_4, changing divergent residues to the corresponding *Eco* σ^{70}_4 residue in the context of an otherwise wild-type α-*Sau* σ^A_4 fusion protein and looking for the combination of substitutions that resulted in the loss of the interaction with the λCI-gp67 fusion protein in the two-hybrid system without compromising structural integrity (measured by the two-hybrid interaction with the λCI-*Sau* β-flap fusion protein and the one-hybrid interaction with the -35 element, as before). In the second study, we sequentially changed *Eco* σ^{70}_4 residues to the corresponding *Sau* residue in the context of an otherwise wild-type α-*Eco* σ^{70}_4 fusion protein (bearing the D581G substitution) to determine which substitutions enabled *Eco* σ^{70}_4 to bind gp67. A preliminary mutational study where we made substitutions in *Sau* σ^A_4 residues 323-327 revealed that single substitutions at residues 323 and 325, and a triple substitution at residues 325, 326, and 327 affected the structural integrity of the fusion proteins (data not shown). Thus, we focused on *Sau* σ^A residues 309, 312, 313, and 335 (corresponding to *Eco* σ^{70} residues 553, 556, 557, and 579) for our studies.

We began our analysis with *Sau* σ^A_4, making various combinations of substitutions at residues 309, 312, 313, and 335, and looking to specifically abolish gp67 binding. We found that

a double substitution at *Sau* σ^A residues 312 and 313 to the corresponding *Eco* σ^{70} residues

reduced gp67 binding by approximately twofold relative to wild-type *Sau* σ$^A_4$ in the two-hybrid

system (see α-*Sau*(*Eco* 556, 557) in **Figure 2.6A**), which suggested that either one or both of

the amino acid side chains at residues E312 and N313 in *Sau* σ$^A_4$ are important for the

specificity of gp67 binding. (Since this is a double substitution, we cannot tease apart the

individual effects of changes in each residue, but from a second substitution series detailed in

Appendix 1 we know that residue E312 contributes to this specificity). The addition of the

substitution at residue 309 reduced gp67 binding by threefold relative to wild-type *Sau* σ$^A_4$ but

did not fully abolish it (see α-*Sau*(*Eco* 553, 556, 557) in **Figure 2.6A**), which implicated the side

chain of residue D309 in determining the specificity of gp67 binding. Incorporating a fourth

substitution at residue 335 fully abolished gp67 binding (see α-*Sau*(*Eco* 553, 556, 557, 579) in

Figure 2.6A), suggesting that the side chain of residue V335 is also involved in establishing this

binding specificity. Importantly, each of these versions of the α-*Sau* σ$^A_4$ fusion is properly folded

and functional, since each can engage in productive interactions with the λCI-*Sau* β-flap fusion

protein in the two-hybrid system (**Figure 2.6A**, light grey bars) and with the -35 element in the

one-hybrid system (**Figure 2.6B**). These data led us to conclude that the side chains of *Sau* σ^A

residues <u>309, 312, 313, and 335</u> contribute to the specificity of gp67 for *Sau* σ$^A_4$. (The

quadruply-substituted version of *Sau* σ$^A_4$, *Sau*(*Eco* 553, 556, 557, 579), was used by

Osmundson *et al.* for the functional characterization of gp67 using a native *Sau* system [37]).

To complement the analyses using *Sau* σ$^A_4$ and provide further support for the role of the

side chains of *Sau* σ^A residues 309, 312, 313, and 335 in the specificity of gp67 binding, we

created a similar substitution series using *Eco* σ$^{70}_4$ where we changed the *Eco* residues to the

corresponding *Sau* residue and looked for the combination of substitutions that enabled gp67 to

bind to otherwise wild-type *Eco* σ$^{70}_4$ bearing the D581G substitution. Based on the results of the

Sau σ$^A_4$ substitution series, we expected that replacing *Eco* σ^{70} residues 553, 556, 557, and 579
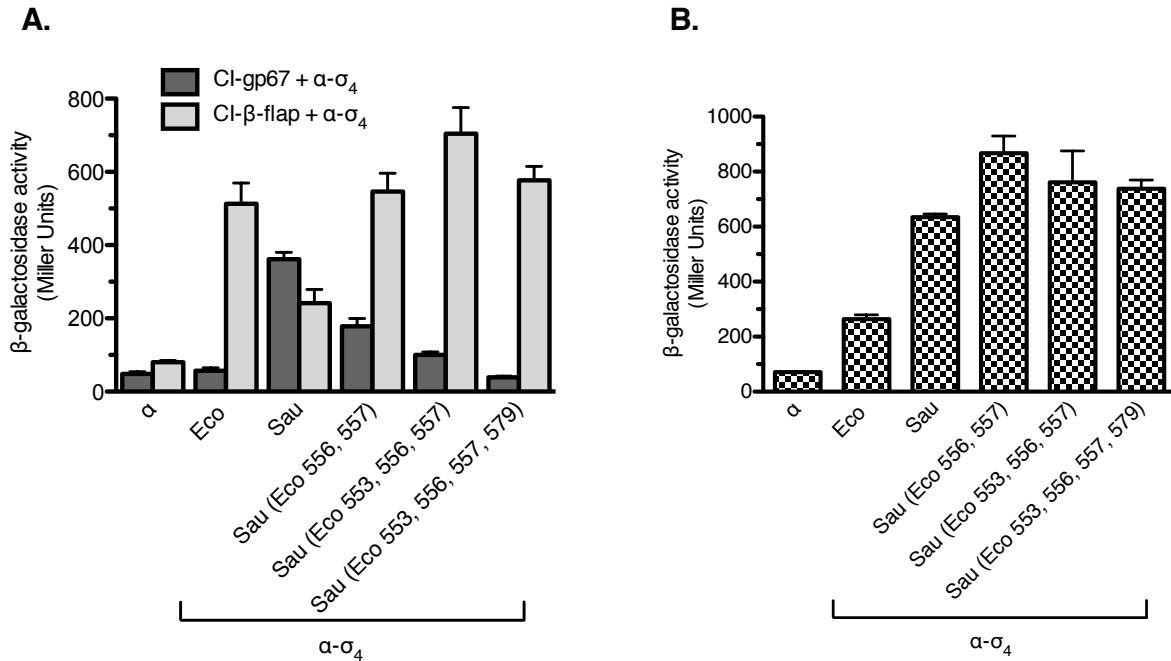
Figure 2.6. The side chains of *Sau* σ^A residues 309, 312, 313, and 335 contribute to the specificity of gp67 binding. β-galactosidase activity is reported in Miller Units. Each bar represents the average Miller Units of three biological replicates assayed in the same experiment; error bars are ± standard deviation (SD). **(A)** Two-hybrid interactions of a substitution series of α-*Sau* σ$^A_4$ fusion proteins, starting with a double substitution at residues 312 and 313 and sequentially incorporating additional substitutions at the remaining divergent residues. The substitutions in each instance are to the residue found in *Eco* σ$^{70}_4$ at the corresponding position. Dark grey bars represent the interaction with the λCI-gp67 fusion protein and light grey bars the interaction with the λCI-*Sau* β-flap fusion protein to assess the structural integrity of the α-*Sau* σ$^A_4$ fusion proteins. The results indicate that a version of *Sau* σ$^A_4$ bearing *Eco* σ^{70} residues at positions 309, 312, 313, and 335 (corresponding to *Eco* σ^{70} 553, 556, 557, and 579) is completely deficient for gp67 binding but is properly folded and functional, suggesting that the identity of the amino acid at these positions is important in establishing the specificity of gp67 for its cognate σ factor, *Sau* σ^A. **(Legend continued on next page.)**

Figure 2.6 (continued) (B) One-hybrid interactions of the substituted α-*Sau* σ$^A_4$ fusion proteins

with an ectopically positioned -35 element, a second test for functionality. The results confirm

that all of the α-*Sau* σ$^A_4$ fusion proteins are properly folded and functional.

with the corresponding four *Sau* σ^A residues would switch the specificity for gp67 without significantly altering the structural integrity of the α-*Eco* σ$^{70}_4$ fusion protein. As seen in **Figure 2.7A**, neither a single substitution at *Eco* σ^{70} residue 553 (see α-*Eco*(*Sau* 309)), nor a double substitution at residues 553 and 556 (see α-*Eco*(*Sau* 309, 312)), nor a triple substitution at residues 553, 556, and 557 (see α-*Eco*(*Sau* 309, 312, 335)) enabled gp67 binding. The quadruply-substituted version of *Eco* σ$^{70}_4$ bearing *Sau* σ^A residues 309, 312, 313, and 335 in place of the corresponding *Eco* residues was fully competent to interact with gp67 (see α-*Eco*(*Sau* 309, 312, 313, 335)). These data raised the possibility that the Q579V substitution alone contributed to the ability of α-*Eco*(*Sau* 309, 312, 313, 335) to bind gp67, but a second substitution series detailed in **Appendix 1** revealed that bearing the *Sau* σ^A residue at *Eco* σ^{70} position 579 does not, by itself, enable gp67 binding.

Unlike what we observed with the quadruply-substituted *Sau* σ$^A_4$, the quadruply-substituted α-*Eco* σ$^{70}_4$ fusion protein appears to be less stably folded: the two-hybrid interaction with the λCI-*Sau* β-flap fusion protein is significantly weaker than it is for the native α-*Sau* σ$^A_4$ or α-*Eco* σ$^{70}_4$ fusion proteins (**Figure 2.7A**, light grey bars), as is its binding to the ectopically positioned -35 element in the one-hybrid system (**Figure 2.7B**). Both of these interactions are above background, and full-length *Eco* σ^{70} bearing these four substitutions (and the D581G substitution) can complement the depletion of wild-type chromosomally-encoded *Eco* σ^{70} and direct transcription from *Eco* σ^{70}-dependent promoters *in vitro* (see **Chapter 3**). Thus, we believe that the α-*Eco*(*Sau* 309, 312, 313, 335) fusion protein is functional, and that the observed interaction with gp67 is specific and is due to the presence of *Sau* σ^A side chains at the four residues identified as important for the specificity of gp67 binding. Versions of *Eco* σ$^{70}_4$ and full-length *Eco* σ^{70} bearing these substitutions were used in subsequent *in vivo* and *in vitro* functional analyses (see **Chapter 3**).
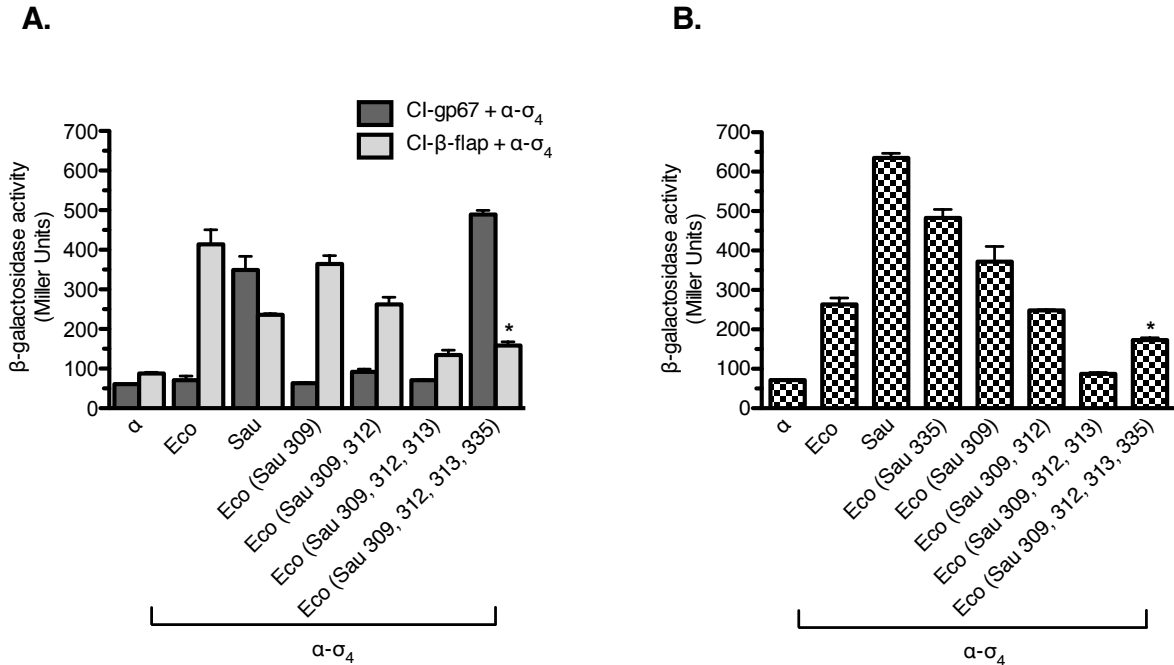
Figure 2.7. *Eco* σ$^{70}_4$ can be modified to bind gp67 with *Sau* σ^A residues 309, 312, 313, and 335. β-galactosidase activity is reported in Miller Units. Each bar represents the average Miller Units of three biological replicates assayed in the same experiment; error bars are ± standard deviation (SD). **(A)** Two-hybrid interactions of a substitution series of α-*Eco* σ$^{70}_4$ fusion proteins, starting with a substitution at residue 553 and sequentially incorporating additional substitutions at the remaining divergent residues. The substitutions in each instance are to the residue found in *Sau* σ$^A_4$ at the corresponding position, and all α-*Eco* σ$^{70}_4$ fusion proteins bear the D581G substitution. Dark grey bars represent the interaction with the λCI-gp67 fusion protein and light grey bars the interaction with the λCI-*Sau* β-flap fusion protein to assess the structural integrity of the α-*Eco* σ$^{70}_4$ fusion proteins. The results indicate that a version of *Eco* σ$^{70}_4$ bearing *Sau* σ^A residues at positions 553, 556, 557, and 579 (corresponding to *Sau* σ^A 309, 312, 313, and 335) is fully competent to bind gp67. The asterisk (*) indicates that, despite a weak interaction of the α-*Eco*(*Sau* 309, 312, 313, 335) and λCI-*Sau* β-flap fusion proteins, this interaction is above background **(Legend continued on next page.)**

79

Figure 2.7 (continued) ($\sim$1.8-fold activation of transcription relative to α) and is similar to what

is observed in the two-hybrid system when using an α-*Eco* σ$^{70}_4$ fusion where the σ$_4$ moiety lacks

the D581G substitution, an interaction known to occur in the holoenzyme (not shown). *In vivo*

and *in vitro* data in **Chapter 3** support our conclusion that the α-*Eco*(*Sau* 309, 312, 313, 335)

fusion protein, though likely less stable than the native α-*Sau* σ$^A_4$ fusion protein or the D581G

version of the α-*Eco* σ$^{70}_4$ fusion protein, is functional. Like the α-*Eco*(*Sau* 309, 312, 313, 335)

fusion protein, the triply substituted α-*Eco*(*Sau* 309, 312, 313) fusion protein interacts relatively

weakly with the λCI-*Sau* β-flap fusion protein, but comparison of the results with the quadruply

and triply substituted variants indicates that the failure of the α-*Eco*(*Sau* 309, 312, 313) fusion

protein to interact with the λCI-gp67 fusion protein is interpretable. **(B)** One-hybrid interactions

of the substituted α-*Eco* σ$^{70}_4$ fusion proteins with an ectopically positioned -35 element, a

second test for functionality. The asterisk (*) indicates that, despite a weaker interaction of the

α-*Eco*(*Sau* 309, 312, 313, 335) fusion protein with the -35 element, this interaction is above

background ($\sim$2.4-fold activation relative to α); we note that this stimulatory effect is significantly

greater than that of the α-*Eco* σ$^{70}_4$ fusion protein lacking the D581G substitution [35].

Having used genetics to identify *Sau* σ^A residues 309-335 as the specificity-determining region for gp67 binding and the side chains of residues 309, 312, 313, and 335 as important for establishing this specificity, we next wanted to determine whether our results (summarized in **Figure 2.8A**) were corroborated by the *Sau* σ^A_4/gp67 co-crystal structure from the Darst laboratory. We expected the gp67 specificity-determining region to be part of the *Sau* σ^A_4/gp67 interface and the four residues identified as important for this specificity to be surface-exposed and engaged in interactions with gp67 residues. As evidenced in **Figure 2.8B**, the genetically-identified specificity-determining region for gp67 binding on *Sau* σ^A_4 (highlighted in yellow) lies at the interface between *Sau* σ^A_4 and gp67, and the side chains of *Sau* σ^A residues 309, 312, 313, and 335 (represented by yellow sticks) are surface-exposed and are oriented toward gp67. An analysis of the crystallographic data conducted by J. Osmundson revealed that residue D309 makes hydrogen bonds with residues near the N-terminus of gp67; residue E312 forms salt bridges with residues K2 and K195 of gp67 that would be disrupted when residue E312 is replaced with A556 from *Eco* σ^{70}; residue N313 makes a hydrogen bond with residue N188 of gp67 and would likely be repelled by K2 and K195 of gp67 when replaced with K557 from *Eco* σ^{70}; and residue V335 contacts hydrophobic residues F133, Y180, and G184 of gp67, interactions that would be disrupted by the polar side chain of Q579 of *Eco* σ^{70} [J. Osmundson, pers. comm. and [37]]. The concordance between the genetic and structural data provides independent support for the biological relevance of the *Sau* σ^A_4/gp67 complex that was captured in the crystal and further validates the transcription-based bacterial two-hybrid and one-hybrid systems as useful tools to test structural predictions *in vivo* [see, for example, 12, 21, 36, 46].

A.

Specificity determining region

$Eco\ \sigma^{70}_4$ LELPLDSATTESLRAATHDVLAGLTAREAKVLRMRFGIDMNTDYTLEEVGKQFDV

$Sau\ \sigma^{A}_4$ -------------KEQLEDVLDTLTDREENVLRLRFGLDDGRTRTLEEVGKVFGV

(Top scale: 528, 538, 548, 558, 568, 578; bottom scale: 304, 314, 324, 334)

$Eco\ \sigma^{70}_4$ TRERIRQIEAKALRKLRHPSRSEVLRSFLDD

$Sau\ \sigma^{A}_4$ TRERIRQIEAKALRKLRHPSRSKRLKDFMD-

(Top scale: 588, 598, 608; bottom scale: 344, 354, 364)

B.

90°
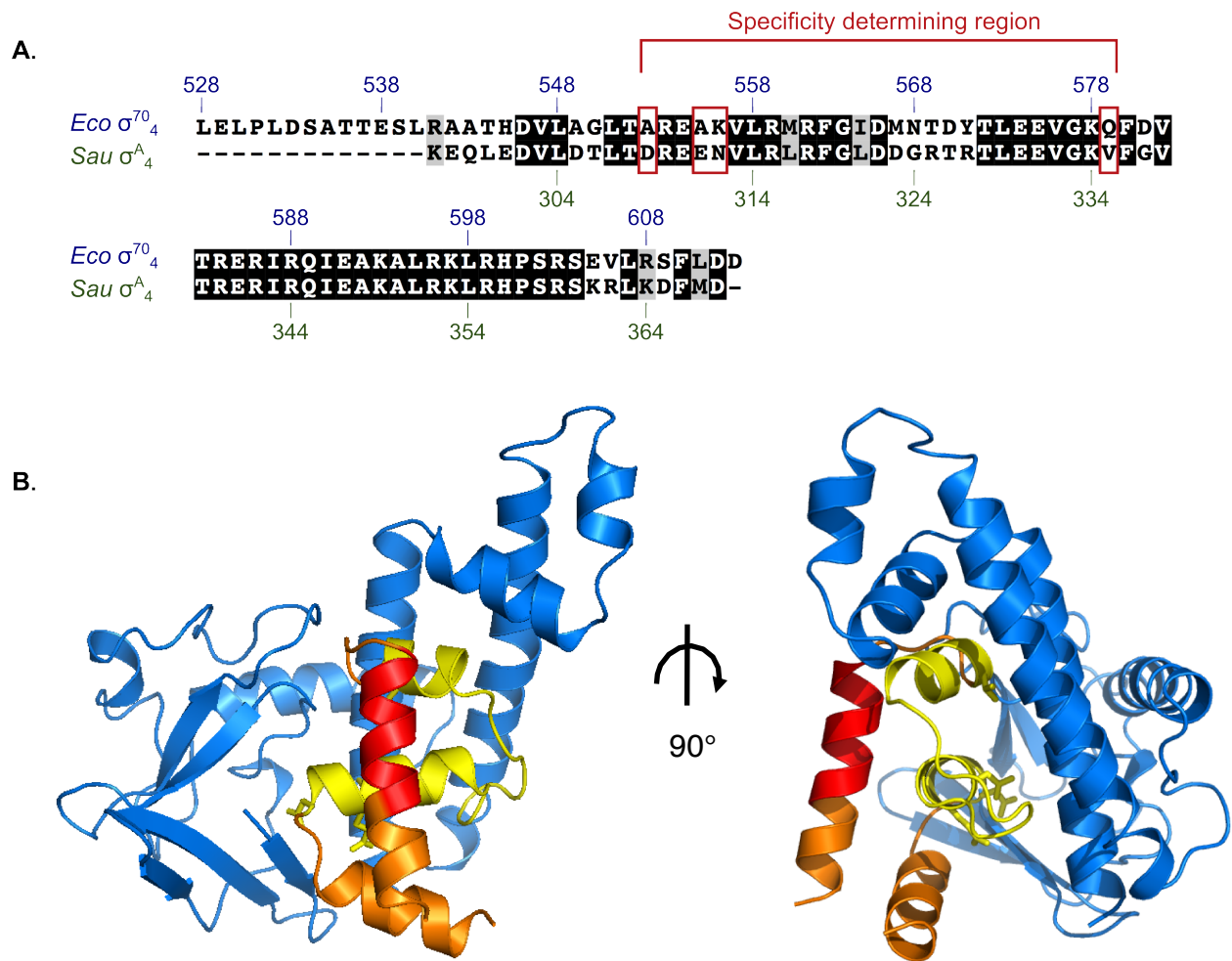
Figure 2.8. Concordance between genetically-deduced results and crystal structure of the *Sau* σ$^A_4$/gp67 complex. (A) Alignment of the protein sequences of σ$_4$ of *Eco* σ^{70} (residues 528-613) and *Sau* σ^A (residues 297-368). Identical amino acids at equivalent positions are indicated by a black background; similar amino acids are indicated by a grey background; gaps in the alignment are indicated by dashes. The region of σ$_4$ identified by our genetic analysis as determining the specificity of gp67 binding (*Sau* σ^A residues 309-335, corresponding to *Eco* σ^{70} residues 553-579) is delineated by a red bracket. Individual nonconserved residues within this region whose side chains play a role in the specificity of gp67 binding (*Sau* σ^A residues 309, 312 313, and 335, which correspond to *Eco* σ^{70} residues 553, 556, 557, and 579) are indicated by red boxes. **(Legend continued on next page.)**

Figure 2.8 (continued) (B) Two orthogonal views of the *Sau* σ^A_4/gp67 co-crystal structure, with both proteins shown in ribbon format. Gp67 is colored in blue; the β-sheet-rich domain comprises the N-terminal region of gp67, and the α-helical domain the C-terminal region. *Sau* σ^A_4 is colored in orange, with the genetically-identified specificity determining region (residues 309-335) highlighted in yellow, and the -35 element-binding recognition helix highlighted in red. The side chains of *Sau* σ^A residues 309, 312, 313, and 335 are shown in stick form and colored in yellow. The structural data reveal that gp67 forms an extensive interface with *Sau* σ^A_4 (~2,800 $Å^2$) through both its N-terminal and C-terminal domains [37]. As expected, the identified specificity-determining region of *Sau* σ^A_4 is part of this interface, and the four residues highlighted in **(A)** make contacts with gp67 residues. The DNA recognition helix is largely distinct from the interface with gp67. See text for details on these interactions. [PyMOL file of the *Sau* σ^A_4/gp67 structure provided by J. Osmundson; subsequent manipulations performed by me using PyMOL].

<u>The identified specificity determinants mediate the binding of gp67 to *B. subtilis* σ^A_4</u>

Based on our observations that we could switch the specificity of gp67 for its target in *Eco* and *Sau* by modifying just four residues, we became interested in further investigating gp67 specificity and the importance of the identified determinants using the primary σ factors of other bacteria. We reasoned that if these four residues were sufficient to specify gp67 binding, any σ_4 bearing the same amino acid side chains as *Sau* σ^A_4 at the corresponding positions would be able to bind to—and be inhibited by—gp67. Furthermore, we might be able to make any primary σ factor a target for gp67 simply by modifying these residues. We aligned the protein sequences of the primary σ factors of various Gram-negative and Gram-positive bacteria and looked at the identity of the residues corresponding to *Sau* σ^A positions 309, 312, 313, and 335. We found that bacteria that were closely related to *E. coli* had the *Eco* residues at the equivalent positions (and would thus be expected to be unable to bind gp67), and that those related to *S. aureus* had the *Sau* residues at the equivalent positions (and thus are predicted to be targeted by gp67).

We chose the primary σ of the low G+C Gram-positive bacterium *B. subtilis* (*Bsu* σ^A), which is nearly identical to *Sau* σ^A in region 4 and is a perfect match to *Sau* at the identified gp67 specificity determinants, and the primary σ of the Gram-negative bacterium *T. aquaticus* (*Taq* σ^A)—which differs from both *Sau* σ^A and *Eco* σ^{70} at most of these positions—for further analysis. The alignment of *Eco* σ^{70}_4, *Sau* σ^A_4, *Bsu* σ^A_4, and *Taq* σ^A_4 is shown in **Figure 2.9**. *Bsu* σ^A_4 is of particular interest in light of *in vitro* transcription results using an *E. coli*-based system presented in **Chapter 3** because the smaller evolutionary distance between *B. subtilis* and *S. aureus* likely results in fewer differences in fine RNAP structure, promoter architecture, and regulation, making a *B. subtilis*-based system less heterologous. Additionally, RNAP purification and *in vitro* transcription protocols are well established in *B. subtilis*, and promoters of various classes are well characterized, so detailed mechanistic studies can be readily conducted in this organism using existing genetic and biochemical tools.
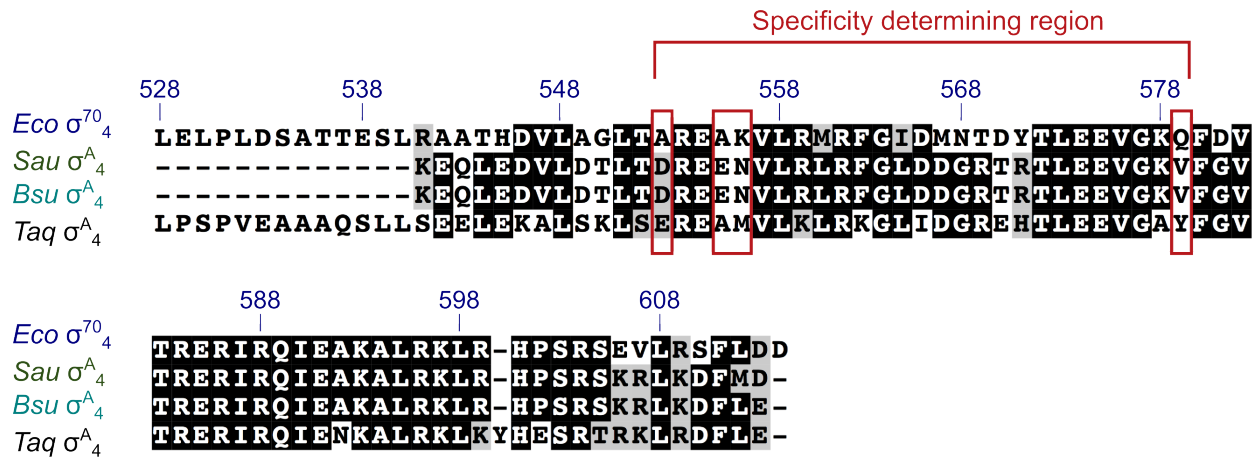
Figure 2.9. Amino acid conservation in region 4 and the gp67 specificity determinants of the primary σ factors of *E. coli*, *S. aureus*, *B. subtilis*, and *T. aquaticus*. Multiple alignment of the protein sequences of region 4 of *Eco* σ^{70} (residues 528-613; labeled in blue), *Sau* σ^{A} (residues 297-368), *Bsu* σ^{A} (residues 300-371), and *Taq* σ^{A} (residues 353-438). The residues used to define region 4 for *Eco* σ^{70} were based on previous work conducted in the laboratory [35]; *Sau* σ^{A}_4 was defined by the fragment used by J. Osmundson for crystallographic experiments; and *Bsu* σ^{A}_4 and *Taq* σ^{A}_4 were defined by the alignments of the full σ factors with *Eco* σ^{70}_4 and *Sau* σ^{A}_4. Identical amino acids at equivalent positions are indicated by a black background, similar amino acids are indicated by a grey background, and gaps in the alignment are indicated by dashes. The specificity-determining region for gp67 binding is indicated by a red bracket, and the four positions identified as important for this specificity are highlighted by red boxes. *Bsu* σ^{A}_4 is identical to *Sau* σ^{A}_4 at these positions (and at most of σ_4), and *Taq* σ^{A}_4 differs from both *Sau* σ^{A}_4 and *Eco* σ^{70}_4 at three of these positions (though the *Taq* residue at the position corresponding to *Sau* σ^{A}_4 D309 of is a similar E).

Since *Bsu* σ^A_4 has the same residues as *Sau* σ^A_4 at the positions corresponding to *Sau* σ^A residues 309, 312, 313, and 335, we predicted that wild-type *Bsu* σ^A_4 would bind gp67 and that a quadruply-substituted version of *Bsu* σ^A_4 bearing the *Eco* σ^{70} residues at the corresponding gp67 specificity determinants would be unable to bind gp67. Conversely, we expected wild-type *Taq* σ^A_4 to be unable to bind gp67, and that substituting the residues at the gp67 specificity determinants with the corresponding *Sau* σ^A residues would enable gp67 binding. To test these predictions, we constructed the α-*Bsu* σ^A_4, α-*Bsu*(*Eco* 553, 556, 557, 579), α-*Taq* σ^A_4, and α-*Taq*(*Sau* 309, 312, 313, 335) fusion proteins and assayed the ability of each to interact with a λCI-gp67 fusion protein in the two-hybrid system. As before, we used the two-hybrid interaction with the λCI-*Sau* β-flap fusion protein and the one-hybrid interaction with an ectopically positioned -35 element to assess the structural integrity of the α-σ_4 fusion proteins. The α-*Taq*(*Sau* 309, 312, 313, 335) fusion protein was severely deficient in β-flap and -35 element binding, suggesting that it is misfolded; it is thus excluded from the data presented.

As expected, the α-*Bsu* σ^A_4 fusion protein interacted with λCI-gp67 as strongly as α-*Sau* σ^A_4, while the quadruply-substituted α-*Bsu*(*Eco* 553, 556, 557, 579) fusion protein was completely deficient for gp67 binding; the α-*Taq* σ^A_4 fusion protein did not interact with λCI-gp67 (**Figure 2.10A**; dark grey bars). Importantly, all of these fusion proteins are properly folded and functional, since they engage in productive interactions with the λCI-*Sau* β-flap fusion protein (**Figure 2.10A**; light grey bars) and the -35 element (**Figure 2.10B**). These data suggest that the interaction between gp67 and *Bsu* σ^A_4 is specified by the same determinants that mediate its interaction with *Sau* σ^A_4, and that the side chains of *Bsu* σ^A residues 312, 315, 316, and 338 (corresponding to *Sau* σ^A residues 309, 312, 313, and 335) contribute significantly to this binding. An experiment performed to look at the *in vivo* effects of gp67 expression in *B. subtilis* is presented in **Chapter 3**, and the further development of a *B. subtilis*-based *in vivo* and *in vitro* system to study gp67-mediated transcription inhibition is discussed in **Chapter 4**.
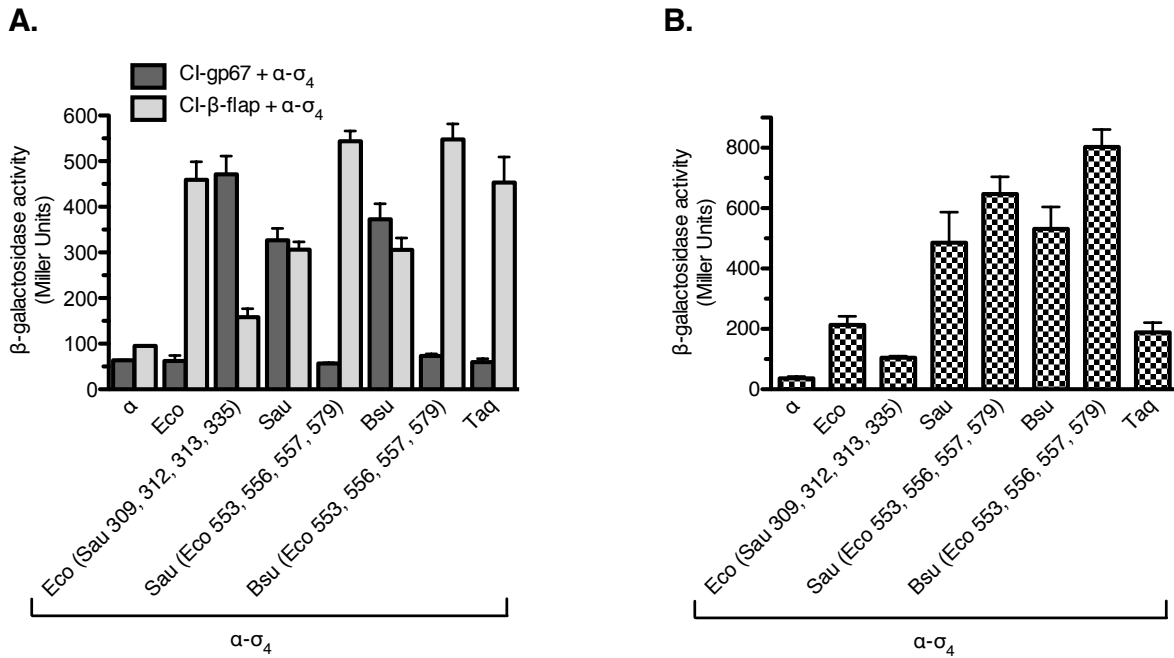
Figure 2.10. The binding of gp67 to *B. subtilis* σ^A_4 is specified by the residues corresponding to *Sau* σ^A 309, 312, 313, and 335. β-galactosidase activity is reported in Miller Units. Each bar represents the average Miller Units of three biological replicates assayed in the same experiment; error bars are ± standard deviation (SD). **(A)** Two-hybrid interactions of α-σ_4 fusion proteins from *Eco* σ^{70}, *Sau* σ^A, *Bsu* σ^A, and *Taq* σ^A, and the corresponding quadruply-substituted versions of each fusion protein. In the case of *Sau* σ^A and *Bsu* σ^A, the substitutions are to the corresponding *Eco* σ^{70} residue at the positions equivalent to *Sau* σ^A 309, 312, 313, and 335; for *Eco* σ^{70} and *Taq* σ^A, the substitutions are to the *Sau* σ^A residue found at these positions. Dark grey bars represent the interaction with the λCI-gp67 fusion protein and light grey bars the interaction with the λCI-*Sau* β-flap fusion protein to assess the structural integrity of the α-σ_4 fusion proteins. The results indicate that *Bsu* σ^A_4 (which has the same residues as *Sau* σ^A_4 at the four identified gp67 specificity determinants) binds gp67, but *Taq* σ^A_4 (which differs from *Sau* σ^A_4 at these four positions) fails to do so. The *Bsu* σ^A_4/gp67 interaction is abolished by replacing these four residues with the corresponding *Eco* σ^{70} residues, as observed with the *Sau* σ^A_4/gp67 interaction. **(Legend continued on next page.)**

Figure 2.10 (continued) The ability of all α-σ$_4$ fusion proteins to interact with the λCI-*Sau* β-flap fusion protein indicates that these proteins are properly folded and stably produced, and that the lack of interaction with gp67 observed for α-*Sau*(*Eco* 553, 556, 557, 579) and α-*Bsu*(*Eco* 553, 556, 557, 579) is not due to misfolding of the α-σ$_4$ fusion proteins. **(B)** One-hybrid interactions of the substituted α-σ$_4$ fusion proteins with an ectopically positioned -35 element, a second test for functionality. As with the two-hybrid interactions with the λCI-*Sau* β-flap fusion protein, the productive interactions of all α-σ$_4$ fusion proteins with the -35 element indicate that they are properly folded and functional.

Identifying _Sau_ σ^A_4-binding determinants in gp67

      To complement our genetic studies of the _Sau_ σ^A_4/gp67 interface, we were also

interested in identifying gp67 residues involved in specifying its binding to _Sau_ σ^A_4, with the goal

of generating a version of gp67 that would be stably folded but unable to bind _Sau_ σ^A_4 (and,

correspondingly, the modified versions of _Eco_ σ^{70}_4). A nonbinding variant of gp67 would serve

as a useful control in subsequent _in vivo_ and _in vitro_ investigations of gp67-mediated

transcription inhibition, allowing us to identify inhibitory effects that are strictly dependent on the

ability of gp67 to bind _Sau_ σ^A_4 (or an appropriately modified _Eco_ σ^{70}_4). An unbiased, random

mutagenesis screening approach to identify gp67 residues that are part of the interface with _Sau_

σ^A_4 is hindered by the lack of an effective counterscreen that would allow us to quickly assess

the structural integrity of the large number of gp67 mutants that would be generated with this

approach: the interaction with _Sau_ σ^A_4 is the only protein/protein contact involving gp67 we have

been able to detect using the two-hybrid system, and the α-gp67 fusion protein does not bind

the -35 element in the one-hybrid system (data not shown). Thus, we used the crystal structure

of the _Sau_ σ^A_4/gp67 complex to identify candidate _Sau_ σ^A_4-binding determinants on gp67 and

then used genetic assays to assess the effects of altering the identified residues.

      In the structure, two charged residues in gp67 (K2 and K195) form salt bridges with _Sau_

σ^A E312, one of the residues in the specificity-determining region identified as important for

gp67 binding by genetic analyses. We wanted to evaluate the functional importance of these

interactions by attempting to construct a mutant-suppressor pair by reversing the charges of

these residues. In particular, we reasoned that charge reversal(s) at either gp67 residues K2

and K195 or _Sau_ σ^A residue E312 would abolish the interaction between _Sau_ σ^A_4 and gp67 and

that the complementary charge reversal(s) in the partner protein would restore the interaction

between the two proteins. The ability of a gp67 double mutant bearing charge reversals at

residues K2 and K195 to interact with _Sau_ σ^A_4 bearing a charge reversal at residue E312 would

indicate that the substitutions in gp67 specifically abolished Sau σ^A_4 binding but did not disrupt

the proper folding of the gp67 protein, providing us with an otherwise functional nonbinding

variant of gp67 that would be unable to inhibit Sau σ^A-dependent transcription. We constructed

the α-Sau σ^A_4 E312K, λCI-gp67 K2D K195D, and λCI-gp67 K2E K195E fusion proteins and

tested their interactions using the two-hybrid system (again using the interaction with the λCI-

Sau β-flap fusion protein and the ectopically positioned -35 element to assess the structural

integrity of the α-Sau σ^A_4 E312K fusion protein). As seen in **Figure 2.11A**, the charge reversal

at Sau σ^A_4 residue 312 severely disrupted the binding of wild-type gp67 (dark grey bar) but had

no effect on Sau β-flap- or -35 element binding (**Figure 2.11B**, light grey and checkered bars,

respectively), confirming that Sau σ^A residue E312 is involved in gp67 binding. Tandem charge

reversals at residues K2 and K195 of gp67 (to either D or E) abolished the two-hybrid

interaction with the wild-type α-Sau σ^A_4 fusion protein (as expected if these gp67 residues lie at

the interface with Sau σ^A_4), but the complementary charge reversal at Sau σ^A residue E312

failed to restore the interaction (**Figure 2.11A**, striped bars). To determine whether the failure of

the gp67 variants to bind Sau σ^A_4 bearing the E312K might be the result of protein misfolding,

we used Western immunoblotting to assess the structural integrity of the λCI-gp67 fusion

proteins. We observed a difference in electrophoretic mobility between the wild-type and

charge-reversal versions of gp67 (**Figure 2.11C**, left panel), which suggests that the λCI-gp67

variants fold differently than wild-type λCI-gp67, perhaps in a conformation incompatible with

binding to the α-Sau σ^A_4 fusion proteins in the two-hybrid system.

      The Sau σ^A_4/gp67 co-crystal structure also identified several residues in gp67 (Y37, S40,

N41, Y43, and N188) that form a pocket into which the side chain of Sau σ^A residue R310

(which is part of the identified specificity-determining region but is conserved between Sau σ^A

and Eco σ^{70}) is tucked upon binding of gp67 to Sau σ^A_4. We wanted to determine whether
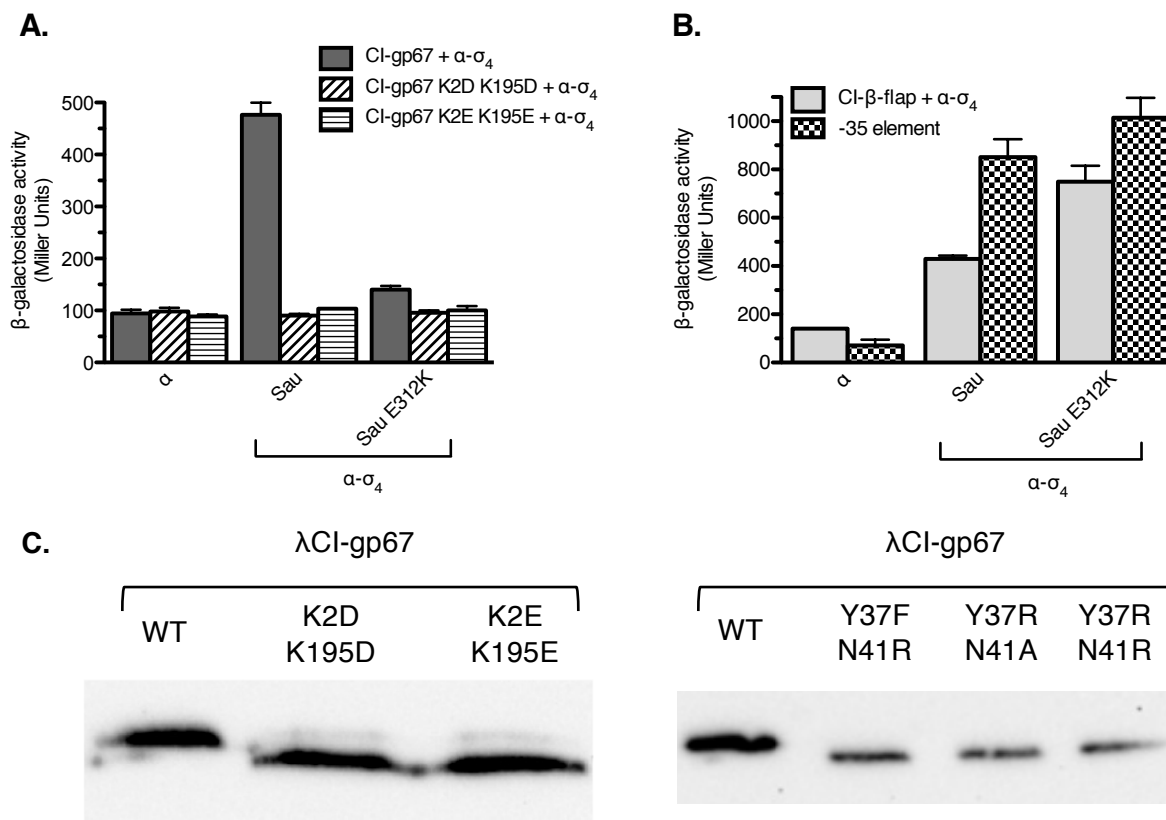
Figure 2.11. Substitutions at gp67 residues structurally predicted to contact *Sau* σ$^A_4$ affect the folding of λCI-gp67 fusion proteins. β-galactosidase activity is reported in Miller Units. Each bar represents the average Miller Units of three biological replicates assayed in the same experiment; error bars are ± standard deviation (SD). **(A)** Two-hybrid interactions of a candidate mutant-suppressor pair bearing charge reversal substitutions at *Sau* σ^A residue E312 and gp67 residues K2 and K195. Charge reversal substitutions were introduced at *Sau* σ^A E312 and gp67 K2 and K195 to determine whether the charge reversal(s) in one partner abolished the *Sau* σ$^A_4$/gp67 interaction, and whether the interaction was restored by the complementary charge reversal(s) in the other partner. The data show that the E312K substitution in *Sau* σ$^A_4$ severely impairs binding of wild-type gp67, but that this binding is not restored by charge reversals at gp67 residues K2 and K195. **(Legend continued on next page.)**

Figure 2.11 (continued) (B) Two-hybrid and one-hybrid interactions of the α-*Sau* σ^A_4 E312K

fusion protein with the λCI-*Sau* β-flap fusion protein (light grey bars) and the -35 element

(checkered bars) to assess the structural integrity of the fusion protein. α-*Sau* σ^A_4 E312K can

engage in productive interactions with both the β-flap and the -35 element, indicating that it is

properly folded and functional, and that the deficiency in gp67 binding is the result of disrupting

one of the crucial *Sau* σ^A_4/gp67 contacts. **(C)** Western blot analysis using an anti-λCI antibody to

assess the structural integrity of various λCI-gp67 fusion proteins bearing substitutions that

abolished the interaction with wild-type α-*Sau* σ^A_4 in the bacterial two-hybrid system. Wild-type

λCI-gp67 has the expected electrophoretic mobility of a protein of the appropriate molecular

weight (~50 kDa), but all λCI-gp67 substitution mutants migrate faster, consistent with a smaller

protein (despite all having the same number of residues) or an altered conformation. There is no

evidence that the mutant λCI-gp67 fusion proteins are undergoing degradation, suggesting that

these gp67 variants fold stably but in an altered conformation that may render them

incompatible for binding the α-*Sau* σ^A_4 fusion protein in the two-hybrid system.

substitutions at these gp67 residues would weaken or abolish binding to *Sau* σ^A_4 without

affecting the stability of gp67, so we constructed λCI-gp67 Y37F N41R, λCI-gp67 Y37R N41A,

and λCI-gp67 Y37R N41R fusion proteins and tested their interaction with the native α-*Sau* σ^A_4

fusion protein in the two-hybrid system. We again used Western blot analysis to assess the

structural integrity of the substituted λCI-gp67 fusion proteins. Like with the substitutions at K2

and K195, we found that all tested substitutions at gp67 residues Y37 and N41 abolished the

two-hybrid interaction with the α-*Sau* σ^A_4 fusion protein (data not shown), but we also observed

the same altered electrophoretic mobility of the λCI-gp67 variants, suggesting that these

proteins fold differently than wild-type λCI-gp67 (**Figure 2.11C**, right panel). Since we could not

evaluate whether the substitutions that were introduced into gp67 directly affected contacts at

the gp67/*Sau* σ^A_4 interface or affected binding indirectly due to altered protein folding, we

abandoned our attempts to use genetic methods to identify (or verify) *Sau* σ^A_4-binding

determinants on gp67.

The *Sau* σ^A_4/gp67 complex binds the *Sau* β-flap at the flap-tip helix

Biochemical and crystallographic data indicate that gp67 forms a ternary complex with

the σ^A-containing *Sau* holoenzyme (E$_{Sau}$•σ^A) and that the binding of gp67 to *Sau* σ^A_4 does not

induce conformational rearrangements in *Sau* σ^A_4 [6, 37]. Further, superimposing the structure

of σ^A_4 bound to the β-flap from the *Taq* holoenzyme structure [32] onto the *Sau* σ^A_4/gp67

structure reveals no steric clashes between the ten-residue β-flap-tip helix (FTH, which provides

the primary interaction determinants for the attachment of σ_4) and gp67 bound to *Sau* σ^A_4 [37],

suggesting that the binding of *Sau* σ^A_4 to the β-flap is not disrupted in the presence of gp67.

This further suggests that the formation of the gp67/E$_{Sau}$•σ^A ternary complex is likely mediated

primarily by the interaction of gp67 with *Sau* σ^A_4 engaged in its normal contacts with core RNAP,

rather than by the interaction of gp67 with specific determinants on core RNAP. [It is unclear

whether gp67 binds to *Sau* σ^A via σ_4 first and then the *Sau* σ^A_4/gp67 complex binds to core

RNAP to form the gp67-containing holoenzyme, or whether $E_{Sau} \cdot \sigma^A$ forms first and then gp67

binds to the holoenzyme through its interaction with *Sau* σ^A_4]. To test the structural prediction

that *Sau* σ^A_4 complexed with gp67 can bind the β-flap *in vivo*, we used a modified version of the

two-hybrid system adapted to detect bridging interactions [49]. We have previously found that

gp67 fails to interact detectably with the β-flap in the two-hybrid system but know that gp67

binds strongly to *Sau* σ^A_4 and that *Sau* σ^A_4 interacts with the β-flap. Thus, we asked whether

unfused *Sau* σ^A_4 could serve as a protein bridge between the gp67 moiety of an α-gp67 fusion

protein and the β-flap moiety of a λCI-*Sau* β-flap fusion protein (see **Figure 2.12A** for a

schematic of this bridging two-hybrid setup). If *Sau* σ^A_4 complexed with gp67 can bind to the β-

flap, as predicted by structural modeling, we would expect to see bridging and significant

transcription activation. As seen in **Figure 2.12B**, the presence of unfused *Sau* σ^A_4 results in

significant transcription activation in this assay (dark grey bars), indicating that the *Sau* σ^A_4/gp67

complex binds the β-flap in a FTH-dependent manner. These results provide strong support for

the structural prediction that the *Sau* σ^A_4/gp67 interaction is compatible with the *Sau* σ^A_4/β-flap

interaction in the context of the holoenzyme and suggest further that, unlike AsiA, gp67 does not

inhibit $E_{Sau} \cdot \sigma^A$-dependent transcription by interfering with the σ_4/β-flap contacts that position σ_4

to functionally engage the promoter -35 element.

Gp67 affects -35 element recognition in the context of the one-hybrid system

      Several structural observations suggested that gp67 does not inhibit transcription simply

by interfering with -35 element recognition: 1) unlike AsiA (which inhibits *Eco* σ^{70}-dependent

transcription at all -10/-35 promoters), gp67 does not rearrange *Sau* σ^A_4 into a conformation

incompatible with -35 element binding; 2) the gp67-binding surface of *Sau* σ^A_4 is largely distinct

from the recognition helix implicated in -35 element recognition (see **Figure 2.8B**): only one of

A.
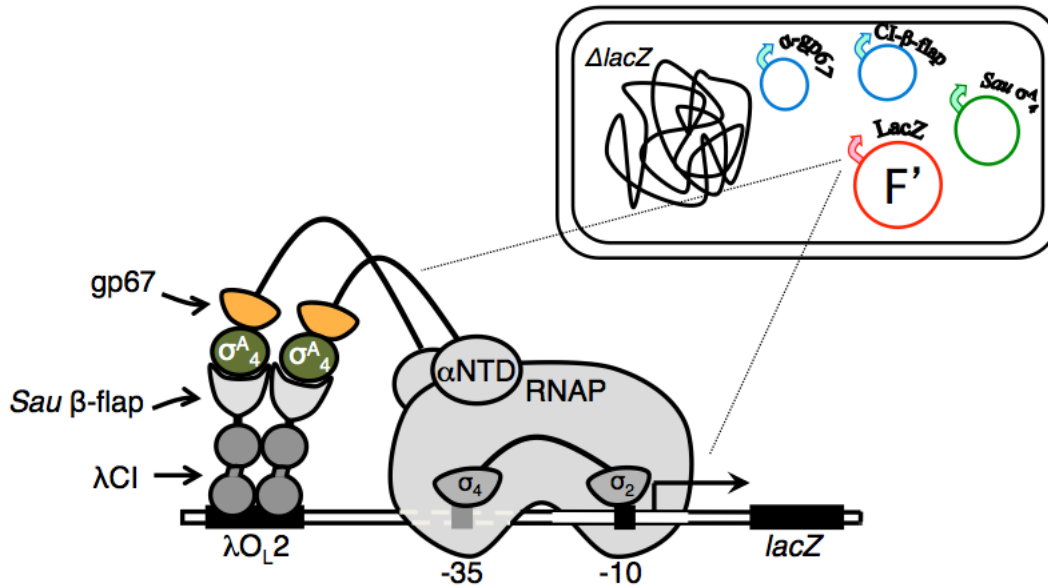
B.
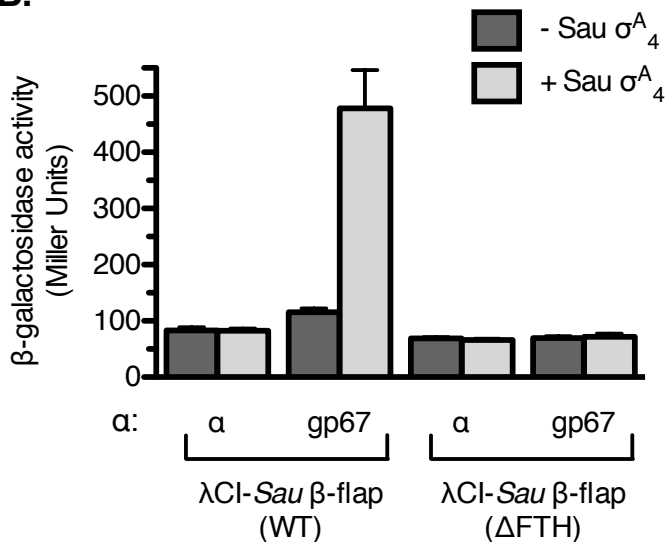
Figure 2.12. The *Sau* σ^A_4/gp67 complex binds the β-flap-tip helix *in vivo*. (A) Bacterial two-hybrid assay adapted to detect bridging interactions. The schematic depicts how unfused *Sau* σ^A_4 bridging an α-gp67 fusion protein and a λCI-*Sau* β-flap fusion protein would activate transcription from the $P_{lac}O_L2$-62 test promoter. (B) β-galactosidase activity is reported in Miller Units. Each bar represents the average Miller Units of three biological replicates assayed in the same experiment; error bars are ± standard deviation (SD). FW102 O_L2-62 cells contain three compatible plasmids: **(Legend continued on next page.)**

Figure 2.12 (continued) 1) a plasmid encoding α alone or the α-gp67 fusion protein; 2) a

plasmid encoding the λCI-*Sau* β-flap fusion protein, with either the complete flap region (WT) or

the flap-tip helix deleted (ΔFTH); and 3) a plasmid encoding either unfused, N-terminally His$_6$-

tagged *Sau* σ$^A_4$ (light grey bars) or no *Sau* σ$^A_4$ (dark grey bars). The results indicate that the α-

gp67/His$_6$-*Sau* σ$^A_4$ complex binds the *Sau* β-flap at the FTH, since the transcription activation

observed in the presence of His$_6$-*Sau* σ$^A_4$ is abolished when the FTH is deleted.

the ten highly-conserved DNA-binding residues of σ_4 (*Sau* σ^A R310, which corresponds to *Eco* σ^{70} R554) is part of the interface with gp67 (**Figure 2.13A**); and 3) superimposing the -35 element onto the *Sau* σ^A_4/gp67 structure reveals only minor steric clashes with the DNA in a region of gp67 that is likely conformationally flexible [37]. This is further supported by the observation that gp67 affects transcription from only ~9% of all *S. aureus* promoters *in vivo*, despite -10/-35 promoters constituting the major class of promoters [37]. We sought to directly test these structural predictions using a modified version of the bacterial one-hybrid system to assess the effect of producing unfused gp67 on the binding of an α-σ_4 fusion protein to the ectopically positioned -35 element. Unexpectedly, we observed that producing unfused gp67 resulted in a reduction in transcription activation from the test promoter specifically when the σ_4 moiety of the α-σ_4 fusion protein could bind gp67 (compare dark grey vs. light grey bars for *Sau* and *Eco* in **Figure 2.13B**). We investigated two possible explanations for this unexpected effect of gp67 on the binding of σ_4 to the -35 element.

First, when looking at the DNA-binding activity of σ_4 in isolation (as in the one-hybrid system), it is possible that the interaction of the σ_4 moiety of the α-σ_4 fusion protein and the ectopic -35 element is stabilized by a protein-protein interaction between σ_4 and the CTD of a native α subunit making nonspecific contacts with DNA upstream of the ectopic -35 element. (Such interactions could occur in the context of a heterodimeric holoenzyme species containing both a native α subunit and an α-σ_4 fusion protein.) If such stabilizing interactions occur, then the inhibitory effect of gp67 on the one-hybrid interaction of σ_4 with the -35 element might be explained by proposing that gp67 sterically occludes the binding of the αCTD just upstream of the ectopic -35 element, where it could interact with the DNA-bound σ_4 moiety. To determine whether αCTD/σ_4 contacts play a role in transcription activation in the one-hybrid system, we took advantage of a previously identified single amino acid substitution in *Eco* σ^{70}_4 that disrupts the *Eco* σ^{70}_4/αCTD interaction (*Eco* σ^{70} R603A [41], corresponding to *Sau* σ^A R359A) and
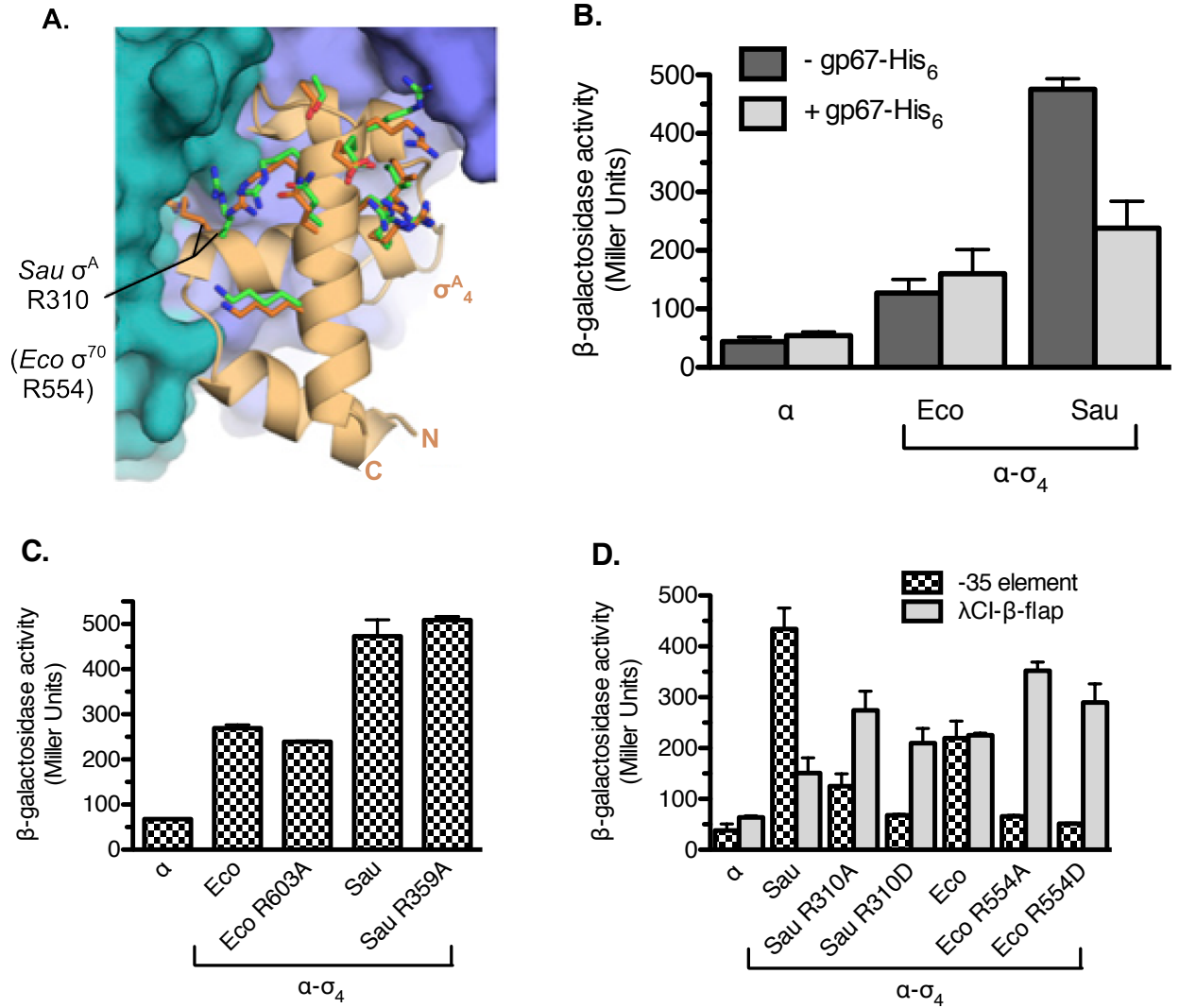
Figure 2.13. Effect of gp67 on -35 element binding. β-galactosidase activity is reported in Miller Units. Each bar represents the average Miller Units of three biological replicates assayed in the same experiment; error bars are ± standard deviation (SD). **(A)** View of *Sau* σ^A_4 (light orange ribbon) in the complex with gp67 (shown as a molecular surface: teal represents the gp67 NTD and blue the gp67 CTD) with the side chains of the ten residues of σ_4 that make direct contacts with -35 element DNA shown as sticks. The conformation of these side chains in the *Sau* σ^A_4/gp67 structure is shown in orange, and the conformation of these same residues from the superimposed *Taq* σ^A_4/-35 element DNA structure [2] is shown in green. **(Legend continued on next page.)**

Figure 2.13 (continued) Only the side chain of *Sau* σ^A residue R310 (corresponding to *Taq* σ^A R379 and *Eco* σ^{70} R554) is affected by gp67 binding. In the complex with -35 element DNA, this residue interacts with the DNA phosphate backbone at the position immediately upstream of the -35 element (termed '-36') [2, 21]. Upon gp67 binding, *Sau* σ^A R310 is redirected away from the DNA-binding interface and is buried in a pocket comprised of gp67 residues Y37, S40, N41, Y43, and N188. (Adapted from [37]). **(B)** Bacterial one-hybrid assay showing the effect of producing unfused gp67 on -35 element recognition by α-σ_4 fusion proteins. FW102 P$_{lac}$Cons-35C cells contain two compatible plasmids: 1) a plasmid encoding α alone or the α-σ_4 fusion protein, and 2) a plasmid encoding either unfused, C-terminally His$_6$-tagged gp67 (light grey bars) or no gp67 (dark grey bars). The results indicate that production of gp67 reduces transcription activation (-35 element binding) by the α-*Sau* σ^A_4 fusion protein, which can bind gp67, but not by the α-*Eco* σ^{70}_4 fusion protein, which cannot. **(C)** Bacterial one-hybrid assay to test whether binding of σ_4 to the -35 element is stabilized by an interaction with the αCTD. *Eco* σ^{70} residue 603 (corresponding to *Sau* σ^A residue 359), which has been implicated in an interaction between σ_4 and the αCTD [41], was mutated to test for an effect on transcription activation from the P$_{lac}$Cons-35C test promoter. The substitution (*Eco* R603A or *Sau* R359A) has no effect on transcription activation by either α-*Eco* σ^{70}_4 or α-*Sau* σ^A_4, which indicates that transcription activation in this one-hybrid system does not depend on stabilizing contacts between the αCTD and the σ_4 moiety of the α-σ_4 fusion protein. **(D)** One-hybrid and two-hybrid interactions of *Eco* and *Sau* α-σ_4 fusion proteins bearing substitutions at *Eco* σ^{70} residue R554 (*Sau* σ^A residue 310) with the -35 element (checkered bars) and the λCI-*Sau* β-flap fusion protein (light grey bars; to assess structural integrity). The substitutions at *Eco* σ^{70} residue R554 (*Sau* σ^A residue 310) abolish transcription activation in the one-hybrid system but have no effect on the two-hybrid interaction with the β-flap. **(Legend continued on next page.)**

Figure 2.13 (continued) These data indicate that the α-σ$_4$ fusion proteins are properly folded and that transcription activation in the one-hybrid system is dependent on the contact between *Eco* σ^{70} residue R554 (*Sau* σ^A residue 310) and the phosphate backbone at position -36.

introduced these substitutions into the corresponding α-σ_4 fusion proteins. If the transcription activation we observe in the one-hybrid system is the result of a σ_4/-35 element interaction stabilized by an αCTD/σ_4 contact, the α-σ_4 fusion proteins bearing the substitutions would be deficient in transcription activation from this reporter. As seen in **Figure 2.13C**, the *Eco* σ^{70} R603A and *Sau* σ^A R359A substitutions had no effect on transcription activation in the one-hybrid system, relative to the wild-type versions of the α-σ_4 fusion proteins. These data strongly suggest that αCTD/σ_4 contacts do not contribute to the transcription activation detected in the one-hybrid system, and, thus, that the (potential) disruption of these contacts by gp67 binding to *Sau* σ^A_4 cannot account for the unexpected reduction in transcription activation observed in the presence of unfused gp67.

Second, it is possible that the contact between residue R554 of *Eco* σ^{70} (corresponding to *Sau* σ^A R310) and the phosphate backbone at the -36 position (see legend of **Figure 2.13A**) is important for efficient initiation at many -10/-35 promoters (though the degree of importance may vary across promoters and bacterial species), and that this contact contributes significantly to the DNA-binding activity of σ_4 when examined in isolation in the one-hybrid system. If the binding energy provided by the *Eco* σ^{70} R554 (*Sau* σ^A R310)/-36 backbone contact is important for efficient -35 element recognition and transcription initiation, particularly when looking at σ_4 in isolation, we would expect α-σ_4 fusion proteins bearing substitutions at residue R554 (for *Eco* σ^{70}_4 proteins) or R310 (for *Sau* σ^A_4 proteins) to be deficient in transcription activation in the one-hybrid system. As seen in **Figure 2.13D**, substitutions at *Sau* σ^A residue 310 and *Eco* σ^{70} residue R554 reduce -35 element recognition substantially in the one-hybrid system (checkered bars), but do not significantly impact the stability of the α-σ_4 fusion proteins as measured by the two-hybrid interaction with the λCI-*Sau* β-flap fusion protein (light grey bars). The defect in transcription activation of the α-*Sau* σ^A_4 fusion proteins bearing substitutions at residue R310 phenocopies the effect of unfused gp67 on the wild-type α-*Sau* σ^A_4 fusion protein/-35 element

interaction. This is consistent with -35 element recognition (transcription activation) in the one-hybrid system being dependent on the *Eco* σ^{70} R554 (*Sau* σ^A R310)/-36 backbone contact. This dependence, in turn, suggests that the unexpected inhibitory effect of unfused gp67 on the binding of the α-*Sau* σ^A_4 fusion protein to the -35 element is due to the loss of the contact between σ^A R310 and the phosphate backbone at the -36 position. [See the **Discussion** section for further treatment of the potential role of this backbone contact in transcription regulation].

Discussion

Detailed studies of bacteriophage-encoded transcription regulators have provided invaluable mechanistic insight into bacterial transcription and have generated tools to probe RNAP function at all steps of the transcription cycle. The gp67 protein encoded by phage G1 of the Gram-positive pathogen *S. aureus* is one recent uncharacterized example: an anti-σ factor specific to the primary σ factor of *S. aureus* (*Sau* σ^A) that has no characterized conserved domains or sequence or structural similarity to other anti-σ factors and may present a novel regulatory paradigm. The results reported in this work describe the genetic dissection of the interface between gp67 and its target, conserved region 4 of *Sau* σ^A. Gp67 binds tightly to *Sau* σ^A_4 but does not interact with *Eco* σ^{70}_4 despite the similarities in the amino acid sequences of the two σ factors in this region. We identified the region consisting of *Sau* σ^A residues 309-335, and specifically the side chains of the residues at positions 309 (aspartate), 312 (glutamate), 313 (asparagine), and 335 (valine), as important for determining the specificity of gp67 for its target. The primary σ factor from another species that bears the same residues at these positions (*Bsu* σ^A) was also found to be a target for gp67, while the primary σ factor from a second species that differed at these positions (*Taq* σ^A) failed to bind gp67. We further demonstrated that substitutions at these four residues alone were sufficient to switch the specificity of gp67 binding: *Eco* σ^{70}_4 bearing the *Sau* σ^A_4 amino acid side chains at these positions was fully

competent for gp67 binding, whilst the corresponding substitutions in *Sau* σ^A_4 (or *Bsu* σ^A_4) abolished gp67 binding. These genetic data are in agreement with the *Sau* σ^A_4/gp67 co-crystal structure [37].

An *E. coli*-based system for the study of gp67 function

Previous reports [6, 37] have demonstrated that gp67 forms a ternary complex with the *Sau* σ^A-containing holoenzyme ($E_{Sau} \cdot \sigma^A$) and that gp67 likely exerts its regulatory effects in the context of this ternary complex. However, gp67 fails to form the ternary complex with a hybrid holoenzyme comprised of *Eco* core RNAP and *Sau* σ^A ($E_{Eco} \cdot \sigma^A$) [37]. Thus, functional studies of gp67 necessitate the use of native *S. aureus* systems *in vivo* and *in vitro*, despite the poorer biochemical characterization of *Sau* RNAP and promoters and the dearth of genetic and biochemical tools in *Sau* for in-depth mechanistic studies of transcription. Our ability to make *Eco* σ^{70} sensitive to gp67 by introducing only four substitutions in σ_4 that do not significantly affect σ^{70} function raises the intriguing possibility of developing a native *E. coli* system that uses this modified σ^{70} to study the mechanism of action of gp67. Such a system would enable us to take advantage of the many genetic and biochemical tools already available in *E. coli* (e.g. the commercial availability of wild-type *Eco* core RNAP, the existence of *Eco* RNAPs bearing modifications in several subunits, characterized mutations in promoters that affect activity and regulation *in vivo* and *in vitro*, etc.) and its greater experimental tractability to thoroughly investigate gp67-mediated transcription inhibition. In order to assess the usefulness of an *Eco*-based system to study gp67, we need to demonstrate that the *Eco* system recapitulates the key behaviors of gp67 in its native *Sau* host, such as gp67-mediated growth inhibition and the formation of a ternary complex with $E_{Eco} \cdot \sigma^{70}_{modified}$. Experiments addressing these questions are presented in **Chapter 3** of this dissertation.

A role for gp67 in modulating -35 element recognition?

Structural modeling of the *Sau* σ^A_4/gp67 complex onto holoenzyme suggested that,

unlike what has been previously observed with other anti-σ factors like FlgM [reviewed in 3] and

AsiA [reviewed in 18, 19], gp67 binding to *Sau* σ^A_4 does not interfere with *Sau* σ^A_4 binding to

core RNAP at the ten-residue β-flap-tip helix (FTH) or -35 element recognition. We used

modified versions of our bacterial two-hybrid and one-hybrid genetic assays to directly test these

structural predictions. In agreement with structural modeling, we found that the *Sau* σ^A_4/gp67

complex bound the *Sau* β-flap in a FTH-dependent manner. While we were unable to

unequivocally exclude the possibility that *Sau* σ^A_4 binding to gp67 induces conformational

changes in gp67 that enable it to specifically and directly contact the β-flap, the wealth of

genetic [23], biochemical [11], and structural [30, 32, 47] data documenting the σ_4/β-flap FTH

interaction strongly suggests that the *Sau* σ^A_4 moiety of the *Sau* σ^A_4/gp67 complex mediates the

binding to core RNAP and ternary complex formation.

Using the one-hybrid system, we found that—contrary to structural modeling

predictions—unfused gp67 affected the recognition of an ectopically positioned consensus -35

element specifically by an α-*Sau* σ^A_4 fusion protein. We demonstrated that the transcription

activation in this one-hybrid system, which reports on the binding of σ_4 to the -35 element in

isolation, is dependent on an interaction between *Eco* σ^{70} residue R554 (corresponding to *Sau*

σ^A residue R310) and the DNA phosphate backbone at the position immediately upstream of the

-35 element (position -36). This residue is the only residue in σ_4 implicated in DNA recognition

that is also a part of the interface with gp67. It is possible that this σ_4/DNA backbone contact is

also important for efficient -35 element recognition and transcription initiation in the context of

holoenzyme (as opposed to σ_4 alone, as in the one-hybrid system), and that this importance

varies across different promoters. Recent work has shown that gp67 exhibits promoter

selectivity by inhibiting the function of the αCTDs in *Sau* [37], but since the binding of gp67 to σ_4

disrupts the *Eco* σ^{70} R554 (*Sau* σ^A R310)/-36 backbone contact, gp67 may also have αCTD-independent effects on -35 element recognition specifically at promoters where this contact plays a key role in efficient initiation. The importance (if any) of the σ$_4$/DNA backbone contact in transcription initiation may also vary by species. Potential αCTD-independent effects of gp67 on transcription are investigated using an *E. coli* system in **Chapter 3**.

Materials and Methods

Plasmids, Strains, and Growth Conditions

A complete list of the bacterial strains used in this chapter of the dissertation is provided in **Table A1.1** in **Appendix 1**. *E. coli* strain NEB5α F'IQ (New England Biolabs) was used as the recipient for all plasmid constructions outlined below.

A complete list of plasmids used in this chapter of the dissertation is provided in **Table A1.2** in **Appendix 1**. Plasmid pBRαLN, a derivative of the medium copy number plasmid pBR322, was used to make fusions to the α subunit [7]. It encodes residues 1-248 of α (consisting of the αNTD and linker) with a NotI restriction site immediately downstream of codon 248 of α and a BamHI restriction site adjacent to the NotI site. DNA fragments encoding protein domains to be used in the two-hybrid and one-hybrid assays were generated by PCR and cloned into pBRαLN as NotI/BamHI restriction fragments, introducing an extra base at the end of the NotI site to generate a linker of three alanine residues and maintain the proper frame for translation of the fusion protein. Plasmid pACλCI32, a derivative of the medium copy number plasmid pACYC184 was used to make fusions to λCI [7]. It encodes full-length λCI with a NotI restriction site immediately downstream of codon 237 of λCI and BstYI and BamHI restriction sites adjacent to the NotI site. DNA fragments encoding protein domains to be used in the two-hybrid assay were generated by PCR and cloned into pACλCI32 as NotI/BamHI restriction

fragments, again with an extra base after the NotI site to create the alanine linker and remain in-frame. The high copy number plasmid pCDF*lac*, a derivative of pCDF-1b (Novagen) modified by P. Deighan to introduce the *lacUV5* promoter upstream of the T7 promoter in pCDF-1b, was used to produce unfused *Sau* σ^A_4 and gp67. Plasmid pCDF*lac* bears a plasmid-encoded N-terminal His_6 tag with unique PmlI and KpnI restriction sites immediately downstream of the His_6 tag, unique BamHI and AvrII sites in the multiple cloning site, and an NdeI site upstream of the plasmid-encoded N-terminal His_6 tag. *Sau* σ^A_4 was amplified by PCR and cloned into pCDF*lac* as a KpnI/BamHI restriction fragment to generate His_6-*Sau* σ^A_4; gp67 was amplified by PCR with primers encoding a C-terminal His_6 tag and cloned into pCDF*lac* as an NdeI/AvrII restriction fragment to generate gp67-His_6. All strains bearing pBRαLN derivatives were grown at 37ºC in LB (broth and plates) supplemented with carbenicillin (100 μg/mL); strains bearing pACλCI32 derivatives were grown in LB supplemented with chloramphenicol (25 μg/mL); and strains bearing pCDF*lac* derivatives were grown in LB supplemented with spectinomycin (50 μg/mL).

β-Galactosidase Assays

For the bacterial two-hybrid assays, reporter strain FW102 O_L2-62 [34] was co-transformed with two compatible multicopy plasmids: one encoding either α alone or the indicated α-σ_4 fusion protein, and another encoding a λCI-gp67 fusion protein, a λCI-gp67 fusion protein bearing substitutions in residues predicted by structural analysis to make direct contacts with *Sau* σ^A_4, or a λCI-*Sau* β-flap fusion protein. These plasmids direct the synthesis of the fusion proteins (or α) under the control of an IPTG-inducible promoter. Individual transformants were selected on plates and three independent isolates of each were grown overnight in LB broth supplemented with kanamycin (50 μg/mL), carbenicillin (100 μg/mL), and chloramphenicol (25 μg/mL) in the absence of IPTG. Overnight cultures were grown at 37ºC using deep well microtiter plates and a microtiter plate incubator shaker set to 900 rpm.

Saturated overnight cultures were diluted 1:100 into fresh LB broth supplemented with

kanamycin, carbenicillin, chloramphenicol, and 20 µM IPTG in microtiter plates and grown to

mid-exponential phase (O.D.$_{600}$ 0.3-0.6). β-galactosidase assays were performed as described

using microtiter plates and a microtiter plate reader [45], and Miller Units calculated as

described [34, 45]. The results shown in all figures are the averages of the three independent

isolates with standard deviations. The two-hybrid assay modified to detect bridging interactions

[49] was performed as described for the two-hybrid assays with the following modifications: one

plasmid encoded either α alone or an α-gp67 fusion protein; the second plasmid encoded either

a λCI-*Sau* β-flap fusion protein or a λCI-*Sau* β-flap ΔFTH fusion protein; and a third compatible

multicopy plasmid encoded unfused N-terminally His$_6$-tagged *Sau* σ$^A_4$ or no protein under the

control of an IPTG-inducible promoter. Plates and cultures in this experiment were additionally

supplemented with spectinomycin (50 µg/mL).

For the bacterial one-hybrid assays to detect -35 element binding [34, 35], reporter strain

FW102 P$_{lac}$Cons-35C was transformed with a single multicopy plasmid encoding either α alone

or the indicated α-σ$_4$ fusion protein under the control of an IPTG-inducible promoter. Individual

transformants were selected on plates and three independent isolates of each were grown

overnight as in the two-hybrid experiments but using LB broth supplemented with kanamycin (50

µg/mL) and carbenicillin (100 µg/mL) in the absence of IPTG. Overnight cultures were diluted

1:100 and grown to mid-exponential phase in LB broth supplemented with kanamycin,

carbenicillin, and 20 µM IPTG. β-galactosidase assays and subsequent computations were

performed as described for the two-hybrid assays. The assay to assess the effect of unfused

gp67 on -35 element binding by σ$_4$ was performed as described for the one-hybrid assays with

the following modifications: a second compatible multicopy plasmid encoding either unfused C-

terminally His$_6$-tagged gp67 or no protein under the control of an IPTG-inducible promoter was

co-transformed into cells; plates and cultures were additionally supplemented with

spectinomycin (50 µg/mL); and saturated overnights were diluted 1:100 and grown to mid-

exponential phase in LB broth supplemented with kanamycin, carbenicillin, spectinomycin, and

5 µM IPTG.

Western Blot Analysis

Western blots were performed using the cell lysates utilized in the β-galactosidase

assays [45]. Briefly, 200 µL of mid-exponential phase cultures were lysed for 30 minutes using

the detergent PopCulture (Novagen) and lysozyme and equal amounts of sample (normalized to

O.D.$_{600}$) were electrophoresed through 12% SDS-polyacrylamide gels and electroblotted to

Hybond C nitrocellulose membrane (Amersham). Membranes were blocked in 5% dried milk in

PBS and then probed with a 1:10,000 dilution of the anti-λCI antibody (generous gift from J.

Beckwith). After incubation with a goat anti-rabbit secondary antibody (Cell Signaling), the

SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific) was used for the

chemiluminescent detection of the bound horseradish peroxidase conjugate.

References

1.   **Browning, D. F. and Busby, S. J.** 2004. The regulation of bacterial transcription initiation. Nat Rev Microbiol **2**(1): 57-65.

2.   **Campbell, E. A.*, et al.*** 2002. Structure of the bacterial RNA polymerase promoter specificity sigma subunit. Mol Cell **9**(3): 527-39.

3.   **Campbell, E. A., Westblade, L. F., and Darst, S. A.** 2008. Regulation of bacterial RNA polymerase sigma factor activity: a structural perspective. Curr Opin Microbiol **11**(2): 121-7.

4.   **Chadsey, M. S. and Hughes, K. T.** 2001. A multipartite interaction between *Salmonella* transcription factor sigma28 and its anti-sigma factor FlgM: implications for sigma28 holoenzyme destabilization through stepwise binding. J Mol Biol **306**(5): 915-29.

5.   **Chadsey, M. S., Karlinsey, J. E., and Hughes, K. T.** 1998. The flagellar anti-sigma factor FlgM actively dissociates *Salmonella typhimurium* sigma28 RNA polymerase holoenzyme. Genes Dev **12**(19): 3123-36.

6.   **Dehbi, M.*, et al.*** 2009. Inhibition of transcription in *Staphylococcus aureus* by a primary sigma factor-binding polypeptide from phage G1. J Bacteriol **191**(12): 3763-71.

7.   **Dove, S. L. and Hochschild, A.** 2004. A bacterial two-hybrid system based on transcription activation. Methods Mol Biol **261**: 231-46.

8.   **Dove, S. L., Joung, J. K., and Hochschild, A.** 1997. Activation of prokaryotic transcription through arbitrary protein-protein contacts. Nature **386**: 627-30.

9.   **Feklistov, A. and Darst, S. A.** 2009. Promoter recognition by bacterial alternative sigma factors: the price of high selectivity? Genes Dev **23**(20): 2371-5.

10.   **Feklistov, A. and Darst, S. A.** 2011. Structural basis for promoter-10 element recognition by the bacterial RNA polymerase sigma subunit. Cell **147**(6): 1257-69.

11.   **Geszvain, K.*, et al.*** 2004. A hydrophobic patch on the flap-tip helix of *E.coli* RNA polymerase mediates sigma 70 region 4 function. J Mol Biol **343**(3): 569-87.

12.   **Gregory, B. D.**, *et al.* 2005. An altered-specificity DNA-binding mutant of *Escherichia coli* sigma70 facilitates the analysis of sigma70 function in vivo. Mol Microbiol **56**(5): 1208-19.

13.   **Gregory, B. D.**, *et al.* 2004. A regulator that inhibits transcription by targeting an intersubunit interaction of the RNA polymerase holoenzyme. Proc Natl Acad Sci U S A **101**(13): 4554-9.

14.   **Grigorova, I. L.**, *et al.* 2006. Insights into transcriptional regulation and sigma competition from an equilibrium model of RNA polymerase binding to DNA. Proc Natl Acad Sci U S A **103**(14): 5332-7.

15.   **Gross, C. A.**, *et al.* 1998. The functional and regulatory roles of sigma factors in transcription. Cold Spring Harb Symp Quant Biol **63**: 141-55.

16.   **Gruber, T. M. and Gross, C. A.** 2003. Multiple sigma subunits and the partitioning of bacterial transcription space. Annu Rev Microbiol **57**: 441-66.

17.   **Helmann, J. D.** 1999. Anti-sigma factors. Curr Opin Microbiol **2**(2): 135-41.

18.   **Hinton, D. M.** 2005. Molecular gymnastics: distortion of an RNA polymerase sigma factor. Trends Microbiol **13**(4): 140-3.

19.   **Hinton, D. M.**, *et al.* 2005. Transcriptional takeover by sigma appropriation: remodelling of the sigma 70 subunit of *Escherichia coli* RNA plymerase by the bacteriophage T4 activator MotA and co-activator AsiA. Microbiology **151**: 1729-40.

20.   **Hughes, K. T. and Mathee, K.** 1998. The anti-sigma factors. Annu Rev Microbiol **52**: 231-86.

21.   **Jain, D.**, *et al.* 2004. Structure of a ternary transcription activation complex. Mol Cell **13**(1): 45-53.

22.   **Jishage, M.**, *et al.* 1996. Regulation of RNA polymerase sigma subunit synthesis in *Escherichia coli*: intracellular levels of four species of sigma subunit under various growth conditions. J Bacteriol **178**(18): 5447-51.

23.   **Kuznedelov, K.**, *et al.* 2002. A role for interaction of the RNA polymerase flap domain with the sigma subunit in promoter recognition. Science **295**: 855-7.

24.   **Kwan, T.***, et al.* 2005. The complete genomes and proteomes of 27 *Staphylococcus aureus* bacteriophages. Proc Natl Acad Sci U S A **102**(14): 5174-9.

25.   **Lambert, L. J.***, et al.* 2004. T4 AsiA blocks DNA recognition by remodeling sigma70 region 4. EMBO J **23**(15): 2952-62.

26.   **Liu, J.***, et al.* 2004. Antimicrobial drug discovery through bacteriophage genomics. Nat Biotechnol **22**(2): 185-91.

27.   **Lonetto, M., Gribskov, M., and Gross, C. A.** 1992. The sigma 70 family: sequence conservation and evolutionary relationships. J Bacteriol **174**(12): 3843-9.

28.   **Maeda, H., Fujita, N., and Ishihama, A.** 2000. Competition among seven *Escherichia coli* sigma subunits: relative binding affinities to the core RNA polymerase. Nucleic Acids Res **28**(18): 3497-503.

29.   **Marr, M. T. and Roberts, J. W.** 1997. Promoter recognition as measured by binding of polymerase to nontemplate strand oligonucleotide. Science **276**: 1258-60.

30.   **Murakami, K. S.** 2013. X-ray crystal structure of Escherichia coli RNA polymerase sigma70 holoenzyme. J Biol Chem **288**(13): 9126-34.

31.   **Murakami, K. S. and Darst, S. A.** 2003. Bacterial RNA polymerases: the wholo story. Curr Opin Struct Biol **13**: 31-9.

32.   **Murakami, K. S., Masuda, S., and Darst, S. A.** 2002. Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 A resolution. Science **296**: 1280-4.

33.   **Nechaev, S. and Severinov, K.** 2003. Bacteriophage-induced modifications of host RNA polymerase. Annu Rev Microbiol **57**: 301-22.

34.   **Nickels, B. E.** 2009. Genetic assays to define and characterize protein-protein interactions involved in gene regulation. Methods **47**: 53-62.

35.   **Nickels, B. E.***, et al.* 2002. Protein-protein and protein-DNA interactions of sigma70 region 4 involved in transcription activation by lambda cl. J Mol Biol **324**: 17-34.

36.   **Nickels, B. E.*, et al.*** 2005. The interaction between sigma70 and the beta-flap of *Escherichia coli* RNA polymerase inhibits extension of nascent RNA during early elongation. Proc Natl Acad Sci U S A **102**(12): 4488-93.

37.   **Osmundson, J.*, et al.*** 2012. Promoter-specific transcription inhibition in *Staphylococcus aureus* by a phage protein. Cell **151**(5): 1005-16.

38.   **Paget, M. S. and Helmann, J. D.** 2003. The sigma70 family of sigma factors. Genome Biol **4**(1): 203.

39.   **Panaghie, G.*, et al.*** 2000. Aromatic amino acids in region 2.3 of *Escherichia coli* sigma 70 participate collectively in the formation of an RNA polymerase-promoter open complex. J Mol Biol **299**(5): 1217-30.

40.   **Record, M. T., Jr.*,** *et al.*, *Escherichia coli RNA polymerase (Eσ70), promoters, and the kinetics of the steps of transcription initiation* in *Escherichia coli and Salmonella: Cellular and Molecular Biology (2nd edn)*, Neidhardt, F. C., et al., Editors. 1996, American Society for Microbiology: Washington, DC. p. 792-820.

41.   **Ross, W.*, et al.*** 2003. An intersubunit contact stimulating transcription initiation by *E coli* RNA polymerase: interaction of the alpha C-terminal domain and sigma region 4. Genes Dev **17**(10): 1293-307.

42.   **Severinova, E., Severinov, K., and Darst, S. A.** 1998. Inhibition of *Escherichia coli* RNA polymerase by bacteriophage T4 AsiA. J Mol Biol **279**(1): 9-18.

43.   **Simeonov, M. F.*, et al.*** 2003. Characterization of the interactions between the bacteriophage T4 AsiA protein and RNA polymerase. Biochemistry **42**(25): 7717-26.

44.   **Somerville, G. A. and Proctor, R. A.** 2009. At the crossroads of bacterial metabolism and virulence factor synthesis in Staphylococci. Microbiol Mol Biol Rev **73**: 233-48.

45.   **Thibodeau, S. A., Fang, R., and Joung, J. K.** 2004. High-throughput beta-galactosidase assay for bacterial cell-based reporter systems. Biotechniques **36**(3): 410-5.

46.   **Twist, K. A.*, et al.*** 2011. Crystal structure of the bacteriophage T4 late-transcription coactivator gp33 with the beta-subunit flap domain of *Escherichia coli* RNA polymerase. Proc Natl Acad Sci U S A **108**(50): 19961-6.

47.   **Vassylyev, D. G.*, et al.** 2002. Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. Nature **417**: 712-9.

48.   **Wosten, M. M.** 1998. Eubacterial sigma-factors. FEMS Microbiol Rev **22**(3): 127-50.

49.   **Yuan, A. H.*, et al.** 2008. Rsd family proteins make simultaneous interactions with regions 2 and 4 of the primary sigma factor. Mol Microbiol **70**(5): 1136-51.

50.   **Yuan, A. H. and Hochschild, A.** 2009. Direct activator/co-activator interaction is essential for bacteriophage T4 middle gene expression. Mol Microbiol **74**(4): 1018-30.

51.   **Yuan, A. H., Nickels, B. E., and Hochschild, A.** 2009. The bacteriophage T4 AsiA protein contacts the beta-flap domain of RNA polymerase. Proc Natl Acad Sci U S A **106**(16): 6597-602.

Chapter 3:

An *E. coli*-based system to study gp67-mediated transcription inhibition

Attributions

The development of an *E. coli*-based system to study gp67 function was undertaken to circumvent some of the technical challenges of working with the bacteriophage G1 native host, *S. aureus*, and to take advantage of the wealth of tools available in *E. coli* for detailed mechanistic studies of transcription. J. Osmundson performed the native gel shift analysis reported in this chapter, and I conducted all of the remaining experiments presented here with extensive troubleshooting guidance from W. Ross and technical assistance from P. Deighan for the *in vitro* transcription experiments. N. Nair constructed the *E. coli* σ^{70}-depletion strain utilized in this chapter to assess the effects of gp67 on growth *in vivo*. I wrote the text of this chapter, with editorial assistance from A. Hochschild. An abbreviated version of this chapter, combined with some data presented in Chapter 2, is currently in press at *Journal of Bacteriology* with the title "Phage-encoded inhibitor of *Staphylococcus aureus* transcription exerts context-dependent effects on promoter function in a modified *E. coli*-based transcription system".

Introduction

Transcription in bacteria is dependent on a single multisubunit core RNAP (subunit composition $\alpha_2\beta\beta'\omega$) that is competent for nucleotide synthesis but incapable of recognizing specific promoter sequences until it binds the dissociable promoter specificity factor, σ, to form the RNAP holoenzyme (subunit composition $\alpha_2\beta\beta'\omega\sigma$) [9, 22]. Bacterial genomes generally encode multiple σ factors with different promoter specificities that direct the transcription of various sets of genes in response to environmental and developmental cues: a single essential primary σ factor responsible for the bulk of transcription during exponential growth, and one or more alternative σ factors that direct the expression of specialized regulons [10, 26, 34]. σ^{70} is the primary σ factor of the Gram-negative model bacterium, *E. coli* (referred to as *Eco* σ^{70}), and σ^A is the primary σ factor of the related low-G+C Gram-positive bacteria, *S. aureus* and *B. subtilis* (referred to as *Sau* σ^A and *Bsu* σ^A, respectively).

Primary σ factors are modular proteins with four regions of conserved sequence ($\sigma_{1.1}$, σ_2, σ_3, and σ_4) that bind promoter DNA in the context of the RNAP holoenzyme, making extensive contacts with both core RNAP subunits and conserved promoter elements. Conserved region 2 (σ_2) binds a coiled-coil motif in the β' subunit (β'cc) and functionally engages the TATAAT hexamer that constitutes the -10 promoter element; conserved region 3.0 contacts the TG dinucleotide of the extended -10 promoter motif (TGnTATAAT); and conserved region 4 (σ_4) contacts a flexible flap domain in the β subunit (β-flap) to be appropriately positioned to engage the TTGACA -35 element hexamer [reviewed in 22]. Additionally, some promoters contain an A+T-rich sequence motif upstream of the -35 element called the UP element that is recognized not by the σ factor, but by the C-terminal domains of the α subunits of RNAP (αCTDs). Several promoters—the best characterized of which are the ribosomal RNA (*rrn*) promoters—depend on sequence-specific UP element/αCTD contacts for full activity, but the αCTDs can also engage in sequence-nonspecific contacts with upstream DNA that facilitate holoenzyme recruitment to

116

promoters lacking UP elements [31]. The αCTDs can also interact with activators bound

upstream of the core promoter to stimulate transcription initiation [reviewed in 2].

Much of the regulation of bacterial gene expression occurs at transcription initiation. The

integral roles of σ factors and the α subunits in this process make these subunits key regulatory

targets. Bacteria and their infecting bacteriophages have evolved a multitude of strategies to

modulate the activities of σ factors and the α subunits as a means to enact the appropriate

transcriptional program. In the case of phages, these modulations repurpose the host

transcription machinery to express viral genes and facilitate viral propagation. Anti-σ factors,

regulators that bind tightly and specifically to their cognate σ factor to inhibit transcription from

promoters dependent on that σ factor for initiation [reviewed in 5, 12, 14], are key players in the

regulation of transcription initiation employed by both bacteria and phages. Anti-σ factors

function by a variety of mechanisms; many (e.g. the *Eco*-encoded stationary phase factor Rsd)

function primarily by sequestering their cognate σ factor and preventing holoenzyme formation,

but others (e.g. the *Eco*-encoded anti-σ^{28} factor FlgM, or the coliphage T4-encoded anti-σ^{70}

AsiA) exert their inhibitory functions in the context of the holoenzyme [reviewed in 5]. The study

of phage-encoded transcription regulators (including anti-σ factors) has provided great

functional insight into the bacterial transcription apparatus, and the sequencing of a multitude of

phage genomes has facilitated the identification of novel regulators that may further enhance

our understanding of transcription and guide the development of new antimicrobials.

Phage G1 gp67 was identified as an inhibitor of growth and RNA synthesis in its native

host, *S. aureus* [16, 19], and was subsequently shown to interact directly with *Sau* σ^A via

conserved region 4 and form a stable ternary complex with the σ^A-containing *Sau* holoenzyme

($E_{Sau} \cdot \sigma^A$) [6, 25]. Osmundson *et al.* [25] then demonstrated that gp67 is not a general

transcription inhibitor but rather exhibits specificity for the *Sau* rRNA promoters and several

other promoters with A+T-rich UP element-like upstream regions. Expression analyses in *Sau*,

biochemical analyses with *Sau* RNAP and structure-based modeling suggested that the promoter selectivity of gp67 stems from its ability to interfere with the function of the αCTDs, specifically inhibiting transcription from promoters that require productive αCTD/UP element interactions for efficient initiation [25]. These data suggest that transcription regulation by gp67 presents a novel paradigm whereby the regulator becomes a stable component of the bacterial transcription machinery through its interaction with one RNAP subunit (the σ factor) but selectively affects the function of another subunit (α).

Gp67 does not interact with *Eco* σ^{70}, so the characterization of gp67 seemed dependent on the development of *Sau*-based *in vivo* and *in vitro* experimental systems despite the poorer characterization of *Sau* RNAP and promoters relative to their *Eco* counterparts. Based on the genetic results presented in **Chapter 2** of this dissertation, we can modify *Eco* σ^{70}_4 such that it binds gp67 in the context of the two-hybrid system. We reasoned that full-length *Eco* σ^{70} bearing these modifications in conserved region 4 would likewise be sensitive to gp67, and that this would enable us to develop an *Eco*-based system to investigate gp67 function that could take advantage of the wide array of tools and reagents available in *Eco* for detailed mechanistic studies of transcription. Here we report that gp67 inhibits growth in *Eco* when transcription of housekeeping genes is directed by a version of *Eco* σ^{70} that binds gp67, and that gp67 forms a stable ternary complex with holoenzyme reconstituted with *Eco* core RNAP and modified *Eco* σ^{70}. We describe the use of this gp67-responsive *Eco* RNAP holoenzyme and otherwise-identical promoters either bearing or lacking an UP element to uncover an important role for the αCTD in efficient transcription initiation in the absence of an UP element. We further report that gp67-mediated transcription inhibition can be either αCTD-dependent or αCTD-independent, depending on the promoter context. We also show that gp67 inhibits the growth of *B. subtilis*, which may facilitate the development of a less-heterologous *Bsu*-based system for further mechanistic studies of gp67.

Results

Gp67 inhibits growth of *E. coli* when transcription is directed by *Eco* σ^{70} modified to bind gp67

Gp67 was originally identified based on its ability to inhibit *Sau* growth [19], so we were interested in determining whether we could recapitulate this behavior in *Eco* by modifying *Eco* σ^{70} using the gp67 binding specificity determinants identified in **Chapter 2**. We reconstituted two variants of full-length *Eco* σ^{70} bearing modifications in conserved region 4: one variant bore *Sau* σ^{A} residues 309-335 (the full gp67 specificity-determining region) in place of *Eco* σ^{70} residues 553-579 (*Eco* σ^{70}_{hybrid}), and the other bore the corresponding *Sau* σ^{A} residues at five positions: *Eco* σ^{70} residues 553, 556, 557, 579, and 581 (*Eco* σ^{70}_{quint}). (The first four substitutions correspond to the identified gp67 specificity determinants; the substitution at residue 581 stabilizes the folding of *Eco* σ^{70}_{4} in the context of the α-σ_{4} fusion protein and was included for consistency with the genetic experiments presented in Chapter 2.) We assessed the functionality of these *Eco* σ^{70} variants using the σ^{70}-depletion strain NUN449, an MG1655 derivative in which the chromosomal wild-type *rpoD* is under the control of the repressible *trp* promoter at its native locus (derived from strain CAG20153, [20]). In this strain, under repressing conditions functional plasmid-encoded *Eco* σ^{70} must be provided to sustain growth. As seen in **Figure 3.1** (compare 'Empty vector' in panels B and C with panel A), both *Eco* σ^{70}_{hybrid} and *Eco* σ^{70}_{quint} fully complemented the depletion of wild-type chromosomally-encoded *Eco* σ^{70}, indicating that these *Eco* σ^{70} variants are fully functional. We then tested the effect on transcription of producing gp67 in σ^{70}-depletion strain cells complemented with either wild-type *Eco* σ^{70} or the *Eco* σ^{70} variants, using growth as the readout. We found that gp67 specifically inhibited *Eco* growth when the available primary σ factor bore the gp67 specificity determinants (**Figure 3.1**), suggesting that gp67 is functional in this modified *Eco* system.

A.

B.

C.
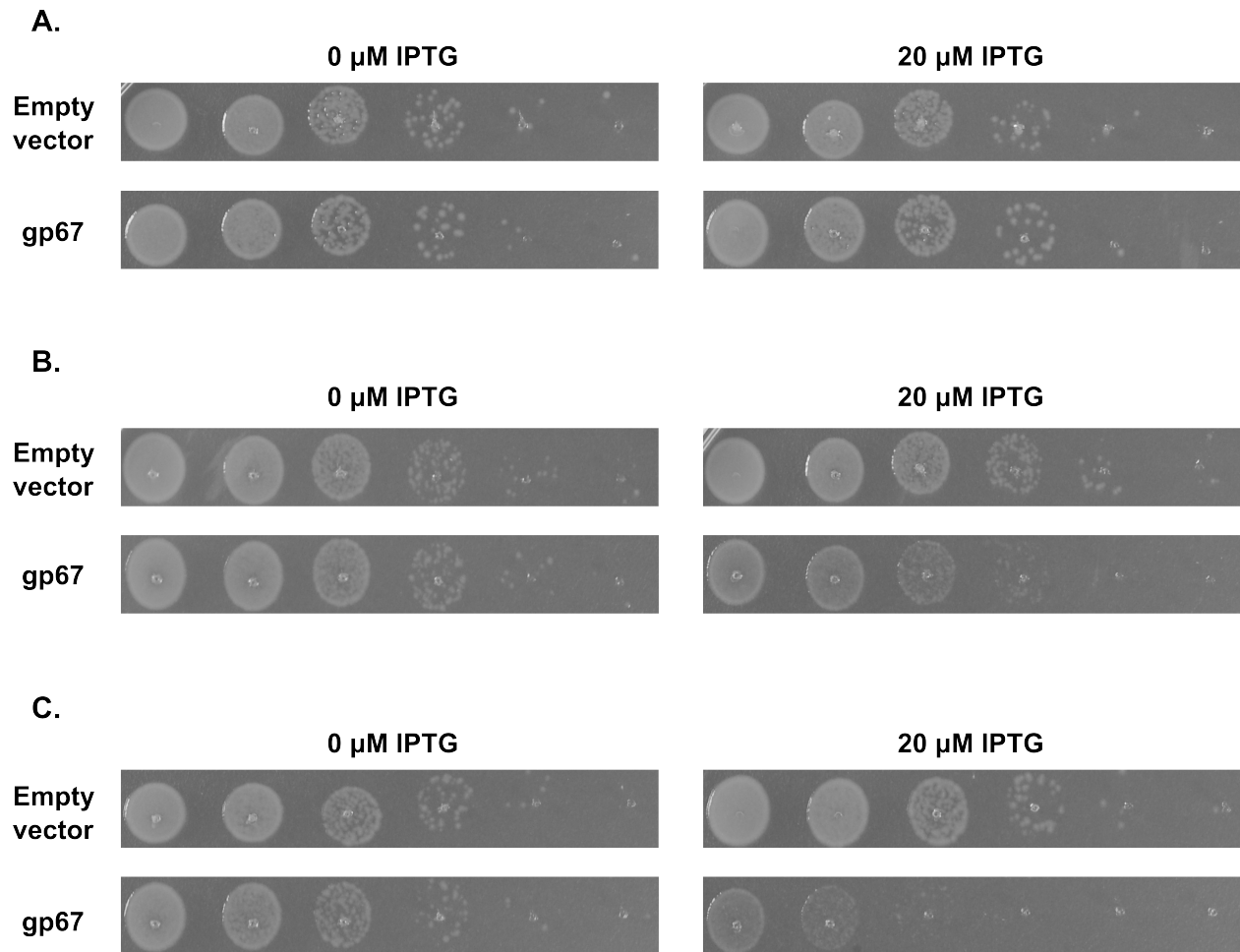
Figure 3.1. Gp67 inhibits *E. coli* growth when σ⁷⁰ bears the gp67 specificity determinants. Serial spot dilutions of the *E. coli* σ⁷⁰-depletion strain NUN449 co-transformed with two compatible plasmids: a plasmid encoding wild-type *Eco* σ^{70} (**A**), *Eco* σ^{70}_{hybrid} (**B**), or *Eco* σ^{70}_{quint} (**C**) under the control of a weak constitutive promoter; and a plasmid encoding either unfused, C-terminally His_6-tagged gp67 or no protein (empty vector) under the control of the IPTG-inducible *lacUV5* promoter. In NUN449, chromosomal *rpoD* is under the control of the repressible *trp* promoter. In the absence of indole-3-acrylic acid (IAA), an inducer of the *trp* promoter, chromosomal *rpoD* expression is repressed and functional plasmid-encoded *Eco* σ^{70} must be provided to sustain growth. **(Legend continued on next page.)**

Figure 3.1 (continued) Both *Eco* σ^{70}_{hybrid} and *Eco* σ^{70}_{quint} complement the depletion of wild-type, chromosomally-encoded *Eco* σ^{70} (panels **B** and **C**, respectively). Production of gp67 specifically inhibits growth when a modified version of *Eco* σ^{70} that bears the gp67 specificity determinants is used (panels **B** and **C**) but has no effect on wild-type *Eco* σ^{70} (panel **A**). These data suggest that gp67 can function as a transcription inhibitor in the context of the *E. coli* transcription machinery if *Eco* σ^{70}_{4} is modified such that it can bind gp67. The growth-inhibitory effect of gp67 is more pronounced at higher IPTG concentrations (not shown). Note that the chromosomal wild-type *rpoD* allele is repressed on this medium (no IAA present).
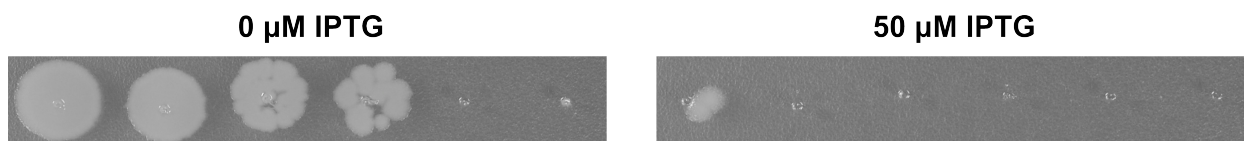
<u>Gp67 inhibits growth of unmodified *B. subtilis*</u>

Bsu σ^A bears the same residues as *Sau* σ^A at the positions corresponding to the gp67 specificity determinants, and we previously showed that *Bsu* σ^A_4 interacted strongly with gp67 in the two-hybrid system in a manner dependent on the identity of the four identified gp67 specificity determinants (see **Chapter 2**). This raised the possibility that *Bsu* could be used as a system that is more experimentally tractable than *Sau* but less heterologous than *Eco* for detailed mechanistic studies of gp67 function. As with the *Eco*-based system, we sought to recapitulate the key behaviors of gp67 reported in *Sau* using *Bsu*. We reasoned that producing gp67 in *Bsu* would inhibit *Bsu* σ^A-dependent transcription and growth, but without the added complexity of having to modify or deplete chromosomally-encoded *Bsu* σ^A. To address this, we constructed a derivative of *Bsu* strain PY79 with a single copy of *orf67* (encoding gp67-His$_6$) under the control of the IPTG-inducible strong *hyper-spank* promoter integrated at the *amyE* locus on the chromosome. As seen in **Figure 3.2**, the production of gp67 resulted in significant growth inhibition, suggesting that gp67 can bind *Bsu* σ^A_4 in the context of native full-length *Bsu* σ^A and inhibit *Bsu* σ^A-dependent transcription, as it does with *Sau* σ^A in its native host.

<u>Gp67 forms a ternary complex with *Eco* holoenzyme reconstituted with modified σ^{70}</u>

Gp67 becomes a stable component of the σ^A-containing *Sau* holoenzyme (E$_{Sau}$•σ^A) [6, 25] but, despite previous reports [6], cannot interact with a hybrid holoenzyme consisting of *Eco* core RNAP in complex with *Sau* σ^A (E$_{Eco}$•σ^A) [25]. Further, gp67 inhibits transcription in *Sau* in the context of this ternary complex [25], so a heterologous system for the study of gp67 must recapitulate this complex formation in order to be mechanistically informative. To determine whether gp67 could interact with holoenzyme consisting of *Eco* core RNAP in complex with *Eco* σ^{70} bearing the gp67 specificity determinants, J. Osmundson performed native gel shift assays using purified *Eco* core RNAP together with wild-type *Eco* σ^{70}, *Eco* σ^{70}_{hybrid}, or *Eco* σ^{70}_{quint}, and

0 μM IPTG   **50 μM IPTG**

Figure 3.2. Gp67 inhibits growth of *B. subtilis*. Serial spot dilutions of *B. subtilis* strain PY79

with a single copy of *orf67* under the control of the IPTG-inducible strong promoter *hyper-spank*

integrated at the *amyE* chromosomal locus. Production of gp67 results in significant growth

inhibition at all IPTG concentrations tested (50 μM, 100 μM, and 500 μM; only the data for

induction of *orf67* expression at 50 μM IPTG is presented), which suggests that gp67 inhibits

Bsu σ^A-dependent transcription and is functional in this experimentally tractable heterologous

system.

gp67. In this assay, ternary complex formation is visualized as a shift of the holoenzyme band in the presence of gp67. As seen in **Figure 3.3**, addition of gp67 resulted in a shift of the *Eco* holoenzyme band only when holoenzyme was reconstituted using a modified version of *Eco* σ^{70} bearing the gp67 specificity determinants, indicating that the functionally relevant gp67/holoenzyme ternary complex forms in this modified *Eco* system. We used *Eco* σ^{70}_{quint} for all subsequent functional analyses.

Gp67 exerts αCTD-dependent and αCTD-independent inhibitory effects in a promoter context-specific manner in an *E. coli*-based transcription system

Previous work using a native *Sau*-based transcription system revealed that gp67 selectively inhibited transcription from the UP element-containing *Sau* rRNA promoters and a number of other promoters with A+T-rich upstream regions proposed to function like UP elements, and that gp67 disrupted contacts between $E_{Sau} \cdot \sigma^A$ and upstream DNA [25]. UP elements and the rRNA (*rrn*) promoters have not been as extensively characterized in *Sau* as they have in *Eco*, where the UP element-containing *rrnB* P1 promoter has been particularly thoroughly studied, so we sought to use our gp67-responsive *Eco* transcription system to test whether gp67 exerted promoter (and UP element)-specific effects in this system. Conducting experiments in *Eco* allowed us to take advantage of otherwise-identical versions of *rrnB* P1 either containing (*rrnB* -61) or lacking (*rrnB* -41) its native UP element [30], as well as a version of *Eco* core RNAP lacking the αCTDs (αΔCTD RNAP core) [15] that is transcriptionally active *in vitro* but cannot discriminate between otherwise-identical promoters bearing or lacking an UP element. The use of holoenzyme reconstituted with αΔCTD core would allow us to directly assess the extent to which observed gp67 inhibitory effects are dependent on αCTD/upstream DNA contacts, something that is not presently technically feasible in the native *Sau* system.
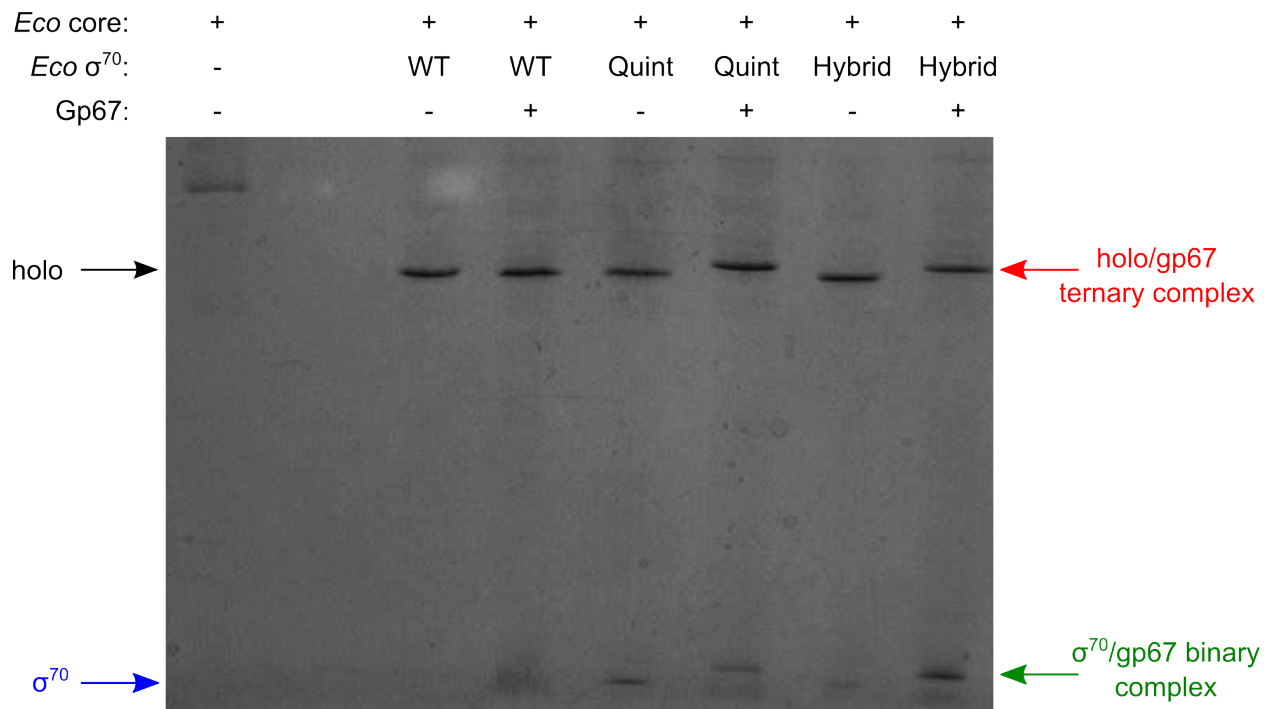
Figure 3.3. Gp67 forms a ternary complex with *Eco* holoenzyme reconstituted with σ^{70} bearing the gp67 specificity determinants. Native gel shift experiments were performed with purified *Eco* core and gp67 plus 1) wild-type *Eco* σ^{70} ('WT'); 2) *Eco* σ^{70} bearing the corresponding *Sau* residues at the four identified gp67 specificity determinants and at residue D581 ('Quint'); or 3) *Eco* σ^{70} bearing *Sau* σ^{A} residues 309-335 (the full gp67 specificity region) in place of *Eco* σ^{70} residues 553-579 ('Hybrid'). Holoenzyme (in the absence of gp67) has an electrophoretic mobility indicated by the black arrow. Ternary complex formation is visualized as a holoenzyme supershift upon addition of gp67 to the reaction and is observed only with the modified versions of *Eco* σ^{70} ('Hybrid' and 'Quint'; bands at the level of the red arrow). The modified versions of *Eco* σ^{70} also form a binary complex with gp67 (bands at level of green arrow), as expected from the two-hybrid data of Chapter 2 indicating that gp67 interacts with σ$_4$ of these versions of *Eco* σ^{70}.

In a first set of experiments, we assayed transcription *in vitro* using the supercoiled DNA templates *rrnB* -41 (no UP element) and *rrnB* -61 (native *rrnB* UP element); these templates contain the native *rrnB* terminators and specify a ~200-nucleotide (nt) transcript. These plasmids also include the RNA-I promoter within the replication origin, a -10/-35 promoter that does not bear an UP element and encodes a 108-nt transcript [21]. We conducted multi-round *in vitro* transcription reactions using either wild-type *Eco* core RNAP or *Eco* αΔCTD core RNAP and wild-type *Eco* σ^{70} or *Eco* σ^{70}_{quint} (represented as: E•σ^{70}, EαΔCTD•σ^{70}, E•σ^{70}_{quint}, and EαΔCTD•σ^{70}_{quint}) to assess the effect of gp67 on transcription by each of these holoenzymes from each of these promoters. As expected, since gp67 does not bind wild-type *Eco* σ^{70} or form a ternary complex with E•σ^{70}, gp67 had no effect on transcription by E•σ^{70} (dark grey bars) or EαΔCTD•σ^{70} (light grey bars) from either *rrnB* -41 (**Figure 3.4A**) or *rrnB* -61 (**Figure 3.4B**). There was also no effect of gp67 on transcription from the RNA-I promoter.

However, when we looked at the effect of gp67 on transcription by E•σ^{70}_{quint}, we saw significant inhibition of transcription at all promoters, irrespective of the presence or absence of the UP element (**Figure 3.5A** and **Figure 3.5B**, dark grey bars). Based on the effects observed with the native *Sau*-based transcription system, the inhibitory effect of gp67 on the promoters lacking UP elements (*rrnB* -41 and RNA-I) was unexpected, especially since gp67-mediated inhibition at *rrnB* -41 was somewhat greater than at *rrnB* -61 (up to 7.6-fold versus up to 4.9-fold). These data indicate that, in an *Eco*-based transcription system, gp67-mediated transcription inhibition is not strictly UP element-dependent.

We then assessed the extent to which the inhibitory effects of gp67 at these promoters depended on the αCTDs using EαΔCTD•σ^{70}_{quint}. We observed only a modest inhibitory effect of gp67 at the RNA-I promoter with EαΔCTD•σ^{70}_{quint} (≤1.9-fold; compare dark grey bars and light grey bars at RNA-I in panels A, B, and C of **Figure 3.5**), suggesting that inhibition at the RNA-I promoter is largely mediated by the disruption of αCTD/upstream DNA contacts by gp67 binding.
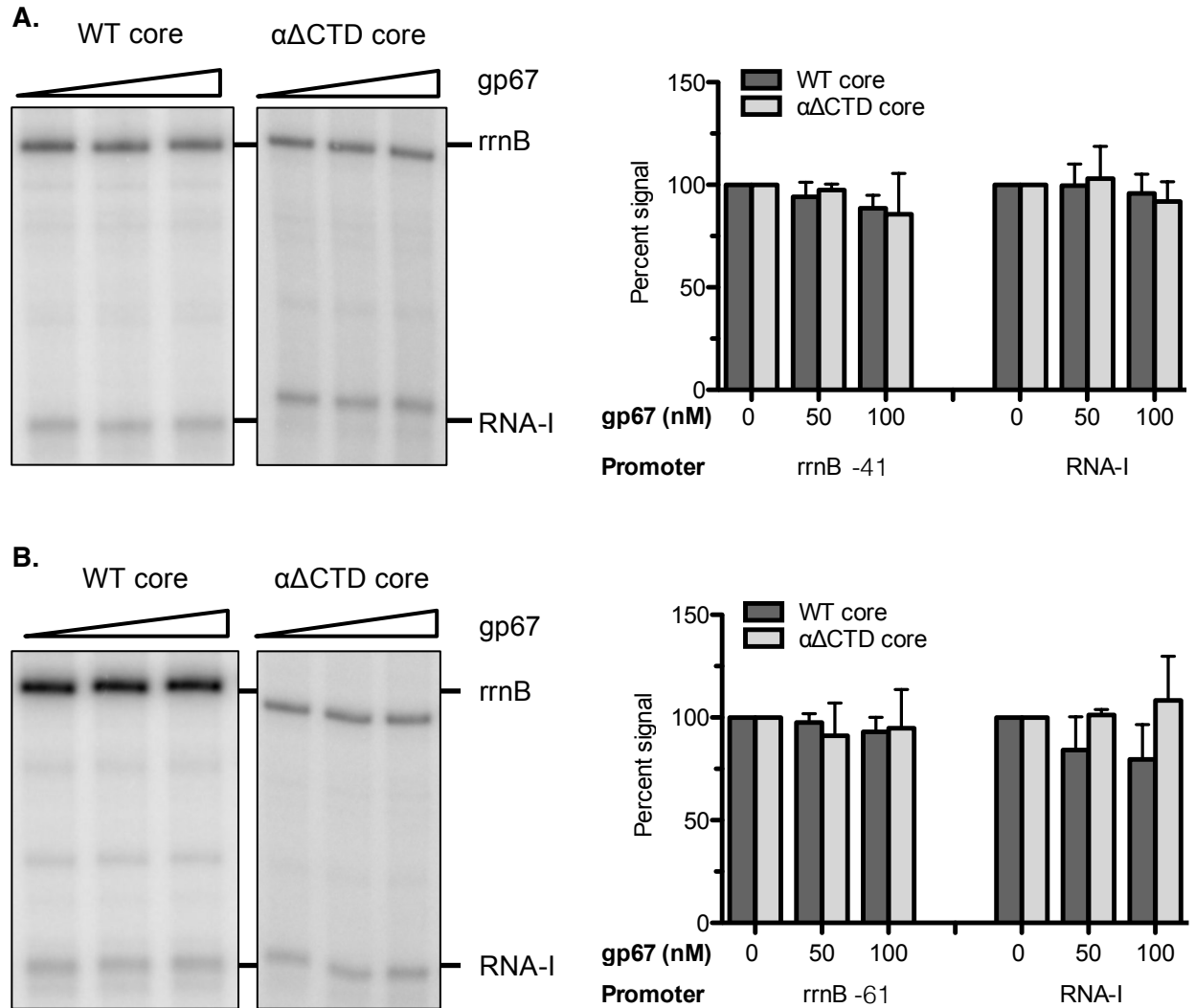
Figure 3.4. Gp67 does not inhibit transcription by holoenzyme reconstituted with wild-type *Eco* σ[70]. A representative gel is shown on the left side of each panel, with the type of *Eco* core RNAP (either wild-type [WT] or αΔCTD) used to reconstitute holoenzyme indicated. Transcripts corresponding to initiation at each of the promoters present on each plasmid template are labeled. The triangles represent increasing concentrations of purified gp67 used in the transcription reactions. At least three independent gels were quantified by phosphorimagery for each experiment. **(Legend continued on next page.)**

Figure 3.4 (continued) For each independent promoter, the amount of transcription (band intensity of corresponding product) in the absence of gp67 was set to 100%. Data reported on the right side of each panel represent the average ± standard deviation of band intensity for each condition for E•σ^{70} (dark grey bars) and EαΔCTD•σ^{70} (light grey bars). All supercoiled plasmid templates contain the RNA-I promoter within the plasmid replication origin. **(A)** *In vitro* transcription experiments using the supercoiled *rrnB* -41 template lacking the UP element. There is no effect of gp67 on transcription from the *rrnB* -41 or RNA-I promoters with either E•σ^{70} or EαΔCTD•σ^{70}. **(B)** *In vitro* transcription experiments using the supercoiled *rrnB* -61 template, which contains the native *rrnB* UP element. Removal of the αCTDs reduces transcription from the *rrnB* promoter on this template (see representative gel on left), consistent with the significant stimulatory effect of the UP element that depends on contacts with the αCTDs, but gp67 again has no effect on transcription.

Figure 3.5. Gp67 exerts inhibitory effects that are independent of the UP element but dependent on promoter context with holoenzyme reconstituted with *Eco* σ^{70}_{quint}. A representative gel is shown in each panel, and at least three independent gels were quantified by phosphorimagery for each experiment. For each independent promoter, the amount of transcription was normalized to that observed in the absence of gp67; data reported represent the average ± standard deviation of band intensity for each condition for $E \cdot \sigma^{70}_{quint}$ (dark grey bars) and $E\alpha\Delta CTD \cdot \sigma^{70}_{quint}$ (light grey bars). All supercoiled plasmid templates contain the RNA-I promoter within the plasmid replication origin. **(A)** Effect of gp67 on transcription from the *rrnB* -41 template. Gp67 significantly inhibits transcription by $E \cdot \sigma^{70}_{quint}$ at the *rrnB* P1 promoter without an UP element and at the RNA-I promoter (also lacking an UP element), indicating that there is no UP element dependence of gp67 inhibition in the *Eco* transcription system. Gp67-mediated transcription inhibition is largely independent of the αCTDs at the *rrnB* P1 promoter but seems to depend significantly on the αCTDs at the RNA-I promoter, suggesting that nonspecific contacts of the αCTDs with upstream DNA are important for initiation at this promoter. These effects are also observed with the *rrnB* -61 template in **(B)**, reinforcing the conclusion that gp67 can exert αCTD-independent inhibitory effects in the *Eco* transcription system. **(C)** Effect of gp67 on transcription from the strong phage λ $P_{R'}$ promoter bearing a consensus extended -10 promoter within its transcribed region (centered at +19). Gp67 has no effect on transcription from the extended -10 promoter, indicating that gp67 does not sequester σ^{70}_{quint} to interfere with holoenzyme formation. The inhibitory effect on the $P_{R'}$ promoter is modest and αCTD-independent (compare dark grey bars and light grey bars at '$P_{R'}$'). Taken together, these data suggest that gp67-mediated inhibition is specific to -10/-35 promoters but is not equivalent across all -10/-35 promoters: depending on the promoter context, inhibitory effects can be mediated by or be independent of contacts between the αCTDs and upstream DNA.
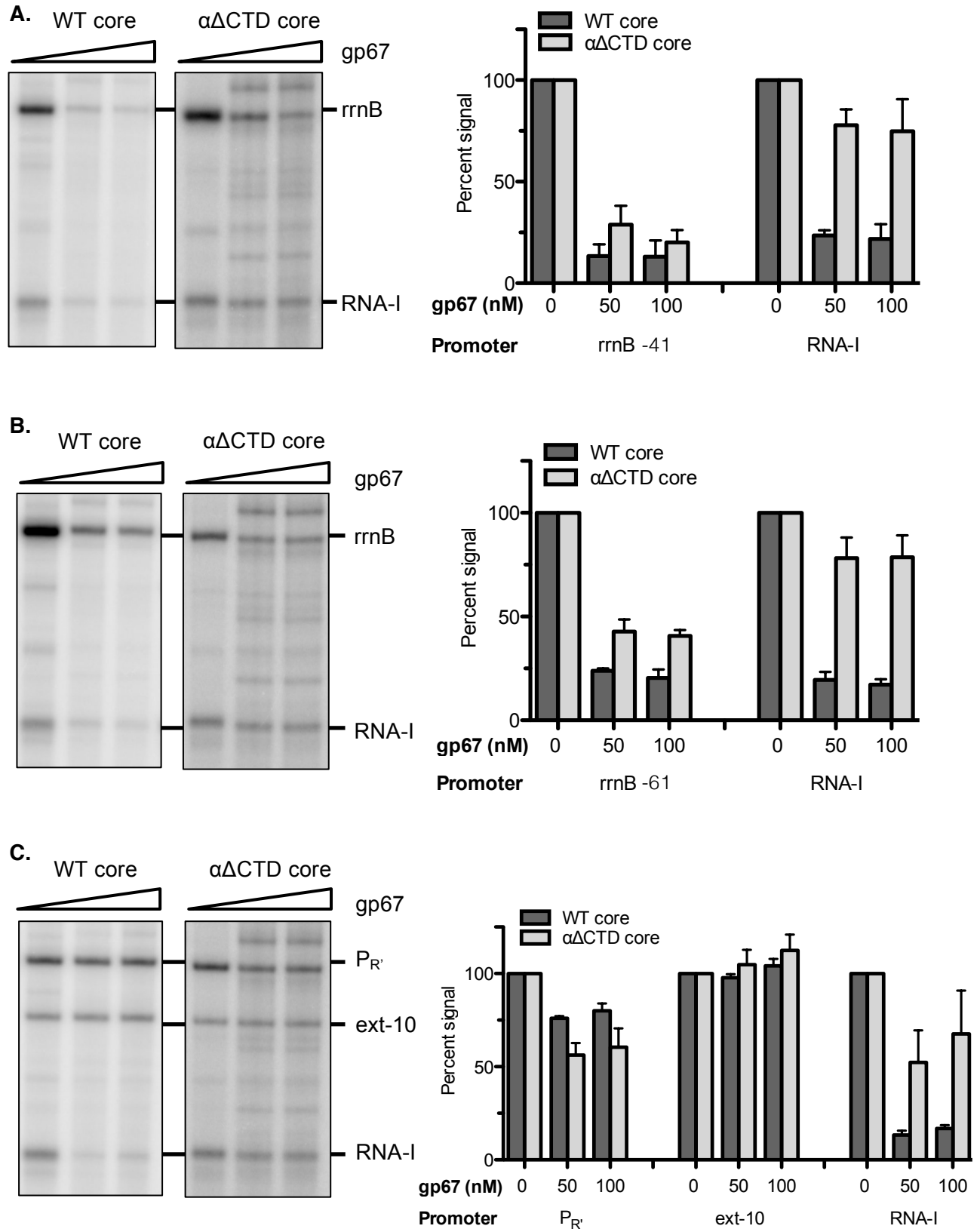
Figure 3.5 (continued)

130

These observations further suggest that, although the RNA-I promoter does not contain an UP element or significantly A+T-rich upstream sequences, sequence-nonspecific interactions of the αCTDs with upstream DNA are important for efficient transcription initiation at this promoter. These effects have not been previously described at the RNA-I promoter, but UP element-independent αCTD-mediated effects on transcription have been described on two well-characterized promoters lacking UP elements (*lacUV5* and λ*P_R*) with kinetic experiments [31]. Our results with the RNA-I promoter suggest that gp67 may be useful as a tool to reveal previously uncharacterized sequence-nonspecific effects of the αCTDs on transcription initiation at specific promoters (see **Discussion**). In contrast, the inhibitory effect of gp67 on transcription from *rrnB* -41 and *rrnB* -61 was only partially dependent on the αCTDs. Specifically, the magnitude of gp67-mediated transcription inhibition was reduced only ~1.8-fold in the absence of the αCTDs for both *rrnB* templates (compare light grey bars and dark grey bars at *rrnB* in **Figure 3.5A** and **Figure 3.5B**).

We also assessed the effect of gp67 on transcription from an additional well-characterized -10/-35 promoter—the bacteriophage λ late promoter $P_{R'}$, a strong promoter that lacks an UP element—and a consensus extended -10 promoter lacking both a -35 element and an UP element. For these experiments, we used supercoiled DNA templates that bore the native $P_{R'}$ promoter in place of the *rrnB* P1 promoter and the extended -10 promoter within the $P_{R'}$ transcribed region (centered at position +19), both upstream of the *rrnB* terminators. Transcription initiation at $P_{R'}$ generates a ~200-nt product (as with the *rrnB* -41 and *rrnB* -61 templates) and initiation at the extended -10 promoter generates a ~170-nt product; both of these transcripts are distinguishable from the 108-nt RNA-I transcript also produced from these templates. As with the *rrnB* -41 and *rrnB* -61 templates, gp67 had no effect on transcription from $P_{R'}$, the extended -10 promoter, or RNA-I by E•σ^{70} or EαΔCTD•σ^{70} (not shown). When we used holoenzyme reconstituted with σ^{70}_{quint}, we observed no gp67-mediated transcription inhibition

with either E•σ$^{70}_{quint}$ or EαΔCTD•σ$^{70}_{quint}$ at the extended -10 promoter (**Figure 3.5C**, dark grey and light grey bars at 'ext-10'), indicating that gp67 does not function as a canonical anti-σ factor to simply sequester σ$^{70}_{quint}$ and prevent holoenzyme formation. Inhibition by gp67 at the P$_{R'}$ promoter was modest and also αCTD-independent—in fact, gp67-mediated transcription inhibition was greater with EαΔCTD•σ$^{70}_{quint}$ (1.7-fold) than with E•σ$^{70}_{quint}$ (1.3-fold). The differential effects of gp67 on transcription from the *rrnB* -41, *rrnB* -61, RNA-I, P$_{R'}$, and extended -10 promoters suggest that gp67 exerts αCTD-dependent and αCTD-independent inhibitory effects at -10/-35 promoters that are largely dependent on promoter context.

Given the unexpected lack of UP element dependence of gp67-mediated transcription inhibition at the *rrnB* P1 promoter and the modest inhibitory effects at P$_{R'}$ in the *Eco* transcription system, we wondered whether the inherent instability of the open complex at *rrnB* P1 [28] might make this promoter uniquely sensitive to gp67 in an UP element- and αCTD-independent manner in *Eco*. To evaluate this possibility, we sought to test the effect of the UP element on gp67-mediated transcription inhibition in the context of a promoter with a stable open complex (P$_{R'}$) by fusing the *rrnB* P1 UP element to the P$_{R'}$ core promoter. (This also addressed the related question of whether the presence of an UP element would confer greater gp67 sensitivity on P$_{R'}$.) The *rrnB* P1 UP element had no significant effect on transcription when fused to the wild-type P$_{R'}$ core promoter (a strong promoter); however, when we created a P$_{R'}$ variant with a weakened -35 element (TTGACT to TTGA<u>TA</u>; substitutions underlined), we were able to detect a modest (~2.5 to 3-fold) stimulatory effect of the appended UP element on transcription. We then tested the effect of gp67 on transcription by E•σ$^{70}_{quint}$ and EαΔCTD•σ$^{70}_{quint}$ from supercoiled templates bearing either the native P$_{R'}$ upstream sequence or the *rrnB* P1 UP element upstream of the weakened P$_{R'}$ promoter; as with the wild-type P$_{R'}$ template, these variants also bore the consensus extended -10 promoter centered at position +19, and the RNA-I promoter within the plasmid replication origin. We found that gp67 significantly inhibited

transcription from the weakened $P_{R'}$ variants (up to 3.7-fold) irrespective of the presence of the

UP element (dark grey bars at '$P_{R'}$*' in **Figure 3.6A** and **Figure 3.6B**), and that this inhibitory

effect was αCTD-independent (light grey bars at '$P_{R'}$*' in **Figure 3.6A** and **Figure 3.6B**). These

data reinforce our previous observations with the *rrnB* P1 templates that gp67 can exert αCTD-

independent inhibitory effects in the context of an *Eco*-based transcription system and further

suggest that these effects are not dependent on the stability of the promoter open complex.

(Although we have not directly shown that the weakened $P_{R'}$ variants do not form less stable

open complexes than wild-type $P_{R'}$, this is unlikely to be the case because open complex

stability is largely determined by the interaction of $\sigma_{1.2}$ with the base two positions downstream

of the -10 element [11]. A cytosine base on the template strand at this position severely

weakens the $\sigma_{1.2}$/ discriminator interaction and results in unstable open complexes, as in *rrnB*

P1; both wild-type $P_{R'}$ and the weakened $P_{R'}$ variant have the highly-favored guanine base at

this position, suggesting that both form stable open complexes). The observation that

weakening the -35 element of the $P_{R'}$ promoter increased its sensitivity to gp67 suggests that

gp67 may interfere with the ability of σ_4 to make optimal contacts with the -35 promoter element,

in addition to disrupting sequence-specific or -nonspecific contacts of the αCTDs with upstream

DNA (see below and **Discussion**).

In the transcription assays with the weakened $P_{R'}$ variants we found that transcription

from the extended-10 promoter actually *increased* in the presence of gp67 with either $E \cdot \sigma^{70}_{quint}$

or $E\alpha\Delta CTD \cdot \sigma^{70}_{quint}$ (dark and light grey bars at 'ext-10' in **Figure 3.6A** and **Figure 3.6B**). We

suspected that this apparent stimulatory effect of gp67 on transcription was the result of

competition between the $P_{R'}$ and extended -10 promoters, which are ~30 nt apart, rather than

actual stimulation by gp67: in the case of the weakened $P_{R'}$ variant, inhibition by gp67 at the $P_{R'}$

promoter is greater than at wild-type $P_{R'}$, which would free up more holoenzyme to initiate

transcription at the extended -10 promoter. To eliminate this potential competition, we

Figure 3.6. Gp67 inhibits transcription by holoenzyme reconstituted with *Eco* σ^{70}_{quint} from a P$_{R'}$ promoter with a weakened -35 element in an αCTD-independent manner. A representative gel is shown for each experiment, and at least three gels were quantified by phosphorimagery in each panel. For each independent promoter, the amount of transcription was normalized to that observed in the absence of gp67, and data represent the average amount of transcription ± standard deviation for E•σ^{70}_{quint} (dark grey bars) and EαΔCTD•σ^{70}_{quint} (light grey bars). All supercoiled plasmid templates contain the RNA-I promoter within the plasmid replication origin. **(A)** Effect of gp67 on transcription from a P$_{R'}$ promoter variant (P$_{R'}$*) with a weakened -35 element (TTGACT to TTGATA) that bears the native upstream P$_{R'}$ sequence and a consensus extended -10 promoter within the P$_{R'}$ transcribed region. Gp67 significantly inhibits transcription from this P$_{R'}$ variant in a manner that is independent of the presence of the αCTDs (compare dark grey bars and light grey bars for 'P$_{R'}$*'); the same is observed in **(B)** for the weakened P$_{R'}$ variant bearing the *rrnB* P1 UP element upstream of its -35 element. An αCTD-dependent inhibitory effect of gp67 at the RNA-I promoter is again observed. Taken together, these data provide further evidence that gp67 can exert αCTD-independent inhibitory effects in the context of an *Eco*-based transcription system and additionally suggest that these effects of gp67 may be mediated through the σ_4/-35 element interaction. **(C)** The apparent stimulatory effect of gp67 on transcription from the extended -10 promoter in the context of the P$_{R'}$ variants is the result of competition between the two promoters. Gp67 has no effect on transcription from the extended -10 promoter when the competing P$_{R'}$ promoter is inactivated, indicating that there is no direct activation of transcription by gp67 at the extended -10 promoter.
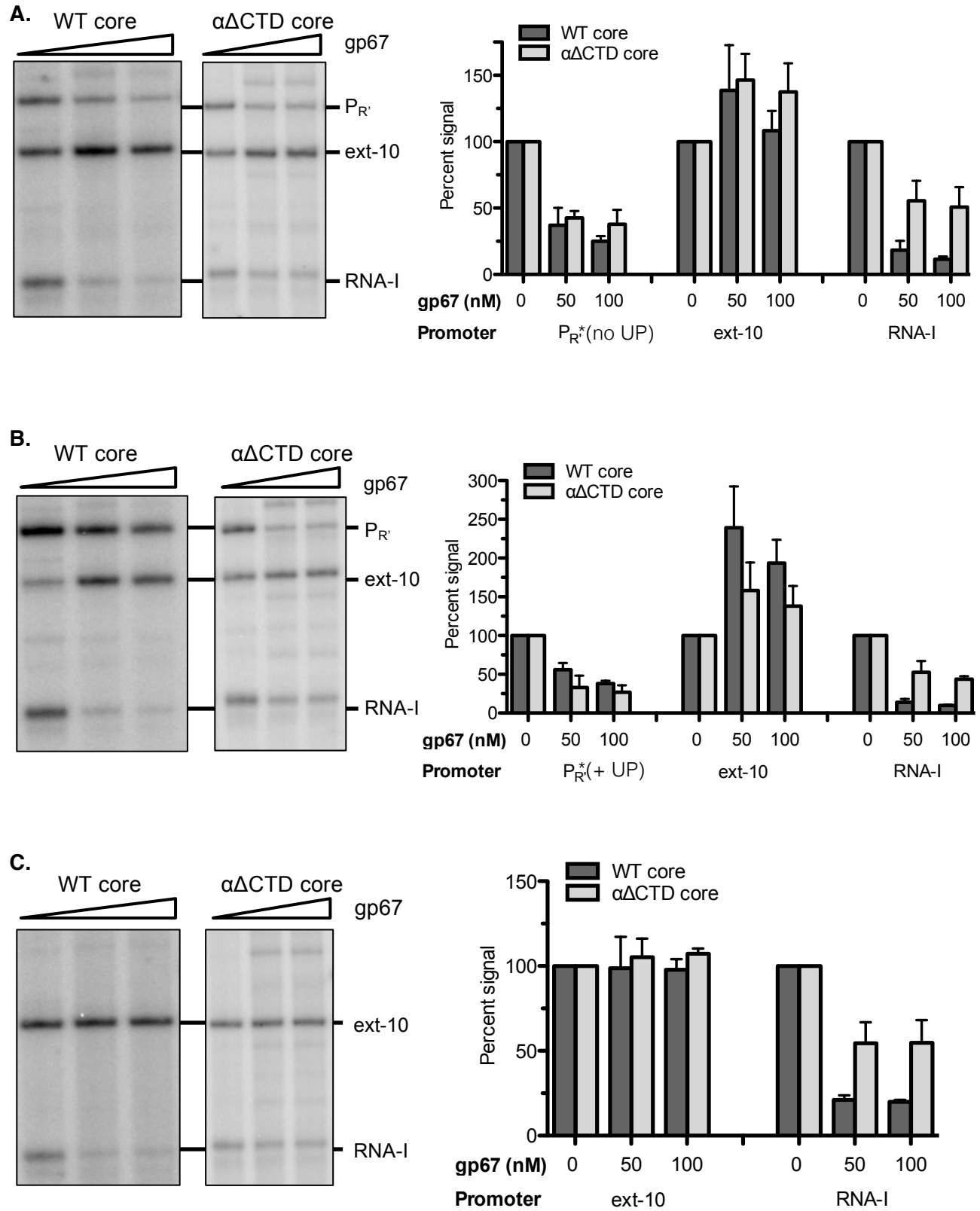
Figure 3.6 (continued)

inactivated the $P_{R'}$ promoter and left the extended -10 promoter intact. As seen in **Figure 3.6C**,

the addition of gp67 had no effect on transcription from the extended -10 promoter in the

absence of the competing weakened $P_{R'}$ promoter, indicating that gp67 does not directly

stimulate transcription from this promoter.

In light of the observed competition between the weakened $P_{R'}$ and extended -10

promoters, we wondered whether competition between the RNA-I promoter and the other

defined promoters on the various plasmids (*rrnB* P1 with or without the UP element; wild-type or

weakened $P_{R'}$; or the extended -10 promoter) resulted in an underestimate of the αCTD

dependence of gp67-mediated inhibition at the RNA-I promoter. To address this question, we

constructed a version of the supercoiled plasmid template that lacked a test promoter upstream

of the *rrnB* terminators but retained the RNA-I promoter within the plasmid origin of replication;

we then conducted multi-round transcription assays as before. If competition between the

various promoters on the previously used plasmids was responsible for the residual (and slightly

variable) αCTD dependence of gp67-mediated inhibition of RNA-I transcription, we would

expect to see no significant effect of gp67 on transcription by $E\alpha\Delta CTD \cdot \sigma^{70}_{quint}$ on the template

bearing only the RNA-I promoter and, possibly, a stronger inhibitory effect with $E \cdot \sigma^{70}_{quint}$. As

seen in **Figure 3.7**, this seems to be the case: with the RNA-I promoter alone, we observe very

significant inhibition by gp67 (up to 11-fold) with $E \cdot \sigma^{70}_{quint}$ (dark grey bars) but observe no

significant effect of gp67 on transcription in the absence of the αCTDs (light grey bars). These

data further substantiate the importance of αCTD/nonspecific DNA contacts for efficient initiation

at the RNA-I promoter and raise the possibility that competition between multiple promoters in

our transcription assays may mask *some* of the αCTD dependence of the gp67-mediated

inhibition reported at the *rrnB* P1, wild-type $P_{R'}$, and weakened $P_{R'}$ promoters (though the bulk of

this inhibition is likely primarily αCTD-independent).
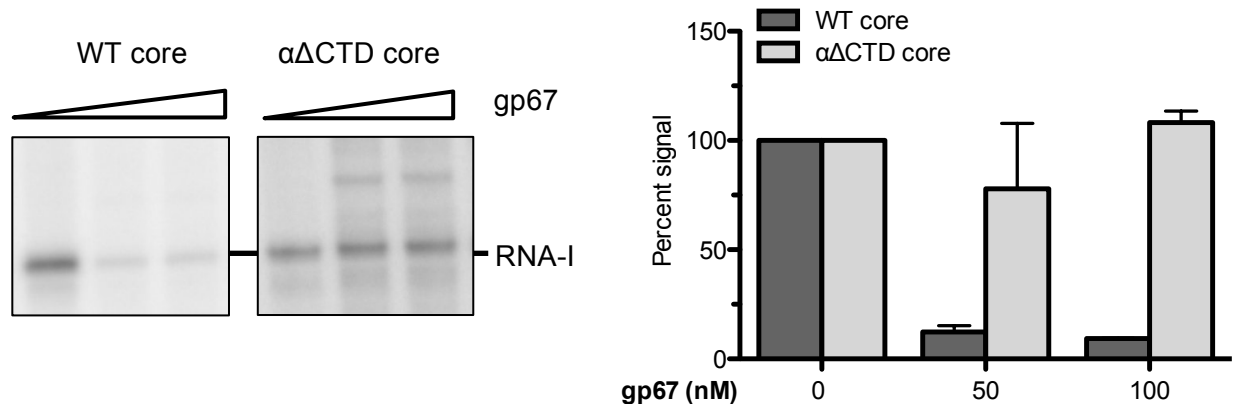
Figure 3.7. Gp67-mediated transcription inhibition at the RNA-I promoter is fully αCTD-dependent in the absence of competing promoters. A representative gel is shown for each experiment, and three independent gels were quantified by phosphorimagery. The amount of transcription was normalized to that observed in the absence of gp67, and the data represent the average amount of transcription ± standard deviation (SD) for $E{\cdot}\sigma^{70}_{quint}$ (dark grey bars) and $E\alpha\Delta CTD{\cdot}\sigma^{70}_{quint}$ (light grey bars). Gp67 significantly inhibits transcription at the isolated RNA-I promoter by $E{\cdot}\sigma^{70}_{quint}$, but has no effect in the absence of the αCTDs. The reduction in signal and large SD for reactions conducted with $E\alpha\Delta CTD{\cdot}\sigma^{70}_{quint}$ and 50 nM of gp67 can be attributed to visually evident underloading of two of the corresponding lanes (not shown). Taken together, these data suggest that contacts between *Eco* holoenzyme and upstream DNA are important for initiation at the RNA-I promoter, and that competition between multiple promoters in our templates led to an apparent residual effect of gp67 in the absence of the αCTDs on RNA-I transcription.

We previously reported that unfused gp67 compromises -35 element recognition by an α-*Sau* σ^A_4 fusion protein in the context of the bacterial one-hybrid system (see **Figure 2.13B** in **Chapter 2** of this dissertation) and demonstrated that the interaction between *Sau* σ^A residue R310 (corresponding to *Eco* σ^{70} residue R554) and the DNA phosphate backbone at the -36 position is important for detectable -35 element recognition in this system (see **Figure 2.13D** in **Chapter 2**). (*Sau* σ^A residue R310 is the only one of the ten DNA-binding residues in σ_4 that is part of the interface with gp67, and it is redirected away from the DNA-binding interface into a deep pocket in gp67 upon gp67 binding [25]). The observation that substitutions at *Sau* σ^A residue R310 in the context of the α-*Sau* σ^A_4 fusion protein in the one-hybrid system phenocopied the transcription-inhibitory behavior of unfused gp67 in this system suggested that this inhibitory effect might be due to the disruption of the contact between *Sau* σ^A residue R310 (*Eco* σ^{70} R554) and the phosphate backbone upon gp67 binding. We postulated that this protein/DNA contact may be important for efficient -35 element recognition (and thus transcription initiation) at some -10/-35 promoters, and that by disrupting this contact, gp67 modulates the binding of σ_4 to the -35 element and exerts transcription-inhibitory effects that are independent of the αCTDs.

We utilized our gp67-responsive *Eco* transcription system to directly address this by reconstituting and purifying full-length *Eco* σ^{70} bearing substitutions to alanine or aspartate at residue R554 (*Eco* σ^{70} R554A and *Eco* σ^{70} R554D) and conducting multi-round transcription experiments using the supercoiled *rrnB* -41 and *rrnB* -61 templates. The use of these templates additionally allowed us to determine whether the *Eco* σ^{70} R554 (*Sau* σ^A R310)/-36 backbone contact plays a role in appropriately positioning the UP element for efficient recognition by the αCTDs, which presents a possible mechanism by which gp67 can specifically inhibit UP-element dependent transcription in *Sau*. If the *Eco* σ^{70} R554/-36 backbone contact is important to achieve the correct holoenzyme/DNA geometry for productive UP element utilization, we

138

would expect the substitutions at *Eco* σ^{70} R554 to result in reduced transcription only from the

rrnB -61 promoter; additionally, we would expect removal of the αCTDs to have no further

detrimental effect on transcription (since the loss of the R554/-36 backbone contact would

prevent the αCTDs from being able to make productive contact with the UP element in the first

place). As seen in **Figure 3.8**, transcription from *rrnB* -41, *rrnB* -61, and RNA-I was significantly

reduced when using either E•σ^{70} R554 (A or D) or EαΔCTD•σ^{70} R554 (A or D), which is

inconsistent with a crucial role of the *Eco* σ^{70} R554/-36 backbone contact in enabling efficient

UP element utilization. Furthermore, inspection of the gels (**Figure 3.8**, panels A and B, left

side) revealed that even with the σ^{70} R554 mutants, removal of the αCTD caused further

reductions in transcription. These observations raised the possibility that the R554 substitutions

compromised the structural integrity of *Eco* σ^{70}, so we looked at the effect of the R554

substitutions on transcription initiation from the consensus extended -10 promoter. We saw no

defect in initiation from this promoter by holoenzyme reconstituted with *Eco* σ^{70} R554A or

R554D (**Figure 3.8C**), which indicated that these proteins are properly folded and functional and

that the initiation defects at the remaining promoters are specific.

The *Eco* σ^{70} R554D substitution resulted in more severe defects in transcription initiation

at the *rrnB* -41 and *rrnB* -61 promoters than the R554A substitution, but the defects are similar

at the two promoters (see 'rrnB' in **Figure 3.8A** and **Figure 3.8B**) and are comparable to those

observed with E•σ$^{70}_{quint}$ in the presence of gp67. These data suggest that the *Eco* σ^{70} R554/-36

backbone contact is important for efficient initiation at these promoters and that the disruption of

this contact by gp67 with E•σ$^{70}_{quint}$ likely accounts for the primarily αCTD-independent inhibitory

effects reported in **Figure 3.5A** and **Figure 3.5B** ('rrnB' bars). We also observed a substantial

(and roughly equivalent) reduction in transcription by both E•σ^{70} R554 (A or D) and EαΔCTD•σ^{70}

R554 (A or D) at the RNA-I promoter, where we previously demonstrated that gp67-mediated

inhibition is dependent on the αCTDs ('RNA-I' bars in **Figures 3.5 - 3.7**). The data with gp67

Figure 3.8. Substitutions at *Eco* σ^{70} residue R554 reduce transcription initiation at the

***rrnB* P1 and RNA-I -10/-35 promoters.** A representative gel is shown for each experiment, and

three gels were quantified by phosphorimagery in each panel. For each promoter, the amount of

transcription was normalized to that observed with wild-type *Eco* σ^{70}, and data represent the

average amount of transcription ± standard deviation for holoenzyme reconstituted with wild-

type core RNAP (dark grey bars) or αΔCTD core RNAP (light grey bars) and wild-type *Eco* σ^{70},

Eco σ^{70} R554A, or *Eco* σ^{70} R554D. All supercoiled plasmid templates contain the RNA-I

promoter within the plasmid replication origin. **(A)** Effect of substitutions at *Eco* σ^{70} R554 on

transcription from the *rrnB* -41 template. Substitutions at *Eco* σ^{70} R554 significantly reduce

transcription by holoenzyme reconstituted with wild-type or αΔCTD core RNAP at the *rrnB* P1

promoter without an UP element and at the RNA-I promoter (also lacking an UP element).

These data indicate that the *Eco* σ^{70} R554/-36 backbone contact does not play a role in

appropriately positioning upstream DNA for engagement by the αCTDs but is important for

efficient recognition of the -35 element at these promoters. These same effects are also

observed with the *rrnB* -61 template in **(B)**, suggesting that the disruption of the *Eco* σ^{70} R554/-

36 backbone contact by gp67 binding to modified *Eco* σ^{70} likely accounts for the αCTD-

independent inhibitory effects of gp67 at the *rrnB* P1 promoter. The reduction in transcription by

EαΔCTD•σ^{70} R554 (A or D) at the RNA-I promoter, where the αCTDs seem to play an important

role in initiation, suggests that contact between *Eco* σ^{70} R554 and the phosphate backbone is

also important at this promoter. Sequence-nonspecific interactions between gp67 and upstream

DNA in the absence of the αCTDs may compensate for some of the reduction in transcription

stemming from at the disruption of the *Eco* σ^{70} R554/-36 backbone contact at the RNA-I

promoter. **(C)** The lack of effect of substitutions at *Eco* σ^{70} R554 on transcription from the

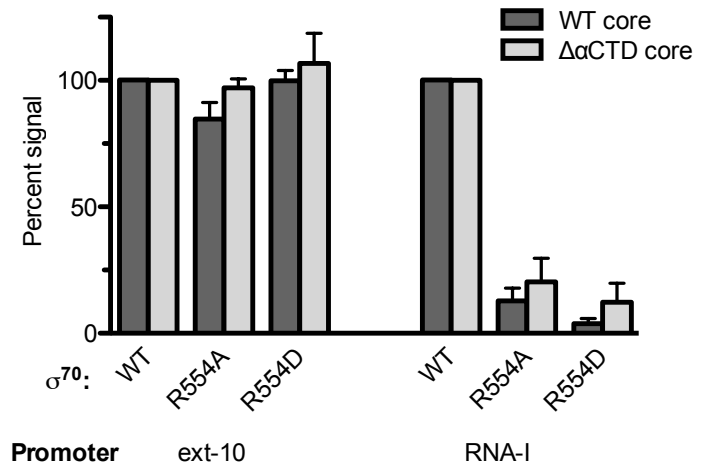extended -10 promoter indicates that these substitutions do not compromise structural integrity.
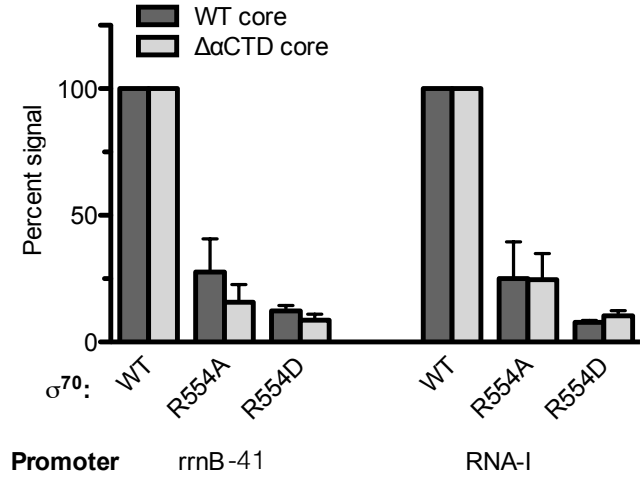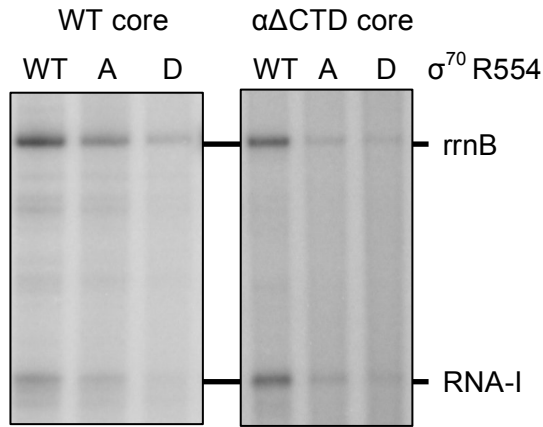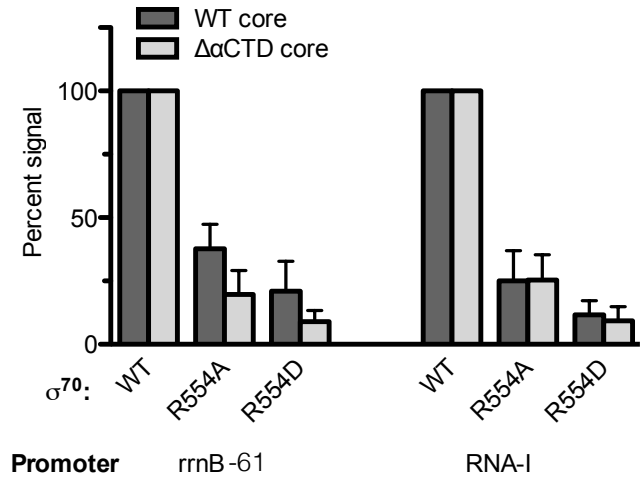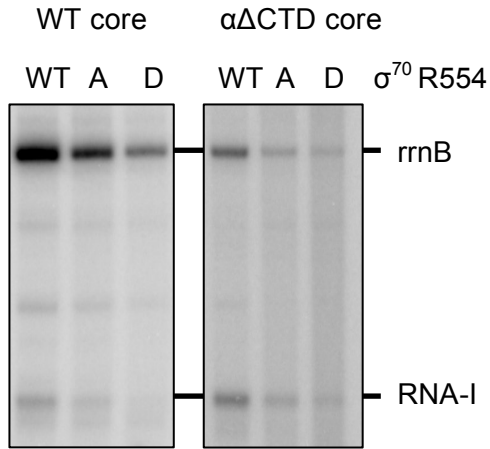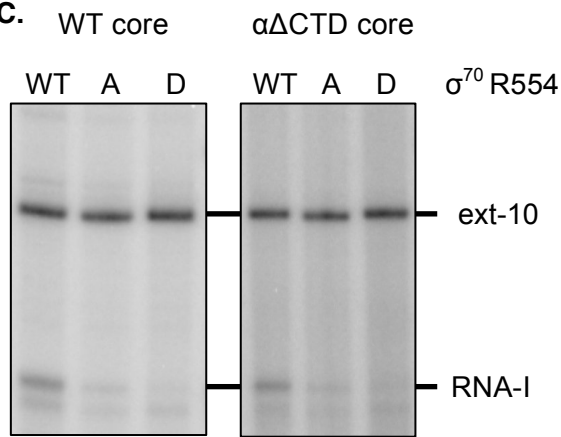
Figure 3.8 (continued)

suggested that sequence-nonspecific contacts of the αCTDs with upstream DNA are important for efficient initiation at this promoter, and that effects of gp67 on -35 element recognition by σ_4 at the RNA-I promoter would likely be minimal.

The apparent discrepancy between the observations with Eco σ^{70} R554 (A or D) at the RNA-I promoter and those with gp67 might be explained as follows. Specifically, the observations with Eco σ^{70} R554 (A or D) imply that the Eco σ^{70} R554/-36 backbone contact is important for transcription initiation at this promoter (as with the $rrnB$ P1 promoter), so transcription is reduced when this contact is disrupted either by substitutions at Eco σ^{70} R554 or by the binding of gp67 in the context of wild-type core RNAP. To explain the lack of gp67-mediated inhibition in the absence of the αCTDs, we suggest that gp67 may make sequence-nonspecific contacts with upstream DNA in the context of the ternary RNAP complex that stabilize holoenzyme binding at the RNA-I promoter, thus compensating for the transcription defect resulting from the loss of the Eco σ^{70} R554/-36 backbone contact. These data suggest that the RNA-I promoter is particularly dependent on stabilizing contacts with upstream DNA by either the αCTDs or another factor like gp67 for efficient transcription initiation. There is no direct biochemical evidence that gp67 binds DNA alone or in the context of the ternary RNAP complex, but gp67 is located close to upstream DNA in the model of the Sau σ^{A}_{4}/gp67 complex bound to promoter DNA, with the DNA-facing surface rich in basic residues [25]. Furthermore, we see the appearance of novel bands—presumably representing transcription initiation events—in all lanes containing gp67 for all supercoiled plasmid DNA templates specifically when using EαΔCTD•σ^{70}_{quint} (see gel images in **Figures 3.4 - 3.7**). This may be indicative of a sequence-nonspecific DNA-binding activity of gp67 in the context of the ternary RNAP holoenzyme complex, which may play a role in the infectious cycle of phage G1 (see **Discussion**).

Discussion

      Bacteriophages often depend on the host bacterial transcription machinery to execute the transcription program required for viral propagation, and much of the regulation of phage development occurs at the level of viral transcription. As such, phages have developed a wide array of mechanisms to modify and appropriate host RNAP to serve viral needs, typically through direct covalent modifications of host RNAP or through specific RNAP-binding proteins [reviewed in 24]. The phage G1-encoded transcription regulator gp67 has been postulated to inhibit growth of its pathogenic host, *S. aureus*, by inhibiting new rRNA synthesis by a novel mechanism: it becomes a stable component of the holoenzyme by interacting with conserved region 4 of the primary σ factor in *Sau* (σ^A) but inhibits transcription by preventing the functional engagement of UP elements by the RNAP αCTDs [25]. The rRNA promoters are the most well-characterized UP element-containing promoters; these promoters have been shown to depend on the UP element for efficient initiation *in vivo* and *in vitro* in *E. coli* [29, 30], but have not been as thoroughly studied in *Sau*.

      Here we describe the development of a gp67-responsive *Eco* transcription system to thoroughly investigate the reported UP element and αCTD dependence of gp67 promoter selectivity. We demonstrated that, like in its native host, gp67 inhibited growth in *Eco* when transcription was directed by a modified version of σ^{70} that binds gp67, and that gp67 forms the functionally relevant ternary complex with this modified *Eco* holoenzyme. We used well-characterized reagents in *Eco* for studying UP element-dependent transcription—chiefly αΔCTD core RNAP and the model rRNA promoter *rrnB* P1 with or without its UP element—to reveal a strong inhibitory effect of gp67 on transcription from -10/-35 promoters (but not an extended -10 promoter) that, unexpectedly, did not depend on the presence of an UP element. We observed significant gp67-mediated transcription inhibition at the *rrnB* P1 promoter and a version of $P_{R'}$ with a weakened -35 element with and without an UP element, and at the RNA-I promoter

(which lacks an UP element). At all promoters except RNA-I, gp67-mediated inhibition was largely αCTD-independent, prompting the conclusion that—at least in the context of the *Eco* transcription machinery—gp67 can inhibit transcription in both an αCTD-dependent (in the case of RNA-I promoter) and an αCTD-independent (in the case of the *rrnB* P1 and $P_{R'}$ promoters) manner. Using genetic data from **Chapter 2**, structural data on the *Sau* σ^A_4/gp67 and σ_4/DNA interfaces [25], and our *Eco* transcription system, we assessed the importance of a contact between *Eco* σ^{70} residue R554 (corresponding to *Sau* σ^A residue R310) and the DNA phosphate backbone at position -36 in transcription initiation from these promoters and as a possible explanation for αCTD-independent gp67-mediated inhibition.

Gp67 as a tool to uncover sequence-nonspecific effects of the αCTDs on promoter activity

The importance of the sequence-specific binding of the RNAP αCTDs to the UP element for efficient initiation at promoters like the rRNA promoters in *Eco* is well documented: sequences upstream of the *rrn* P1 core promoters in *Eco* have been shown to stimulate transcription by up to 300-fold [8, 18, 29, 30], with the UP element alone resulting in ≥30-fold stimulation *in vivo* and *in vitro* in the presence of only the basal transcription machinery [29, 30]. Ribosome biogenesis—limited by rRNA transcription—is crucial for cellular metabolism [reviewed in 28], so the αCTDs are essential for cell viability; however, αΔCTD core RNAP can be used to assemble a transcriptionally-active holoenzyme *in vitro* that cannot discriminate between otherwise-identical promoters bearing or lacking an UP element. The αΔCTD-containing holoenzyme could, in principle, be used to probe the role of the αCTDs in initiation at promoters that do not have UP elements (mediated by stabilizing sequence-nonspecific contacts of the αCTDs with upstream DNA), but direct comparisons between transcription reactions conducted with wild-type and αΔCTD-holoenzymes are problematic since one must account for potential differences in the activities of the protein preparations. Kinetic studies

provide one possible way to circumvent this problem and have been used to demonstrate important effects of sequence-nonspecific interactions of the αCTDs with upstream DNA on the rate of RNAP association at two promoters lacking UP elements [31]. Gp67 provides a potential alternative to kinetic assays in this respect: since gp67 compromises the function of the αCTDs, *in vitro* transcription reactions conducted with the same wild-type core RNAP preparation in the presence and absence of gp67 could be used to determine whether sequence-nonspecific αCTD/DNA contacts play a role in promoter function at any promoter. In this sense, gp67 and the αΔCTD core RNAP are functionally equivalent in the modified *Eco* transcription system. We note, however, that the unanticipated αCTD-independent inhibitory effects of gp67 observed at the *rrnB* P1 and weakened $P_{R'}$ promoters limit the usefulness of this approach in the context of *Eco* RNAP (see **Chapter 4** for further discussion of gp67 as a potential tool to uncover sequence-nonspecific effects of the αCTDs on promoter activity).

Comparison of the effects of gp67 in the context of *Sau* RNAP and *Eco* RNAP

Gp67 forms a ternary complex with the σ^A-containing *Sau* holoenzyme via its interaction with *Sau* σ^A_4, but structural modeling of the *Sau* σ^A_4/gp67 complex onto the σ_4/DNA complex suggests that gp67 binding does not interfere with -35 element recognition by *Sau* σ^A_4 [25]. Combined with the observations that gp67 affected transcription from only ~9% of *Sau* promoters and that gp67-sensitive promoters contained A+T-rich upstream regions, the data in *Sau* suggest that gp67 exerts its transcription-inhibitory effects not via the σ subunit, but rather by interfering with interactions between the αCTDs and DNA upstream of the promoter [25]. Using a bacterial one-hybrid system to detect σ_4/-35 element interactions, we unexpectedly observed that unfused gp67 specifically reduced transcription activation from the reporter by an α-*Sau* σ^A_4 fusion protein (see **Chapter 2**). In the modified *Eco* transcription system, we further noted that gp67 significantly inhibited transcription from both the *rrnB* P1 promoter and a $P_{R'}$

145

variant with a weakened -35 element in a manner that was wholly independent of the presence of an UP element and only partially dependent on the αCTDs. These αCTD-independent effects are specific to -10/-35 promoters and not simply the result of *Eco* σ[70] sequestration by gp67, since gp67 had no effect on transcription from an extended-10 promoter lacking a -35 element. These data, coupled with the observation that weakening the -35 element of $P_{R'}$ enhanced the sensitivity of this promoter to gp67, suggest that gp67 can interfere with -35 element recognition by σ_4 in addition to disrupting contacts between the αCTDs and upstream DNA in the context of *Eco* RNAP.

These unexpected αCTD-independent inhibitory effects and less stringent promoter selectivity of gp67 in our modified *Eco* system may be the result of structural differences between the *Eco* and *Sau* RNAP holoenzymes. The crystal structure of the *Sau* σ^A_4/gp67 complex revealed that only one of the ten σ_4 residues implicated in promoter recognition is part of the interface with gp67: *Sau* σ[A] residue R310, which corresponds to *Eco* σ[70] R554 [25]. This residue contacts the DNA phosphate backbone at position -36 immediately upstream of the -35 element [4, 23], and we have shown that this contact is important for efficient -35 element recognition in the context of σ_4 alone, as assayed in our one-hybrid assay (see **Chapter 2**). Thus, one possible explanation for the difference between the action of gp67 in the *Eco* and *Sau* transcription systems is that the *Eco* σ[70] R554/backbone contact contributes more significantly to promoter recognition in the *Eco* promoters we assayed (*rrnB* P1, $P_{R'}$ and a weakened $P_{R'}$ variant, and RNA-I) than the corresponding *Sau* σ[A] R310/backbone contact does at the majority of *Sau* promoters. Many promoters in Gram-positive bacteria contain extended -10 motifs in combination with consensus or near-consensus -10 and -35 elements and A+T-rich upstream regions [3, 13, 33], so it is possible that a multitude of stabilizing interactions between *Sau* holoenzyme and promoter DNA limits the importance of the *Sau* σ[A] R310/backbone contact in transcription initiation.

It is important to note that structure-based modeling of the gp67/holoenzyme/melted

promoter complex revealed minor steric clashes between the presumably conformationally-

flexible helical tower of the gp67 CTD and the -35 element that could be relieved by small

rearrangements in either the gp67 CTD or the DNA [25]. These rearrangements may be

accommodated more easily by *Sau* RNAP than by *Eco* RNAP, which could result in more

severe effects on -35 element recognition in the context of *Eco* RNAP, as observed in our

experiments. It is also possible that the more restrictive action of gp67 on transcription observed

in the *Sau* system is the result of sequence-independent interactions of gp67 with upstream

DNA in the context of the ternary *Sau* holoenzyme complex. Gp67 may disrupt sequence-

specific or -nonspecific contacts between the *Sau* αCTDs and/or the *Sau* σ^A R310/-36 backbone

contact at all σ^A-dependent *Sau* promoters but may compensate for the loss of these contacts

(and the subsequent reduction in transcription) with nonspecific stabilizing contacts it itself

makes at the majority of *Sau* promoters. Consistent with this possibility, structural data positions

the gp67 NTD close to DNA upstream of the -35 element, with the surface facing the DNA rich

in basic residues [25]. Due to possible differences in promoter architecture and/or structural

features of RNAP between *Eco* and *Sau*, these potential gp67-mediated stabilizing contacts

may not occur as readily in the context of *Eco* holoenzyme so gp67 may exert more global

inhibitory effects on -10/-35 promoters in the *Eco* system.

The potential DNA-binding activity of gp67 in the context of the *Sau* holoenzyme may

have implications for the life cycle of the phage. Since the bulk of RNAP holoenzyme is engaged

in transcribing rRNA during rapid growth [reviewed in 28], gp67-mediated inhibition of rRNA

promoters would free up a substantial amount of holoenzyme that could be redirected to phage

genes. By engaging in sequence-nonspecific interactions with upstream DNA, gp67 may be

directly involved in the repurposing of *Sau* RNAP by stabilizing E$_{Sau}$•σ^A at phage promoters in a

promoter context-dependent manner. Alternatively, gp67 may work in conjunction with another

phage-encoded factor to redirect $E_{Sau} \cdot \sigma^A$ to specific phage promoters where gp67/DNA contacts can further stabilize the binding of $E_{Sau} \cdot \sigma^A$ and stimulate viral transcription.

A *B. subtilis*-based system for the study of gp67 function

Our observations of the action of gp67 in the *Eco*-based transcription system differ from those made in the *Sau* system and could be attributed to specific fine-structure features of the enzymes or promoters that result from the evolutionary distance between *Eco* and *Sau*. The spore-forming low-G+C Gram-positive bacterium *B. subtilis* is significantly more closely related to *Sau* than *Eco* and has the advantage of an abundance of genetic and biochemical tools, including well-established protocols for the purification of RNAP components and well-characterized promoters (including the rRNA promoters) and UP elements. Thus, *Bsu* presents an attractive alternative to the *Eco* system for detailed mechanistic studies of how gp67 antagonizes the function of the αCTDs and exhibits promoter selectivity. If the αCTD-independent effects of gp67 are strictly a feature of the *Eco* transcription system, the development of the *Bsu* system would allow us to effectively use gp67 as a tool to uncover sequence-nonspecific effects of the αCTDs that contribute significantly to promoter function. As a first step in developing a *Bsu*-based *in vivo* and *in vitro* system to study gp67, we have demonstrated that gp67 binds *Bsu* σ^A_4 in a manner dependent on the identity of the amino acid side chains at the four gp67 specificity determinants identified in *Sau* σ^A_4 and that the production of unfused gp67 in *Bsu* cells inhibits growth.

Materials and Methods

Plasmids, Strains, and Growth Conditions

A complete list of the bacterial strains used in this chapter of the dissertation is provided in **Table 3.1**. The *E. coli* σ^{70}-depletion strain NUN449 was constructed by introducing the $\Omega(Cam^r)P_{trp}$-*rpoD* allele from strain CAG20153 [20] into MG1655 *ΔlacZYA* by P1-mediated transduction and identifying Cam^r indole-3-acrylic acid (IAA)-dependent colonies. Cultures of NUN449 were grown as serial dilutions at 37ºC in LB broth supplemented with chloramphenicol (25 µg/mL) and IAA (200 µM) to produce unsaturated overnight cultures. Cultures of NUN449 co-transformed with compatible plasmids directing the synthesis of *Eco* σ^{70} derivatives (weak constitutive promoter) and gp67 (or no protein; IPTG-inducible promoter) were additionally supplemented with carbenicillin (100 µg/mL) and spectinomycin (50 µg/mL) and had no need for IAA. Plates containing NUN449 were grown at 30ºC with the appropriate antibiotics (and IPTG as necessary) for fewer than 16 hours.

The *B. subtilis* strain CMD37 is a derivative of the prototrophic wild-type strain PY79 [35] bearing phage G1 *orf67* (encoding gp67-His$_6$) at the nonessential *amyE* locus on the *Bsu* chromosome. Gp67-His$_6$ was generated by PCR and cloned into the ectopic integration vector pDR111 (a kind gift from D. Rudner) as a HindIII/NheI restriction fragment. Correct amylase-negative integrants were identified on starch plates. CMD37 was grown at 37ºC in LB broth or plates supplemented with spectinomycin (100 µg/mL) and IPTG as indicated to induce the production of gp67.

E. coli strain NEB5α F'IQ (New England Biolabs) was used as the recipient for all plasmid constructions. A complete list of plasmids used in this chapter of the dissertation is provided in **Table 3.2**. The construction of several classes of plasmids is outlined below.

Table 3.1. List of strains used in this study.

Strain	Relevant details	Reference
NEB5α F' I^q	*E. coli lacI^q* host strain for plasmid construction	New England Biolabs
BL21(DE3)	*E. coli* strain for protein expression & purification from plasmids that require T7 RNA polymerase for expression	New England Biolabs
CAG20153	*E. coli* strain in which chromosomal *rpoD* is under the control of the repressible *trp* promoter	[20]
NUN449	MG1655 *ΔlacZYA E. coli* derivative bearing the Ω(Cam^r)P$_{trp}$-*rpoD* allele from strain CAG20153	N. Nair
PY79	*B. subtilis* prototrophic wild-type strain used to generate strains with gene of interest integrated on chromosome at nonessential *amyE* locus	[35]
CMD37	Derivative of *B. subtilis* PY79 with a single copy of *orf67* (encoding gp67-His$_6$) under the control of the IPTG-inducible *hyper-spank* strong promoter integrated at the *amyE* locus on the chromosome	This work

Table 3.2. List of plasmids used in this study.

Plasmid	Relevant details	Reference
pBR-*Eco* σ$^{70}_{WT}$	Encodes full-length, unfused *Eco* σ^{70} (WT) under the control of a weak constitutive synthetic promoter with sequence <u>TTTACA</u>ACATGAAGTAACTTCTCG<u>CATTAT</u>GTCTCGA (-35 & -10 sequences underlined); confers Carb^R; medium-copy	[17]
pBR-*Eco* σ$^{70}_{hybrid}$	Derivative of pBR-*Eco* σ$^{70}_{WT}$ where region 4 has *Sau* σ^A residues 309-335 in place of *Eco* σ^{70} residues 553-579; confers Carb^R; medium-copy	This work
pBR-*Eco* σ$^{70}_{quint}$	Derivative of pBR-*Eco* σ$^{70}_{WT}$ bearing the following substitutions in σ^{70}: A553D, A556E, K557N, Q579V and D581G; confers Carb^R; medium-copy	This work
pCDF*lac*	High-copy plasmid with multiple cloning site; encodes no functional protein; confers Spec^R	Modified from pCDF-1b (Novagen)
pCDF*lac*-gp67-His$_6$	High-copy plasmid; encodes gp67 bearing a C-terminal His$_6$ tag under the control of the *lacUV5* promoter; confers Spec^R	This work
pDR111	*B. subtilis* vector for integration at *amyE* locus on chromosome; directs expression of gene of interest under control of strong IPTG-inducible *hyper-spank* promoter; confers Carb^R and Spec^R	Gift from D. Rudner
pLHN12-His$_6$-*Eco* σ^{70}	Encodes full-length wild-type *Eco* σ^{70} bearing an N-terminal His$_6$ tag under the control of a T7 promoter; confers Carb^R	[27]
pLHN12-His$_6$-*Eco* σ$^{70}_{hybrid}$	Encodes full-length *Eco* σ$^{70}_{hybrid}$ bearing an N-terminal His$_6$ tag under the control of a T7 promoter; confers Carb^R	This work
pLHN12-His$_6$-*Eco* σ$^{70}_{quint}$	Encodes full-length *Eco* σ$^{70}_{quint}$ bearing an N-terminal His$_6$ tag under the control of a T7 promoter; confers Carb^R	This work
pLHN12-His$_6$-*Eco* σ^{70} R554A	Encodes full-length *Eco* σ^{70} R554A bearing an N-terminal His$_6$ tag under the control of a T7 promoter; confers Carb^R	This work
pLHN12-His$_6$-*Eco* σ^{70} R554D	Encodes full-length *Eco* σ^{70} R554D bearing an N-terminal His$_6$ tag under the control of a T7 promoter; confers Carb^R	This work
pRLG770-*rrnB*-41	Supercoiled plasmid for *in vitro* transcription reactions; contains *Eco rrnB* P1 promoter sequence from -41 to +50 (no UP element), part of the 5S rRNA transcript, and the native *rrnB* terminators; plasmid origin encodes RNA-I transcript; confers Carb^R	[1]

Table 3.2. List of plasmids used in this study. (Continued)

Plasmid	Relevant details	Reference
pRLG770-*rrnB* -61	Supercoiled plasmid for *in vitro* transcription reactions; contains *Eco rrnB* P1 promoter sequence from -61 to +50 (contains UP element), part of the 5S rRNA transcript, and the native *rrnB* terminators; plasmid origin encodes RNA-I transcript; confers Carb^R	R. Gourse
pRLG770-P$_{R'}$+19 WT-35	pRLG770 derivative bearing native λ P$_{R'}$ promoter sequence from -61 to +15 with wild-type -35 element (TTGACT) and consensus extended -10 promoter positioned 30 bp downstream of P$_{R'}$; plasmid origin encodes RNA-I transcript; confers Carb^R	This work
pRLG770-P$_{R'}$+19 -35 TTGATA (no UP)	pRLG770 derivative bearing native λ P$_{R'}$ promoter sequence from -61 to +15 with weakened -35 element (TTGA<u>TA</u>, substitutions underlined) and consensus extended -10 promoter positioned 30 bp downstream of P$_{R'}$; plasmid origin encodes RNA-I transcript; confers Carb^R	This work
pRLG770-P$_{R'}$+19 -35 TTGATA (*rrnB* P1 UP)	pRLG770 derivative bearing *rrnB* P1 sequence from -61 to -36 (UP element) and P$_{R'}$ promoter sequence from -35 to +15 with weakened -35 element (TTGA<u>TA</u>) and consensus extended -10 promoter positioned 30 bp downstream of P$_{R'}$; plasmid origin encodes RNA-I transcript; confers Carb^R	This work
pRLG770-extended -10	pRLG770 derivative with P$_{R'}$ -35 element mutated but consensus extended -10 promoter positioned 30 bp downstream of P$_{R'}$ intact; plasmid origin encodes RNA-I transcript; confers Carb^R	This work
pRLG770-RNA-I	pRLG770 derivative bearing no test promoter upstream of the *rrnB* terminators but with the plasmid origin-encoded RNA-I transcript intact; confers Carb^R	This work

Plasmids for complementation of* Eco σ^{70} *depletion. Plasmids used to produce *Eco* σ^{70} derivatives bearing substitutions in conserved region 4 to assess their functionality were all derived from the medium-copy number plasmid pBR-*Eco* σ^{70} [17], which encodes full-length, unfused wild-type *Eco* σ^{70} under the control of a weak constitutive synthetic promoter. Conserved region 4 of *Eco* σ^{70} is demarcated by XhoI and HindIII restriction sites in this plasmid, so derivatives of this plasmid encoding *Eco* σ^{70}_{hybrid} or *Eco* σ^{70}_{quint} were generated by amplifying the corresponding σ_4 fragments by PCR from the corresponding α-σ_4 fusion protein plasmids and cloning them into pBR-*Eco* σ^{70} as XhoI/HindIII restriction fragments. All strains carrying these plasmids were grown in LB supplemented with carbenicillin (100 µg/mL).

Plasmids for the expression of unfused gp67. The high-copy number plasmid pCDF*lac*, a derivative of pCDF-1b (Novagen) modified by P. Deighan to introduce the *lacUV5* promoter upstream of the T7 promoter in pCDF-1b, was used to produce unfused gp67. Plasmid pCDF*lac* bears a plasmid-encoded N-terminal His$_6$ tag with a unique AvrII site in the multiple cloning site downstream of the tag and an NdeI site upstream of the plasmid-encoded N-terminal His$_6$ tag. Gp67 was amplified by PCR with primers encoding a C-terminal His$_6$ tag and cloned into pCDF*lac* as an NdeI/AvrII restriction fragment to generate gp67-His$_6$. Strains bearing pCDF*lac* derivatives were grown in LB supplemented with spectinomycin (50 µg/mL).

Plasmids for purification of* Eco σ^{70} *variants. Plasmid pLHN12-His$_6$-*Eco* σ^{70} [27] was used for the production of all *Eco* σ^{70} variants for purification and use in *in vitro* transcription experiments. This plasmid encodes wild-type N-terminally-His$_6$-tagged *Eco* σ^{70} under the control of the T7 promoter, and σ_4 of *Eco* σ^{70} is demarcated by XhoI and HindIII restriction sites. Derivatives of this plasmid encoding *Eco* σ^{70}_{quint}, *Eco* σ^{70} R554A, and *Eco* σ^{70} R554D were generated by amplifying the corresponding σ_4 fragments by PCR and cloning them into pLHN12-His$_6$-*Eco* σ^{70} as XhoI/HindIII restriction fragments. All strains carrying these plasmids were grown in LB supplemented with carbenicillin (100 µg/mL).

***Plasmids for* in vitro *transcription experiments*.** All supercoiled templates for multi-round *in vitro* transcription experiments were constructed using the pRLG770 vector [32], which contains a polylinker for the insertion of promoter fragments, the majority of the *Eco* 5S rRNA transcript, and the transcription terminators from the *rrnB* operon following the 5S gene; in addition, the plasmid ColE1 origin includes a -10/-35 promoter that produces the 108-nt RNA-I transcript [21]. Versions of this plasmid with the *Eco rrnB* P1 promoter either containing (*rrnB* -61) or lacking (*rrnB* -41) its native UP element were a kind gift from W. Ross. We constructed pRLG770 derivatives where the *rrnB* promoter was replaced by: 1) the bacteriophage λ strong promoter $P_{R'}$ (no UP element) and an extended -10 promoter positioned within the $P_{R'}$ transcribed region; 2) a consensus extended -10 promoter; 3) a $P_{R'}$ variant with a weakened -35 element (no UP element) and the extended -10 promoter positioned within the $P_{R'}$ transcribed region; or 4) the $P_{R'}$ variant with a weakened -35 element and the extended -10 promoter positioned within the $P_{R'}$ transcribed region bearing the *rrnB* P1 UP element. We generated each promoter fragment by PCR and cloned each into pRLG770 as EcoRI/HindIII restriction fragments. All of these pRLG770 derivatives include the RNA-I promoter. We also generated a derivative of pRLG770 with no promoter insert (bearing only the RNA-I promoter at the plasmid replication origin) by digesting pRLG770 with EcoRI and HindIII, filling the overhangs using T4 DNA Polymerase, and ligating. All strains carrying these plasmids were grown in LB supplemented with carbenicillin (100 μg/mL).

Functionality of *Eco* σ[70] Derivatives and gp67-Mediated Growth Inhibition

To assess the functionality of the various full-length *Eco* σ[70] variants and gp67-mediated growth inhibition in *E. coli*, the σ[70]-depletion strain NUN449 was transformed with two compatible multicopy plasmids: one encoding wild-type *Eco* σ[70], *Eco* σ[70]$_{hybrid}$, *Eco* σ[70]$_{quint}$, or no protein under the control of a weak constitutive promoter, and another encoding either gp67-His$_6$

or no protein under the control of an IPTG-inducible promoter. NUN449 strains were grown at 37°C in LB supplemented with chloramphenicol (25 μg/mL), carbenicillin (100 μg/mL), and spectinomycin (50 μg/mL) with no IAA or IPTG as sets of serial dilutions. Cultures that were in mid-exponential phase (O.D.$_{600}$ 0.3-0.6) after overnight growth were then serially diluted in fresh LB broth supplemented with chloramphenicol, carbenicillin, and spectinomycin and 5 μL of each of these dilutions spotted onto plates supplemented with chloramphenicol, carbenicillin, and spectinomycin plus 1) either 0 μM or 200 μM IAA and 2) either 0 μM or 20 μM IPTG. Only the NUN449 strain transformed with the plasmid encoding no *Eco* σ^{70} protein required IAA for growth. Plates were incubated at 30°C for fewer than 16 hours.

For the *B. subtilis* experiment, strain CMD37 was grown overnight at 37°C to saturation in LB broth supplemented with spectinomycin (100 μg/mL) in the absence of IPTG. The overnight culture was then serially diluted into fresh medium supplemented with spectinomycin, and 5 μL of each of these dilutions spotted onto plates supplemented with spectinomycin and 0 μM, 50 μM, 100 μM, or 500 μM IPTG. Plates were incubated overnight at 37°C.

Protein Expression and Purification

Wild-type and mutant *Eco* σ^{70} proteins bearing an N-terminal His$_6$ tag were purified from inclusion bodies in BL21(DE3) cells transformed with plasmid pLHN12-His$_6$-*Eco* σ^{70} or its His$_6$-*Eco* σ^{70} mutant derivatives as described [27]. Wild-type *Eco* core RNAP enzyme for *in vitro* transcription experiments was purchased from Epicentre. Gp67-His$_6$ and αΔCTD *Eco* core RNAP enzyme were generous gifts from J. Osmundson and S. Darst.

Native Gel Shift Analysis

The ability of purified gp67 to interact with *Eco* holoenzyme reconstituted with *Eco* core and wild-type *Eco* σ^{70}, *Eco* σ$^{70}_{hybrid}$, or *Eco* σ$^{70}_{quint}$ to form a stable ternary complex was assayed

using native gels. Wild-type *Eco* σ^{70}, *Eco* σ$^{70}_{hybrid}$, or *Eco* σ$^{70}_{quint}$ (2 µM) was incubated with

gp67-His$_6$ (2 µM) for 10 minutes on ice before the addition of *Eco* core RNAP (1 µM).

Complexes were visualized by Coomassie staining on a 4%-12% native PhAST gel using a

PhastSystem unit (GE Healthcare Life Sciences).

In vitro Transcription Assays

Multi-round *in vitro* transcription reactions using supercoiled templates were performed

essentially as described in [30] with some modifications. Wild-type *Eco* σ^{70} or *Eco* σ$^{70}_{quint}$ (40

nM) was pre-incubated with gp67 (0 nM, 50 nM, or 100 nM) for 10 minutes on ice in 1X

transcription buffer [30 mM KCl, 40 mM tris-acetate pH 7.9, 10 mM MgCl$_2$, 1 mM dithiothreitol

(DTT), and bovine serum albumin (BSA) at 100 µg/mL]. Wild-type or αΔCTD *Eco* core RNAP (4

nM) was then added and the mixtures incubated for 10 minutes at 24ºC to form holoenzyme.

Holoenzyme was added to a reaction containing 1X transcription buffer, 50 ng (0.6 nM)

supercoiled plasmid DNA, 500 µM ATP, 200 µM CTP, 200 µM GTP, 10 µM UTP, and [α-^{32}P]-

UTP (Perkin Elmer) to initiate transcription (final reaction volume 20 µL) and the reaction

allowed to proceed at 24ºC for 15 minutes. Reactions were stopped by the addition of 20 µL of

loading buffer (95% (vol/vol) formamide, 20 mM EDTA, 0.05% (wt/vol) bromophenol blue, and

0.05% (wt/vol) xylene cyanol] [7]), heated for 1 minute at 90ºC, and electrophoresed on a 6%

Urea-PAGE gel. Supercoiled plasmid template DNA was purified as described [32]. *In vitro*

transcription reactions using holoenzyme(s) reconstituted with *Eco* σ^{70} R554A or *Eco* σ^{70} R554D

were conducted as described, but with no pre-incubation of *Eco* σ^{70} R554A or *Eco* σ^{70} R554D

with gp67 on ice. Bands were visualized by phosphorimagery and the data were quantified

using ImageQuant software. The data reported for all experiments are the averages of at least

three independent experiments with standard deviations, normalized at each promoter to the

band intensity at 0 nM gp67 (when looking at the effect of gp67 on transcription) or wild-type

Eco σ^{70} (when assessing the effects of substitutions at *Eco* σ^{70} residue R554).

References

1.   **Aiyar, S. E., Gourse, R. L., and Ross, W.** 1998. Upstream A-tracts increase bacterial promoter activity through interactions with the RNA polymerase alpha subunit. Proc Natl Acad Sci U S A **95**(25): 14652-7.

2.   **Browning, D. F. and Busby, S. J.** 2004. The regulation of bacterial transcription initiation. Nat Rev Microbiol **2**(1): 57-65.

3.   **Camacho, A. and Salas, M.** 1999. Effect of mutations in the "extended -10" motif of three *Bacillus subtilis* sigmaA-RNA polymerase-dependent promoters. J Mol Biol **286**(3): 683-93.

4.   **Campbell, E. A.*, et al.*** 2002. Structure of the bacterial RNA polymerase promoter specificity sigma subunit. Mol Cell **9**(3): 527-39.

5.   **Campbell, E. A., Westblade, L. F., and Darst, S. A.** 2008. Regulation of bacterial RNA polymerase sigma factor activity: a structural perspective. Curr Opin Microbiol **11**(2): 121-7.

6.   **Dehbi, M.*, et al.*** 2009. Inhibition of transcription in *Staphylococcus aureus* by a primary sigma factor-binding polypeptide from phage G1. J Bacteriol **191**(12): 3763-71.

7.   **Deighan, P.*, et al.*** 2008. The bacteriophage lambda Q antiterminator protein contacts the beta-flap domain of RNA polymerase. Proc Natl Acad Sci U S A **105**(40): 15305-10.

8.   **Gourse, R. L., de Boer, H. A., and Nomura, M.** 1986. DNA determinants of rRNA synthesis in *E. coli*: growth rate dependent regulation, feedback inhibition, upstream activation, antitermination. Cell **44**(1): 197-205.

9.   **Gross, C. A.*, et al.*** 1998. The functional and regulatory roles of sigma factors in transcription. Cold Spring Harb Symp Quant Biol **63**: 141-55.

10.   **Gruber, T. M. and Gross, C. A.** 2003. Multiple sigma subunits and the partitioning of bacterial transcription space. Annu Rev Microbiol **57**: 441-66.

11.   **Haugen, S. P.*, et al.*** 2006. rRNA promoter regulation by nonoptimal binding of sigma region 1.2: an additional recognition element for RNA polymerase. Cell **125**(6): 1069-82.

12.   **Helmann, J. D.** 1999. Anti-sigma factors. Curr Opin Microbiol **2**(2): 135-41.

13.   **Helmann, J. D.** 1995. Compilation and analysis of *Bacillus subtilis* sigma A-dependent promoter sequences: evidence for extended contact between RNA polymerase and upstream promoter DNA. Nucleic Acids Res **23**(13): 2351-60.

14.   **Hughes, K. T. and Mathee, K.** 1998. The anti-sigma factors. Annu Rev Microbiol **52**: 231-86.

15.   **Igarashi, K. and Ishihama, A.** 1991. Bipartite functional map of the *E. coli* RNA polymerase alpha subunit: involvement of the C-terminal region in transcription activation by cAMP-CRP. Cell **65**(6): 1015-22.

16.   **Kwan, T.*, et al.* 2005. The complete genomes and proteomes of 27 *Staphylococcus aureus* bacteriophages. Proc Natl Acad Sci U S A **102**(14): 5174-9.

17.   **Leibman, M. and Hochschild, A.** 2007. A sigma-core interaction of the RNA polymerase holoenzyme that enhances promoter escape. EMBO J **26**(6): 1579-90.

18.   **Leirmo, S. and Gourse, R. L.** 1991. Factor-independent activation of *Escherichia coli* rRNA transcription. I. Kinetic analysis of the roles of the upstream activator region and supercoiling on transcription of the rrnB P1 promoter in vitro. J Mol Biol **220**(3): 555-68.

19.   **Liu, J.*, et al.* 2004. Antimicrobial drug discovery through bacteriophage genomics. Nat Biotechnol **22**(2): 185-91.

20.   **Lonetto, M. A.*, et al.* 1998. Identification of a contact site for different transcription activators in region 4 of the *Escherichia coli* RNA polymerase sigma70 subunit. J Mol Biol **284**(5): 1353-65.

21.   **Morita, M. and Oka, A.** 1979. The structure of a transcriptional unit on colicin E1 plasmid. Eur J Biochem **97**(2): 435-43.

22.   **Murakami, K. S. and Darst, S. A.** 2003. Bacterial RNA polymerases: the wholo story. Curr Opin Struct Biol **13**: 31-9.

23.   **Murakami, K. S.*, et al.* 2002. Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. Science **296**: 1285-90.

24.   **Nechaev, S. and Severinov, K.** 2003. Bacteriophage-induced modifications of host RNA polymerase. Annu Rev Microbiol **57**: 301-22.

25.   **Osmundson, J.*, et al.* 2012. Promoter-specific transcription inhibition in *Staphylococcus aureus* by a phage protein. Cell **151**(5): 1005-16.

26.   **Paget, M. S. and Helmann, J. D.** 2003. The sigma70 family of sigma factors. Genome Biol **4**(1): 203.

27.   **Panaghie, G.*, et al.* 2000. Aromatic amino acids in region 2.3 of *Escherichia coli* sigma 70 participate collectively in the formation of an RNA polymerase-promoter open complex. J Mol Biol **299**(5): 1217-30.

28.   **Paul, B. J.*, et al.* 2004. rRNA transcription in *Escherichia coli*. Annu Rev Genet **38**: 749-70.

29.   **Rao, L.*, et al.* 1994. Factor independent activation of *rrnB* P1. An "extended" promoter with an upstream element that dramatically increases promoter strength. J Mol Biol **235**(5): 1421-35.

30.   **Ross, W.*, et al.* 1993. A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. Science **262**: 1407-13.

31.   **Ross, W. and Gourse, R. L.** 2005. Sequence-independent upstream DNA-alphaCTD interactions strongly stimulate *Escherichia coli* RNA polymerase-*lacUV5* promoter association. Proc Natl Acad Sci U S A **102**(2): 291-6.

32.   **Ross, W.*, et al.* 1990. *E.coli* Fis protein activates ribosomal RNA transcription *in vitro* and *in vivo*. EMBO J **9**(11): 3733-42.

33.   **Voskuil, M. I. and Chambliss, G. H.** 1998. The -16 region of *Bacillus subtilis* and other gram-positive bacterial promoters. Nucleic Acids Res **26**(15): 3584-90.

34.   **Wosten, M. M.** 1998. Eubacterial sigma-factors. FEMS Microbiol Rev **22**(3): 127-50.

35.   **Youngman, P. J., Perkins, J. B., and Losick, R.** 1983. Genetic transposition and insertional mutagenesis in *Bacillus subtilis* with *Streptococcus faecalis* transposon Tn917. Proc Natl Acad Sci U S A **80**(8): 2305-9.

Chapter 4:

Summary and Future Directions

Summary

The striking sequence and structural conservation of multisubunit RNA polymerases (RNAPs) across all domains of life suggests that detailed mechanistic studies using experimentally tractable bacterial model systems can provide global insight into gene expression and its regulation. Transcription initiation, the first step in gene expression, is a key regulatory step in bacteria and higher organisms that ensures the coordinated expression of the appropriate specific sets of genes under any environmental or developmental condition. In bacteria, σ factors are involved in every step of transcription initiation: they make extensive contacts with various subunits of the core RNAP enzyme (subunit composition $\alpha_2\beta\beta'\omega$) to form the initiation-competent holoenzyme (subunit composition $\alpha_2\beta\beta'\omega\sigma$) [reviewed in 10, 21, 28], are responsible for the sequence-specific recognition of bacterial promoters by holoenzyme [reviewed in 10, 21, 28], and are directly involved in promoter melting and open complex formation [8, 19, 24]. Bacterial genomes generally encode multiple σ factors: a single, essential primary σ factor that directs the expression of housekeeping genes and the majority of genes during exponential growth, and one or more alternative σ factors that direct the expression of specific regulons [reviewed in 11, 23, 28]. The interplay between the various σ factors determines the cell's transcriptional profile at any time.

The α subunits of RNAP also have important roles in transcription initiation. The C-terminal domains of these subunits (αCTDs) can interact with activators bound upstream of the core promoter to stimulate transcription initiation by facilitating the recruitment of holoenzyme to specific promoters or stabilizing the holoenzyme/promoter complex [reviewed in 4]. Additionally, some promoters contain an A+T-rich sequence motif upstream of the -35 element called the UP element that is specifically recognized by the αCTDs. Several promoters—including the ribosomal RNA *rrn* promoters [reviewed in 25]—depend on sequence-specific αCTD/UP element contacts for full activity, but the αCTDs can also participate in sequence-nonspecific

interactions with upstream DNA that facilitate holoenzyme recruitment to promoters lacking UP elements [27]. The integral roles of σ factors and the α subunits in transcription initiation make them excellent targets for regulation. Bacteria and their infecting bacteriophages have evolved many strategies to modulate the activities of σ factors and the α subunits to enact particular transcriptional programs. Anti-σ factors, regulators that bind tightly and specifically to their cognate σ factor to inhibit transcription from promoters dependent on that σ factor for recognition [reviewed in 5, 12, 13], are key players in transcription regulation used by both bacteria and phages. Many (e.g. the *E. coli*-encoded stationary phase factor Rsd) function primarily by sequestering their cognate σ factor and preventing holoenzyme formation, but others (e.g. the *E. coli*-encoded anti-σ^{28} factor FlgM or the coliphage T4-encoded anti-σ^{70} AsiA) exert their inhibitory functions as components of the holoenzyme [reviewed in 5]. Prior to this work, Rsd and AsiA were the only two anti-σ factors specific to a primary σ that had been characterized. The differences in their mechanisms of action raised the possibility of the existence of a broader class of primary σ factor-specific anti-σ factors that may help guide the development of novel small molecule-derived prophylactic or therapeutic alternatives to antibiotics. The largely unexplored proteomic landscape resulting from the sequencing of a large number of phage genomes provides an excellent source for potential novel anti-σ factors and other transcription regulators.

Bacteriophage G1 gp67 is one example of such a novel anti-σ factor: it was identified as an inhibitor of growth and RNA synthesis in *S. aureus* [16, 17] and subsequently shown to interact directly with *Sau* σ^A (but not with *Eco* σ^{70}) via conserved region 4 to form a stable ternary complex with the σ^A-containing *Sau* holoenzyme [7, 22]. Osmundson *et al.* [22] solved the high-resolution X-ray crystal structure of the *Sau* σ^A_4/gp67 complex to provide insight into the mechanism of action of gp67. In **Chapter 2** of this dissertation, I conducted an unbiased structure-independent genetic analysis to thoroughly characterize the *Sau* σ^A_4/gp67 interface

and investigate the specificity of gp67 for Sau σ^A_4 binding. I constructed a series of Eco σ^{70}_4/Sau σ^A_4 chimeras and used the bacterial two-hybrid system to assess the ability of each of these chimeras to bind gp67. In this analysis, I established that the region spanning Sau σ^A residues 309-335 (corresponding to Eco σ^{70} residues 553-579) determines the specificity of gp67 for Sau σ^A_4. By making $Eco \leftrightarrow Sau$ substitutions at each of the divergent residues in this region, I further identified the side chains of Sau σ^A residues D309, E312, N313, and V335 (corresponding to Eco σ^{70} residues A553, A556, K557, and Q579) as important for gp67 binding specificity. These data are in agreement with the crystallographic data presented in [22] and provide independent support for the biological relevance of the crystallized Sau σ^A_4/gp67 complex. Further, these results establish the viability of developing Eco-based *in vivo* and *in vitro* systems to study gp67.

I extended the genetic analysis to the primary σ factors of other Gram-positive and Gram-negative bacteria and used the two-hybrid system to assess the abilities of the primary σ factors of *B. subtilis* (Bsu σ^A, which is identical to Sau σ^A at the four identified gp67 specificity determinants) and *T. aquaticus* (Taq σ^A, which differs from both Eco σ^{70}_4 and Sau σ^A_4 at three of these positions) to bind gp67. As expected, Bsu σ^A_4 bound gp67 in a manner dependent on the identity of the residues corresponding to Sau σ^A D309, E312, N313, and V335 and Taq σ^A_4 failed to interact with gp67, though the instability of the quadruply-substituted α-Taq σ^A_4 fusion protein (bearing Sau residues at the positions corresponding to Sau σ^A D309, E312, N313 and V335) did not allow me to determine whether these substitutions would enable Taq σ^A_4 to bind gp67. The ability of Bsu σ^A_4 to bind gp67 suggests that Bsu may provide an *in vivo* and *in vitro* system for the detailed mechanistic study of gp67 that is more experimentally tractable than Sau but is less heterologous than the Eco system discussed in Chapter 3.

Using genetics and a structure-based candidate approach, I also sought to identify gp67 residues involved in specifying its binding to Sau σ^A_4, with the goal of generating an otherwise

functional nonbinding mutant of gp67 to use as a control in subsequent *in vivo* and *in vitro* experiments. I introduced charge reversals at two positions in gp67 (K2 and K195) that form salt bridges with *Sau* σ^A_4 E312 (one of the genetically-identified gp67 specificity determinants) and attempted to construct a mutant-suppressor pair to assess the functional importance of these interactions. I also substituted two residues of gp67 (Y37 and N41) that are part of a pocket into which *Sau* σ^A_4 residue R310—part of the identified gp67 specificity-determining region but conserved between *Eco* and *Sau*—is tucked upon gp67 binding. Substitutions at each of these gp67 residues abolished *Sau* σ^A_4 binding but also altered the electrophoretic mobility of the gp67 mutants, so I could not evaluate whether these substitutions directly affected contacts at the gp67/*Sau* σ^A_4 interface or affected binding indirectly due to altered protein folding.

Biochemical and crystallographic data indicate that gp67 forms a ternary complex with the σ^A-containing *Sau* holoenzyme and that the binding of gp67 to *Sau* σ^A_4 does not induce conformational rearrangements in *Sau* σ^A_4 [7, 22]. Two structural predictions follow from these observations: 1) gp67-bound *Sau* σ^A_4 can interact with its primary binding determinant on core RNAP, the flap-tip helix (FTH) on the β subunit; and 2) gp67 binding does not interfere with -35 element recognition. I used modified versions of the two-hybrid and one-hybrid assays that incorporated an additional plasmid directing the synthesis of either unfused *Sau* σ^A_4 or gp67 to directly test these structural predictions *in vivo*. I found that, as expected, the *Sau* σ^A_4/gp67 complex bound the β-flap in a FTH-dependent manner. Unexpectedly, I also observed that unfused gp67 disrupted the recognition of an ectopically positioned -35 element by an α-σ_4 fusion protein specifically when the σ_4 moiety bore the appropriate gp67 specificity-determining residues. I demonstrated that transcription activation in the one-hybrid system depends on the interaction between *Eco* σ^{70} residue R554 (*Sau* σ^A residue R310) and the DNA phosphate backbone immediately upstream of the -35 element. This residue is the only residue in σ_4 implicated in both DNA recognition and the interface with gp67, which strongly suggested that

the unexpected inhibitory effect of gp67 in this system was due to the loss of the contact

between *Sau* σ^A R310 and the phosphate backbone. Additional *in vitro* transcription data from

Chapter 3 suggest that this contact may be important in initiation at several *Eco* promoters and

that, in the context of *Eco* RNAP, gp67 may thus affect -35 element recognition.

In experiments conducted using a native *Sau* system, Osmundson *et al.* [22]

demonstrated that gp67 exhibits specificity for the *Sau* rRNA promoters and other promoters

with A+T-rich UP element-like upstream regions, and that this selectivity is the result of gp67

disrupting contacts between the αCTDs and upstream DNA necessary for efficient initiation at

these promoters. In **Chapter 3** of this dissertation, I developed a gp67-responsive *Eco*-based

transcription system to take advantage of the wealth of well-characterized genetic and

biochemical tools and reagents available in *Eco* for detailed mechanistic studies of transcription.

I demonstrated that gp67 inhibited growth of an *Eco* σ^{70}-depletion strain specifically when the

available primary σ factor bore the appropriate gp67 specificity determinants. Gp67 was also

toxic in *Bsu*, consistent with its ability to bind *Bsu* σ^A_4 in the two-hybrid system. Furthermore,

gp67 formed a stable ternary complex with holoenzyme reconstituted with *Eco* core RNAP and

Eco σ^{70} modified to bear the appropriate gp67 specificity determinants in region 4. Together,

these data suggested that gp67 is functional in the modified *Eco* system and that this

heterologous system could be mechanistically informative.

I then used the gp67-responsive *Eco* transcription system and took advantage of the

well-characterized UP element-containing *Eco* rRNA promoter *rrnB* P1 and a version of *Eco*

core RNAP lacking the αCTDs (αΔCTD RNAP core) [14] to address whether gp67 exerted

promoter (and UP element)-specific effects in the context of the *Eco* system. I performed multi-

round *in vitro* transcription reactions using supercoiled plasmid DNA templates bearing a variety

of promoters either containing or lacking an UP element: two otherwise-identical versions of

rrnB P1 with (*rrnB* -61) or without (*rrnB* -41) its native UP element; a version of the

bacteriophage λ $P_{R'}$ promoter with a consensus extended -10 promoter within the $P_{R'}$ transcript

and lacking an UP element; two otherwise-identical versions of $P_{R'}$ with a weakened -35 element

and either bearing ($P_{R'}$* UP) or lacking ($P_{R'}$* no UP) the *rrnB* P1 UP element; a consensus

extended -10 promoter (with no UP element); and the RNA-I promoter, a -10/-35 promoter

lacking an UP element. I found that gp67 significantly inhibited transcription from all -10/-35

promoters tested (except $P_{R'}$, where the effect was modest) specifically when holoenzyme was

reconstituted with wild-type core RNAP and a version of *Eco* σ^{70} that binds gp67 (*Eco* σ^{70}_{quint}),

irrespective of the presence or absence of the UP element. Gp67 had no effect on transcription

from the extended -10 promoter, indicating that it does not inhibit transcription by sequestering

its cognate σ factor and preventing holoenzyme formation. Furthermore, I found that the extent

to which the inhibitory effects of gp67 at -10/-35 promoters depended on the presence of the

αCTDs varied: inhibition at the RNA-I promoter was almost exclusively αCTD-dependent, but

the effects at the other promoters were largely αCTD-independent. Together, these data

suggested that sequence-nonspecific interactions of the αCTDs with upstream DNA can be

important for efficient transcription initiation at promoters lacking UP elements, and that gp67

can exert both αCTD-dependent and αCTD-independent effects on transcription that depend on

promoter context.

The observations that weakening the -35 element of $P_{R'}$ increased its sensitivity to gp67-

mediated inhibition and that unfused gp67 inhibited the binding of σ_4 to the -35 element as

detected in the one-hybrid assay (see Chapter 2) suggested that the αCTD-independent effects

of gp67 may be the result of compromised -35 element recognition due to the loss of an

energetically significant contact between *Eco* σ^{70} R554 (*Sau* σ^{A} R310) and the phosphate

backbone at position -36. I addressed this by assessing the effect of substitutions at *Eco* σ^{70}

residue R554 on transcription from the *rrnB* -41, *rrnB* -61, extended -10, and RNA-I promoters. I

observed a significant reduction in transcription from the *rrnB* -41, *rrnB* -61, and RNA-I

promoters but not from the extended -10 promoter, indicating that the *Eco* σ^{70} R554A and

R554D proteins were functional but defective in initiation from these -10/-35 promoters. These

data suggested that the *Eco* σ^{70} R554/-36 backbone contact is important for initiation at these

promoters and that the disruption of this contact by gp67 binding likely accounts for the

observed αCTD-independent inhibitory effects. The *Eco* σ^{70} R554/-36 backbone contact also

likely contributes significantly to efficient -35 element recognition at the RNA-I promoter, where I

previously observed αCTD-dependent inhibitory effects of gp67. Disruption of this contact by

either gp67 binding or substitutions at *Eco* σ^{70} R554 reduces initiation at the RNA-I promoter, but

gp67 may make stabilizing sequence-nonspecific contacts with upstream DNA at this promoter

in the absence of the αCTDs that compensate for the transcription defect caused by the loss of

the *Eco* σ^{70} R554/-36 backbone contact.

The results presented in this dissertation describe the usefulness of genetic analyses in

probing the biological relevance of crystallized protein complexes and in developing

heterologous systems for detailed mechanistic studies of novel transcription regulators from

organisms that may be experimentally intractable. These approaches allowed us to utilize the

wealth of tools and reagents available in *E. coli* to directly test structural and mechanistic

predictions about the behavior of gp67, a phage protein proposed to present a novel regulatory

paradigm: forming a stable complex with the RNAP holoenzyme by interacting with the σ

subunit but exerting promoter selectivity by antagonizing the function of the α subunit. Our *E.

coli*-based experimental system revealed an unexpected complexity in the mechanism of gp67

action. In particular, we found that gp67 exerted effects that were promoter context-dependent

and involved both the σ and α subunits of RNAP. Future studies will be required to determine

whether or not the current picture of gp67 function in *Sau* is incomplete. Nonetheless, our

findings with gp67 in the *E. coli*-based system have highlighted aspects of promoter function that are worthy of further study.

Future Directions

Development of a *B. subtilis* transcription system for the study of gp67

Results from Chapters 2 and 3 of this dissertation indicate that gp67 exhibits less stringent promoter selectivity in an *E. coli* transcription system than was observed in its native host, and, additionally, exerts αCTD-independent inhibitory effects that likely result from disrupting -35 element recognition. These discrepancies between the behavior of gp67 in *Sau* and *Eco* could be a function of the evolutionary distance between the two species that results in fine-structure differences in the transcription apparatus (RNAP and/or promoter architecture). To address this question, we became interested in developing a gp67-responsive *B. subtilis* transcription system for *in vivo* and *in vitro* experiments. This experimental system would be less heterologous than the *Eco* system (due to the closer evolutionary distance between *Sau* and *Bsu*) but still have the benefit of well-developed genetic and biochemical tools (e.g. well-characterized promoters, effective protocols for RNAP purification, a well-established *in vitro* transcription system) that are not readily available in *Sau*. As in the *Sau* and *Eco* systems, gp67 should inhibit growth and form the functionally relevant ternary complex in *Bsu* in order for this system to be mechanistically informative. We have already demonstrated gp67-mediated growth inhibition in *Bsu* and would need to conduct native gel shift analyses using purified *Bsu* RNAP, *Bsu* σ^A, and gp67 to confirm the expected formation of the $E_{Bsu} \cdot \sigma^A$/gp67 complex before looking at the effect of gp67 on transcription by $E_{Bsu} \cdot \sigma^A$ *in vitro* from a variety of promoters containing or lacking UP elements. If the *Bsu* system closely recapitulates the reported behavior of gp67 in

Sau, we would expect to observe strictly UP element- and αCTD-dependent inhibitory effects of gp67 on transcription by $E_{Bsu} \cdot \sigma^A$. (Given the relative ease of purification of *Bsu* RNAP (e.g. [2]), it should be possible to purify *Bsu* αΔCTD core RNAP for *in vitro* transcription experiments).

Studies of the *Bsu* σ^A-dependent *rrnB* P1 promoter revealed that its native UP element results in only 2-3-fold stimulation of transcription *in vivo* and 1.3-1.6-fold stimulation *in vitro* [15] (versus the 20-50-fold stimulatory effect of the UP element at *Eco rrnB* P1 *in vitro* [26]); these weak UP element effects are likely the result of the extended -10 motif and optimal 17 bp spacer present in *Bsu rrnB* P1 that afford a strong core promoter. Based on our success in making the $P_{R'}$ promoter responsive to the presence of an UP element in *Eco* by weakening its -35 element, we could introduce similar substitutions into the *Bsu rrnB* P1 -35 element to generate a version of the promoter where the UP element stimulates transcription more strongly, and then use otherwise identical versions of this weakened *Bsu rrnB* P1 promoter either bearing or lacking the UP element to assess the UP element dependence of gp67-mediated inhibition in *Bsu*. In these experiments, we would expect gp67 to specifically inhibit transcription from the weakened *Bsu rrnB* P1 promoter bearing the UP element. *Bsu* RNAP does not efficiently transcribe many *Eco* σ^{70}-dependent promoters *in vitro* [20], so it is unlikely that we would be able to use the RNA-I promoter present in our supercoiled plasmid templates as a control -10/-35 promoter lacking an UP element; however, we could use the core promoter regions of characterized σ^A-dependent promoters lacking the extended -10 motif (e.g. P_{veg} or $P_{gcaD/tms}$). Additionally, we could fuse the UP element of the $\sigma^{D(28)}$-dependent *hag* promoter to P_{veg} or $P_{gcaD/tms}$ to generate non-rRNA, strongly UP element-responsive promoters [6]. If gp67 exhibits the same stringent promoter selectivity in *Bsu* as has been reported in *Sau*, we should only observe transcription inhibition at these promoters in the presence of the *hag* UP element.

The *Bsu* system would also allow us to look at the effect of gp67 on transcription *in vivo* without the technical complications of working with the *Eco* σ^{70}-depletion strain. To look at the

effect of gp67 on UP element utilization, we could integrate single-copy promoter-*lacZ* fusions containing or lacking an UP element into the *Bsu* chromosome and assess the effect of gp67 production on *lacZ* transcription by a β-galactosidase assay or quantitative real-time PCR. Preliminary attempts to address this question in *Eco* using versions of NUN449 with λ lysogens bearing *rrnB* -41-*lacZ* or *rrnB* -61-*lacZ* fusions (constructed by T. Gaal) revealed no significant effect of gp67 on transcription (not shown), but these analyses are complicated by the potential effect of gp67 at many -10/-35 promoters that may include the reference transcript used for normalization (*clpX*). If our *in vitro* transcription experiments in *Bsu* reveal a strict UP element and αCTD dependence of gp67-mediated transcription inhibition, the *Bsu* system would allow us to circumvent these problems and conduct a variety of *in vivo* experiments.

Identifying promoter characteristics that confer susceptibility to gp67

RNA-seq experiments performed in *Sau* to identify gp67-sensitive promoters genome-wide revealed an effect of gp67 on ~9% of total *Sau* promoters, with 4% of all promoters inhibited [22]. These promoters were enriched in A+T-rich sequences upstream of the -35 element, leading to the model that gp67 specifically inhibits transcription at promoters where UP element-like sequences are required for full activity [22]. Conventional approaches to global gene expression analysis assume that the total RNA content per cell is similar under each condition tested (absence or presence of gp67 in this case) and, thus, that using equivalent amounts of total RNA for normalization accurately reflects the relative levels of each transcript in the two conditions [18]. Crystallographic data of the *Sau* σ^A_4/gp67 complex that suggested that gp67 binding would not disrupt -35 element recognition by *Sau* σ^A_4 [22] made this a reasonable assumption: the majority of *Sau* promoters are -10/-35 promoters, and if gp67 does not interfere with -35 element recognition, it would be expected to affect transcription from only a small subset of promoters where -35 element recognition is not the most significant contributor to

transcription initiation. In our *Eco* transcription system, gp67 has inhibitory effects that are independent of the UP element and the αCTDs and likely result from effects on -35 element recognition by σ_4, which suggests that gp67 may have a more widespread effect on transcription and that standard normalization techniques may give an incomplete picture of the number and type(s) of promoters that are susceptible to gp67.

Thus, we want to conduct an analogous global transcriptome analysis using RNA-seq in the *Eco* σ^{70}-shutoff strain NUN449 bearing *Eco* σ^{70}_{quint} and in *Bsu* in the absence and presence of gp67, using spiked-in RNA standards that allow for normalization to cell number [18] to obtain interpretable results. These experiments would help identify promoter characteristics that specify gp67 susceptibility in both a Gram-positive and a Gram-negative organism and determine whether the more widespread effects of gp67 when tested *in vitro* with our modified *E. coli*-based components (relative to the *Sau* and, presumably, *Bsu* components) occur genome-wide. One possible characteristic of promoters that are especially susceptible to gp67 is the importance of the interaction between σ_4 and the DNA backbone at the -36 position to -35 element recognition and transcription initiation. Promoters where this σ_4/backbone contact is an important determinant of promoter binding would be particularly sensitive to gp67, since binding of gp67 to σ_4 redirects a conserved arginine residue at position 554 in *Eco* σ^{70} (corresponding to R310 in *Sau* σ^A and R313 in *Bsu* σ^A) away from the σ_4/DNA interface and into a pocket in gp67 [22]. This inhibitory effect of gp67 would be mediated through σ_4 and is completely independent of the αCTDs.

Once candidate gp67-responsive promoters have been identified by RNA-seq, we can use the *Eco* and *Bsu in vitro* transcription systems to determine whether the effects of gp67 are direct, dependent on the αCTDs (using the *Eco*—and potentially *Bsu*—αΔCTD core RNAP), or mediated through the σ_4/backbone contact. In the *Eco* system, we expect substitutions at σ^{70} R554 to phenocopy the inhibitory behavior of gp67 at promoters where the σ_4/backbone contact

contributes significantly to promoter binding, and for these substitutions to have no effect on transcription at promoters where this interaction is not a key determinant of promoter binding (gp67 should inhibit transcription by $E \cdot \sigma^{70}_{quint}$ R554A(D) at these promoters, however). The existence of this latter class of promoters would provide strong support for the idea that gp67 can exert both αCTD-dependent and αCTD-independent (mediated by σ_4) transcription inhibitory effects. Additionally, if the proposed *in vitro* transcription experiments in *Bsu* reveal a strict αCTD dependence of gp67-mediated inhibition, the RNA-seq analysis in *Bsu* would allow us to use gp67 as a tool on a global scale to identify promoters where sequence-nonspecific interactions of the αCTDs with upstream DNA are important for initiation. (The same analysis in *Eco* is complicated by the αCTD-independent effects of gp67 on transcription).

DNase I footprinting of gp67-sensitive promoters in *E. coli*

DNase I footprinting experiments at the *Sau rrnA* promoter—where Osmundson *et al.* observed UP element-dependent (and presumably αCTD-dependent) inhibition by gp67— revealed that gp67 alters the interaction of *Sau* RNAP with DNA upstream of the -35 element [22]. The DNase I footprint of *Eco* RNAP at the *rrnB* P1 promoter has been thoroughly investigated (e.g. [1, 3, 9, 26]), so conducting these analyses at *rrnB* -41, *rrnB* -61, and the $P_{R'}$ promoters with the weakened -35 element ($P_{R'}$*) bearing or lacking the UP element could provide insight into the effects of gp67 on RNAP binding at these promoters. In the *Eco in vitro* transcription system, we observed significant gp67-mediated inhibition at *rrnB* -61 and P_R* UP, though this inhibition appeared to be largely αCTD-independent. Thus, it would be of great interest to determine whether, in fact, gp67 disrupts the αCTD/UP element contacts of $E \cdot \sigma^{70}_{quint}$ at these promoters by comparing the footprint of this holoenzyme in the absence and presence of gp67. If inhibitory effects are mediated by interfering with -35 element recognition, the footprints in the absence and presence of gp67 would likely be the same. The same would be

173

true of the footprints of EαΔCTD•σ^{70}$_{quint}$ at *rrnB* -61 and P$_R$·* UP in the absence and presence of gp67, though there would be no protection of upstream sequences even in the absence of gp67 due to the lack of the αCTDs. The footprints at *rrnB* -41 and P$_R$·* no UP would be expected to look the same for both E•σ^{70}$_{quint}$ and EαΔCTD•σ^{70}$_{quint}$ in the absence or presence of gp67, and to resemble those of EαΔCTD•σ^{70}$_{quint}$ at the promoters bearing the UP element. Differences from these predictions would suggest that gp67 has some effect on the interaction of RNAP with upstream DNA at these promoters.

DNase I footprinting at the RNA-I promoter would allow us to directly address whether the αCTDs make sequence-nonspecific contacts with upstream DNA at this promoter that are important for initiation—as suggested by the *in vitro* transcription results—and whether gp67 inhibits transcription by disrupting these contacts. Comparing the footprints of E•σ^{70}$_{quint}$ in the absence and presence of gp67 would answer this question: if the αCTDs make nonspecific contacts with upstream DNA at the RNA-I promoter, we would expect to see protection of these sequences in the footprint in the absence of gp67 and the loss of this protection in the presence of gp67. The footprint of EαΔCTD•σ^{70}$_{quint}$ in the absence of gp67 should look the same as the E•σ^{70}$_{quint}$ + gp67 footprint, since the absence of the αCTDs would also result in the loss of protection of upstream sequences. Additionally, the footprint of EαΔCTD•σ^{70}$_{quint}$ in the presence of gp67 would allow us to determine whether gp67 can make sequence-nonspecific contacts with upstream DNA in the context of the ternary complex, as suggested by: 1) the appearance of novel bands in all *in vitro* transcription reactions performed with EαΔCTD•σ^{70}$_{quint}$ in the presence of gp67 (Chapter 3); 2) the proximity of gp67 to upstream DNA in structural models and the basic nature of the DNA-facing surface of gp67 [22]; and 3) the comparison of the effects of gp67 and substitutions at *Eco* σ^{70} R554 on RNA-I transcription *in vitro* (where compensatory protein/DNA contacts by gp67 may stabilize *Eco* RNAP holoenzyme binding at this promoter and manifest as a reduction in inhibition; Chapter 3). These potential gp67/DNA

contacts may be facilitated by the lack of competition with the αCTDs in EαΔCTD•σ$^{70}_{quint}$ and

would be evidenced by some restoration of protection of upstream sequences when compared

to the E•σ$^{70}_{quint}$ + gp67 or EαΔCTD•σ$^{70}_{quint}$ - gp67 footprints. If gp67 contacts upstream DNA

nonspecifically and this is functionally relevant in the context of the phage life cycle, it is likely

that structural differences between *Eco* and *Sau* RNAPs allow gp67 to engage in these contacts

at some promoters in *Sau* in the presence of the αCTDs. (We observe the appearance of novel

bands in the *Eco* system only with the αΔCTD core, suggesting that the αCTDs readily

outcompete gp67 for access to upstream DNA in the *Eco* enzyme; this would evidently not be

the case with the *Sau* enzyme if these gp67/DNA contacts are biologically relevant since the

αCTDs are necessarily present *in vivo*).

Investigating the importance of the *Sau* σ^A R310/backbone contact in transcription initiation

One possible explanation for the apparent differences in the behavior of gp67 in the *Eco*

and *Sau* transcription systems—αCTD-independent inhibition and less stringent promoter

selectivity in *Eco*—is that the σ$_4$ R554 (R310)/backbone contact contributes more significantly to

transcription initiation at *Eco* promoters than at *Sau* promoters. If this is the case, we would

expect to see a more widespread effect of gp67 on transcription in *Eco* in the RNA-seq

experiments. The lack of a significant contribution of the σ^A R310/backbone contact to promoter

binding in *Sau* would indicate that gp67 does not exert any transcription inhibitory effects

through σ$^A_4$ and that all transcription inhibition is mediated by interfering with αCTD function.

Since there is no currently available *Sau* αΔCTD core RNAP, we cannot test the αCTD

dependence of gp67-mediated transcription inhibition in *Sau* directly. However, we can

determine whether the σ^A R310/ backbone contact contributes to promoter recognition in *Sau* by

looking at the effect of substitutions at residue R310 on transcription from gp67-sensitive

promoters (e.g. *rrnA*, *rrnB*, *csp1* [22]). If this contact does not contribute significantly to

175

transcription from the majority of *Sau* promoters, we would expect transcription from all identified gp67-sensitive promoters to be <u>unaffected</u> by substitutions at *Sau* σ^A R310, which would suggest that gp67 does not exert any inhibitory effects through σ_4 in *Sau*. [Note that gp67 would still inhibit transcription by E•σ^A R310A(D) at these promoters in an αCTD-dependent manner, since these substitutions do not affect gp67 binding in the context of the two-hybrid assay; not shown]. However, if we do observe an effect of the R310 substitutions on transcription from at least some of these promoters, this would suggest that gp67 has some <u>σ_4-dependent</u> inhibitory effects in *Sau*. The results of the RNA-seq analysis discussed above, which indicated that only ~9% of Sau promoters are gp67-susceptible [22], imply either that the σ^A R310/backbone contact is an important determinant of promoter binding at no more than a small subset of *Sau* promoters or that gp67 can make compensatory contacts with upstream DNA, thus masking the effect of the loss of the σ^A R310/backbone contact.

In the absence of *in vitro* transcription data with a *Sau* αΔCTD core RNAP or with *Sau* σ^A R310 mutants, we cannot exclude the possibility that gp67 may mediate some of its inhibitory effects on -35 element recognition by σ^A_4 in *Sau* despite data suggesting that gp67 functions exclusively by disrupting αCTD function. If *in vitro* transcription analysis with the *Sau* σ^A R310 mutants identifies any gp67-sensitive promoters where the *Sau* σ^A R310/backbone contact contributes to promoter activity, it may be worthwhile to repeat the RNA-seq analysis in *Sau* using spike-ins for normalization [18] to identify potential αCTD-independent susceptible promoters in the native host.

References

1.   **Aiyar, S. E., Gourse, R. L., and Ross, W.** 1998. Upstream A-tracts increase bacterial promoter activity through interactions with the RNA polymerase alpha subunit. Proc Natl Acad Sci U S A **95**(25): 14652-7.

2.   **Anthony, L. C.*, et al.* 2000. Rapid purification of His(6)-tagged *Bacillus subtilis* core RNA polymerase. Protein Expr Purif **19**(3): 350-4.

3.   **Blatter, E. E.*, et al.* 1994. Domain organization of RNA polymerase alpha subunit: C-terminal 85 amino acids constitute a domain capable of dimerization and DNA binding. Cell **78**(5): 889-96.

4.   **Browning, D. F. and Busby, S. J.** 2004. The regulation of bacterial transcription initiation. Nat Rev Microbiol **2**(1): 57-65.

5.   **Campbell, E. A., Westblade, L. F., and Darst, S. A.** 2008. Regulation of bacterial RNA polymerase sigma factor activity: a structural perspective. Curr Opin Microbiol **11**(2): 121-7.

6.   **Caramori, T. and Galizzi, A.** 1998. The UP element of the promoter for the flagellin gene, *hag*, stimulates transcription from both SigD- and SigA-dependent promoters in *Bacillus subtilis*. Mol Gen Genet **258**(4): 385-8.

7.   **Dehbi, M.*, et al.* 2009. Inhibition of transcription in *Staphylococcus aureus* by a primary sigma factor-binding polypeptide from phage G1. J Bacteriol **191**(12): 3763-71.

8.   **Feklistov, A. and Darst, S. A.** 2011. Structural basis for promoter-10 element recognition by the bacterial RNA polymerase sigma subunit. Cell **147**(6): 1257-69.

9.   **Gaal, T.*, et al.* 1996. DNA-binding determinants of the alpha subunit of RNA polymerase: novel DNA-binding domain architecture. Genes Dev **10**(1): 16-26.

10.   **Gross, C. A.*, et al.* 1998. The functional and regulatory roles of sigma factors in transcription. Cold Spring Harb Symp Quant Biol **63**: 141-55.

11.   **Gruber, T. M. and Gross, C. A.** 2003. Multiple sigma subunits and the partitioning of bacterial transcription space. Annu Rev Microbiol **57**: 441-66.

12.   **Helmann, J. D.** 1999. Anti-sigma factors. Curr Opin Microbiol **2**(2): 135-41.

13. **Hughes, K. T. and Mathee, K.** 1998. The anti-sigma factors. Annu Rev Microbiol **52**: 231-86.

14. **Igarashi, K. and Ishihama, A.** 1991. Bipartite functional map of the *E. coli* RNA polymerase alpha subunit: involvement of the C-terminal region in transcription activation by cAMP-CRP. Cell **65**(6): 1015-22.

15. **Krasny, L. and Gourse, R. L.** 2004. An alternative strategy for bacterial ribosome synthesis: *Bacillus subtilis* rRNA transcription regulation. EMBO J **23**(22): 4473-83.

16. **Kwan, T.*, et al.*** 2005. The complete genomes and proteomes of 27 *Staphylococcus aureus* bacteriophages. Proc Natl Acad Sci U S A **102**(14): 5174-9.

17. **Liu, J.*, et al.*** 2004. Antimicrobial drug discovery through bacteriophage genomics. Nat Biotechnol **22**(2): 185-91.

18. **Loven, J.*, et al.*** 2012. Revisiting global gene expression analysis. Cell **151**(3): 476-82.

19. **Marr, M. T. and Roberts, J. W.** 1997. Promoter recognition as measured by binding of polymerase to nontemplate strand oligonucleotide. Science **276**: 1258-60.

20. **Moran, C. P., Jr.*, et al.*** 1982. Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis*. Mol Gen Genet **186**(3): 339-46.

21. **Murakami, K. S. and Darst, S. A.** 2003. Bacterial RNA polymerases: the wholo story. Curr Opin Struct Biol **13**: 31-9.

22. **Osmundson, J.*, et al.*** 2012. Promoter-specific transcription inhibition in *Staphylococcus aureus* by a phage protein. Cell **151**(5): 1005-16.

23. **Paget, M. S. and Helmann, J. D.** 2003. The sigma70 family of sigma factors. Genome Biol **4**(1): 203.

24. **Panaghie, G.*, et al.*** 2000. Aromatic amino acids in region 2.3 of *Escherichia coli* sigma 70 participate collectively in the formation of an RNA polymerase-promoter open complex. J Mol Biol **299**(5): 1217-30.

25. **Paul, B. J.*, et al.*** 2004. rRNA transcription in *Escherichia coli*. Annu Rev Genet **38**: 749-70.

26.   **Ross, W., et al.** 1993. A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. Science **262**: 1407-13.

27.   **Ross, W. and Gourse, R. L.** 2005. Sequence-independent upstream DNA-alphaCTD interactions strongly stimulate *Escherichia coli* RNA polymerase-*lacUV5* promoter association. Proc Natl Acad Sci U S A **102**(2): 291-6.

28.   **Wosten, M. M.** 1998. Eubacterial sigma-factors. FEMS Microbiol Rev **22**(3): 127-50.

Appendix 1:

Supplemental data for Chapter 2 - Genetic dissection of the *Sau* σ^A_4 / gp67 interface

Supplemental Results

The side chain of *Sau* σ^A residue 313 may not contribute to gp67 binding specificity

We identified the region spanning *Sau* σ^A residues 309-335 (corresponding to *Eco* σ^{70} residues 553-579) as the region responsible for establishing the specificity of gp67 for *Sau* σ$^A_4$. We further identified the side chains of individual residues that are important to this binding specificity by introducing sequential substitutions in *Sau* σ$^A_4$ (and the corresponding substitutions in *Eco* σ$^{70}_4$) and looking for the minimum number of substitutions that abolished (or enabled, in the case of substitutions in *Eco* σ$^{70}_4$) the two-hybrid interaction between the α-σ$_4$ and λCI-gp67 fusion proteins. Our initial analysis, which started with a double substitution at *Sau* σ^A residues 312 and 313, identified the side chains of *Sau* σ^A residues 309, 312, 313, and 335 as contributors to the specificity of gp67 binding; the corresponding substitution series in *Eco* σ$^{70}_4$ corroborated these conclusions (see **Figures 2.6** and **2.7** for these data). The initial substitution series in *Eco* σ$^{70}_4$ raised the possibility that the Q579V substitution alone was the sole determinant of the switch in gp67 binding specificity, since the α-*Eco*(*Sau* 309, 312, 313) fusion protein did not bind gp67 but the α-*Eco*(*Sau* 309, 312, 313, 335) did (see **Figure 2.7A**).

To further investigate the role of *Sau* σ^A residue 335 (corresponding to *Eco* σ^{70} residue 579) in the specificity of gp67 binding, we constructed an additional substitution series using *Sau* σ$^A_4$ (and the corresponding complementary series using *Eco* σ$^{70}_4$) that started with the substitution at *Sau* σ^A V335 (or *Eco* σ^{70} Q579) and sequentially incorporated substitutions at residues 309, 312, and 313 (corresponding to *Eco* σ^{70} residues 553, 556, and 557, respectively). If the identity of the residue at *Sau* σ^A position 335 (*Eco* σ^{70} position 579) is the main determinant for gp67 binding specificity, we expected the α-*Sau*(*Eco* 579) fusion protein to be severely deficient in interacting with the λCI-gp67 fusion protein in the two-hybrid system, and, conversely, the α-*Eco*(*Sau* 335) fusion protein to bind gp67. As seen in **Figure A1.1**, the V335Q substitution in *Sau* σ$^A_4$ results in only a ~1.6-fold reduction in transcription activation in the two-
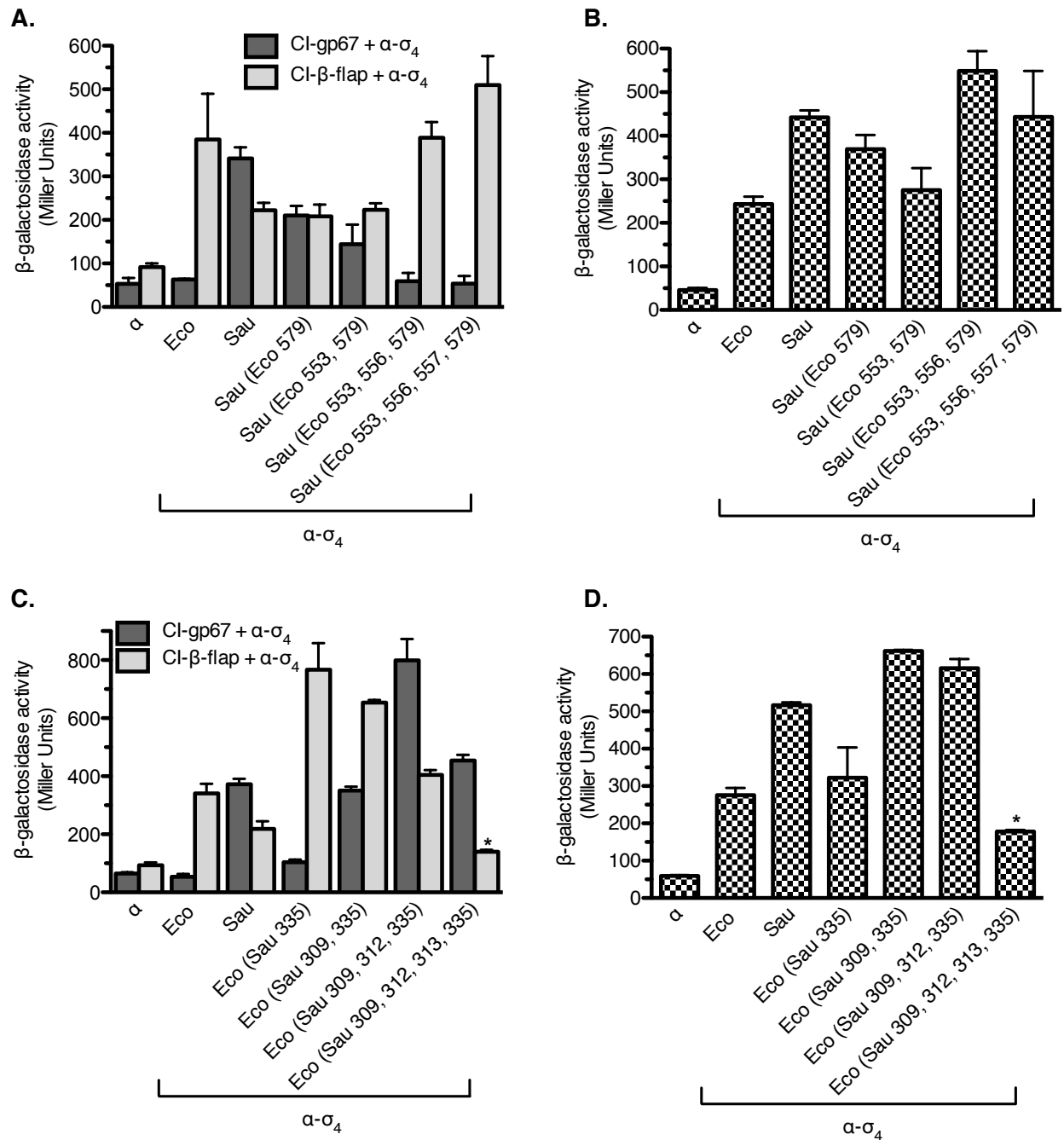
Figure A1.1. *Sau* σ^A residue N313 may not contribute to gp67 binding specificity. β-galactosidase activity is reported in Miller Units. Each bar represents the average Miller Units of three biological replicates assayed in the same experiment; error bars are ± standard deviation (.

(A) Two-hybrid interactions of a substitution series of α-*Sau* σ^A₄ fusion proteins, starting with a substitution at residue 335 **(Legend continued on next page.)**

Figure A1.1 (continued) and sequentially incorporating additional substitutions at the remaining divergent residues. The substitutions in each instance are to the residue found in *Eco* σ^{70}_4 at the corresponding position. Dark grey bars represent the interaction with the λCI-gp67 fusion protein and light grey bars the interaction with the λCI-*Sau* β-flap fusion protein to assess the structural integrity of the α-*Sau* σ^A_4 fusion proteins. The results indicate that *Sau* σ^A_4 bearing *Eco* σ^{70} residues at positions 309, 312, and 335 (corresponding to *Eco* σ^{70} 553, 556, and 579) is completely deficient for gp67 binding but is properly folded and functional, suggesting that the identity of the amino acid at these positions is important in establishing the specificity of gp67 for *Sau* σ^A and that residue 313 may not play an important role. **(B)** One-hybrid interactions of the substituted α-*Sau* σ^A_4 fusion proteins with an ectopically positioned -35 element, a second test for functionality. The results confirm that all of the α-*Sau* σ^A_4 fusion proteins are properly folded and functional. **(C)** Two-hybrid interactions of the complementary substitution series of α-*Eco* σ^{70}_4 fusion proteins. The substitutions in each instance are to the residue found in *Sau* σ^A_4 at the corresponding position, and all α-*Eco* σ^{70}_4 fusion proteins bear the D581G substitution. Dark and light grey bars are as in **(A)**. The results indicate that *Eco* σ^{70}_4 bearing *Sau* σ^A residues at positions 553, 556, and 579 (corresponding to *Sau* σ^A 309, 312, and 335) is fully competent to bind gp67 and is more stably folded than the previously described quadruply-substituted fusion protein (indicated by the asterisk (*) and described in **Figure 2.7**). *Eco* σ^{70}_4 bearing *Sau* σ^A residues at positions 553 and 556 alone is also competent for gp67 binding and properly folded. **(D)** One-hybrid interactions of the substituted α-*Eco* σ^{70}_4 fusion proteins with the ectopically-positioned -35 element. The data confirm that the α-*Eco* σ^{70}_4 fusion proteins are properly folded and functional. The functional quadruply-substituted fusion protein is indicated by (*) and described in **Figure 2.7** and the accompanying text in Chapter 2.

hybrid assay (see α-*Sau*(*Eco* 579) in **Figure A1.1A**, dark grey bars), and the Q579V

substitution alone in *Eco* σ^{70}_4 does no enable gp67 binding (see α-*Eco*(*Sau* 335) in **Figure**

A1.1C, dark grey bars). Together, these data indicated that the side chain of *Sau* σ^A residue 335

(*Eco* σ^{70} residue 579) is not the sole determinant for gp67 binding specificity, and that the side

chains of at least some of the other identified positions play an important role in this specificity.

This second substitution series in *Sau* σ^A_4 also enabled us to establish that the side

chain at *Sau* σ^A residue 312 (which in the original analysis was part of a double substitution with

residue 313 so we could not tease apart the effects of each) does contribute to gp67 binding

specificity. As seen in **Figure A1.1A**, substitutions at *Sau* σ^A residues 309 and 335 result in a

~2.4-fold reduction in transcription activation relative to wild-type *Sau* σ^A_4, and the addition of

the substitution at residue 312 abolishes gp67 binding in the two-hybrid system (compare α-

Sau(*Eco* 553, 579) and α-*Sau*(*Eco* 553, 556, 579), dark grey bars) without compromising the

structural integrity of the α-σ_4 fusion protein (**Figure A1.1A**, light grey bars, and **Figure A1.1B**).

Interestingly, the triply-substituted *Sau* σ^A_4 fusion protein (α-*Sau*(*Eco* 553, 556, 579)) exhibits

the same behavior with respect to gp67, *Sau* β-flap, and -35 element binding as the quadruply-

substituted *Sau* σ^A_4 originally used for subsequent functional analyses. These data suggested

that, although *Sau* σ^A N313 is surface-exposed and makes a hydrogen bond with residue N188

of gp67, the side chain at this position may not contribute significantly to gp67 binding specificity.

We used the complementary substitution series in *Eco* σ^{70}_4 to further investigate the role

of *Sau* σ^A residue N313 (corresponding to *Eco* σ^{70} residue K557) in determining the specificity of

gp67. We reasoned that if *Sau* σ^A N313 did not contribute significantly to this specificity, *Eco*

σ^{70}_4 bearing the corresponding substitutions at residues 553, 556, and 579 would be fully

competent for gp67 binding. As seen in **Figure A1.1C**, this is indeed the case: the α-*Eco*(*Sau*

309, 312, 335) fusion protein binds gp67 (dark grey bars) and, in fact, appears to be more

stable than the quadruply-substituted α-*Eco*(*Sau* 309, 312, 313, 335) fusion protein, since it

affords more robust interactions with the λCI-*Sau* β-flap fusion protein (light grey bars) and the -35 element (**Figure A1.D**). The apparent stronger interaction of the triply-substituted α-*Eco*(*Sau* 309, 312, 335) fusion protein with the λCI-gp67 fusion protein is likely the result of the increased stability of this α-σ_4 fusion protein rather than a defect in gp67 binding caused by the K557N substitution. This second mutational analysis was conducted after *Eco* σ^{70} bearing the corresponding *Sau* σ^A residues at positions 553, 556, 557, and 579 (and the D581G substitution present in all α-*Eco* σ^{70}_4 fusion proteins) had been purified and utilized in the *in vitro* transcription experiments presented in **Chapter 3**, but using *Eco* σ^{70} (*Sau* 309, 312, 335) for subsequent *in vivo* and *in vitro* functional analyses may have proven more fruitful due to the greater stability and stronger β-flap and -35 element binding. We also note that *Eco* σ^{70}_4 bearing the corresponding *Sau* σ^A residues at positions 553 and 579 alone binds gp67 as well as wild-type *Sau* σ^A_4 in the two-hybrid assay (see α-*Eco*(*Sau* 309, 335) in **Figure A1.1C**, dark grey bars) and is properly folded and functional (**Figure A1.1C**, light grey bars, and **Figure A1.1D**); however, due to the data from the *Sau* σ^A_4 substitution series that implicates the side chain of *Sau* σ^A_4 residue 312 as an important contributor to the specificity of gp67 binding, we believe that using *Eco* σ^{70}(*Sau* 309, 312, 335) would be preferable.

Supplemental Materials and Methods

Plasmids, Strains, and Growth Conditions

A complete list of the bacterial strains used **Chapter 2** and **Appendix 1** of the dissertation is provided in **Table A1.1**. A complete list of plasmids used in **Chapter 2** and **Appendix 1** of the dissertation is provided in **Table A1.2**. Plasmids and strains were constructed as outlined in the Materials and Methods section of Chapter 2.

Table A1.1. List of strains used in this study.

Strain	Relevant details	Reference
NEB5α F' I^q	*E. coli lacI^q* host strain for plasmid construction	New England Biolabs
FW102	*E. coli* host strain for promoter-*lacZ* fusions on single-copy F' episomes bearing either a kanamycin resistance gene (Kan) or a tetracycline resistance gene (Tet)	[5]
FW102 P$_{lac}$O$_L$2-62	FW102 harboring an F′ Kan bearing test promoter PlacO$_L$2-62 linked to *lacZ*	[1]
BN317	FW102 harboring an F′ Kan bearing test promoter P$_{lac}$Cons-35C, which bears a consensus -35 element (cTTGACA) upstream of the core *lac* promoter linked to *lacZ*	[3]

Table A1.2. List of plasmids used in this study.

Plasmid	Relevant details	Reference
pBRα	Encodes full-length α under the control of tandem *lpp* and *lacUV5* promoters; confers Carb[R]	[2]
pBRα-*Sau* σ$^A_4$	Encodes residues 1-248 of α fused by 3 alanine residues to residues 297-368 of *S. aureus* σ^A under the control of tandem *lpp* and *lacUV5* promoters; confers Carb[R]	[4]
pBRα-*Eco* σ$^{70}_4$	Encodes residues 1-248 of α fused by 3 alanine residues to residues 528-613 of *Eco* σ^{70} under the control of tandem *lpp* and *lacUV5* promoters; bears a mutation in the σ^{70} moiety specifying the D581G substitution; confers Carb[R]	[3]
pBRα-*Eco*(*Sau* 297-335)	Encodes residues 1-248 of α fused by 3 alanine residues to *Eco* σ$^{70}_4$ with *Sau* σ^A residues 297-335 in place of *Eco* σ^{70} residues 541-579 under the control of tandem *lpp* and *lacUV5* promoters; bears a mutation in the σ^{70} moiety specifying the D581G substitution; confers Carb[R]	This work
pBRα-*Eco*(*Sau* 336-368)	Encodes residues 1-248 of α fused by 3 alanine residues to *Eco* σ$^{70}_4$ with *Sau* σ^A residues 336-368 in place of *Eco* σ^{70} residues 580-613 under the control of tandem *lpp* and *lacUV5* promoters; confers Carb[R]	This work
pBRα-*Sau*(*Eco* 561-613)	Encodes residues 1-248 of α fused by 3 alanine residues to *Sau* σ$^A_4$ with *Eco* σ^{70} residues 561-613 in place of *Sau* σ^A residues 317-368 under the control of tandem *lpp* and *lacUV5* promoters; bears a mutation in the σ^{70} moiety specifying the D581G substitution confers Carb[R]	This work
pBRα-*Eco*(*Sau* 317-368)	Encodes residues 1-248 of α fused by 3 alanine residues to *Eco* σ$^{70}_4$ with *Sau* σ^A residues 317-368 in place of *Eco* σ^{70} residues 561-613 under the control of tandem *lpp* and *lacUV5* promoters; confers Carb[R]	This work
pBRα-*Eco*(*Sau* 305-368)	Encodes residues 1-248 of α fused by 3 alanine residues to *Eco* σ$^{70}_4$ with *Sau* σ^A residues 305-368 in place of *Eco* σ^{70} residues 549-613 under the control of tandem *lpp* and *lacUV5* promoters; confers Carb[R]	This work
pBRα-*Sau*(*Eco* 549-613)	Encodes residues 1-248 of α fused by 3 alanine residues to *Sau* σ$^A_4$ with *Eco* σ^{70} residues 549-613 in place of *Sau* σ^A residues 305-368 under the control of tandem *lpp* and *lacUV5* promoters; bears a mutation in the σ^{70} moiety specifying the D581G substitution confers Carb[R]	This work

Plasmid	Relevant details	Reference
pBRα-*Eco*(*Sau* 309-368)	Encodes residues 1-248 of α fused by 3 alanine residues to *Eco* σ^{70}_4 with *Sau* σ^A residues 309-368 in place of *Eco* σ^{70} residues 553-613 under the control of tandem *lpp* and *lacUV5* promoters; confers Carb^R	This work
pBRα-*Sau*(*Eco* 553-613)	Encodes residues 1-248 of α fused by 3 alanine residues to *Sau* σ^A_4 with *Eco* σ^{70} residues 553-613 in place of *Sau* σ^A residues 309-368 under the control of tandem *lpp* and *lacUV5* promoters; bears a mutation in the σ^{70} moiety specifying the D581G substitution; confers Carb^R	This work
pBRα-*Eco*(*Sau* 309-335)	Encodes residues 1-248 of α fused by 3 alanine residues to *Eco* σ^{70}_4 with *Sau* σ^A residues 309-335 in place of *Eco* σ^{70} residues 553-579 under the control of tandem *lpp* and *lacUV5* promoters; bears a mutation in the σ^{70} moiety specifying the D581G substitution; confers Carb^R	This work
pBRα-*Sau*(*Eco* 553-579)	Encodes residues 1-248 of α fused by 3 alanine residues to *Sau* σ^A_4 with *Eco* σ^{70} residues 553-579 in place of *Sau* σ^A residues 309-335 under the control of tandem *lpp* and *lacUV5* promoters; confers Carb^R	This work
pBRα-*Sau*(*Eco* 556, 557)	Encodes residues 1-248 of α fused by 3 alanine residues to *Sau* σ^A_4 bearing the corresponding *Eco* σ^{70} residues at positions 312 and 313 under the control of tandem *lpp* and *lacUV5* promoters; confers Carb^R	This work
pBRα-*Sau*(*Eco* 553, 556, 557)	Encodes residues 1-248 of α fused by 3 alanine residues to *Sau* σ^A_4 bearing the corresponding *Eco* σ^{70} residues at positions 309, 312, and 313 under the control of tandem *lpp* and *lacUV5* promoters; confers Carb^R	This work
pBRα-*Sau*(*Eco* 553, 556, 557, 579)	Encodes residues 1-248 of α fused by 3 alanine residues to *Sau* σ^A_4 bearing the corresponding *Eco* σ^{70} residues at positions 309, 312, 313, and 335 under the control of tandem *lpp* and *lacUV5* promoters; confers Carb^R	This work
pBRα-*Sau*(*Eco* 579)	Encodes residues 1-248 of α fused by 3 alanine residues to *Sau* σ^A_4 bearing the corresponding *Eco* σ^{70} residue at position 335 under the control of tandem *lpp* and *lacUV5* promoters; confers Carb^R	This work
pBRα-*Sau*(*Eco* 553, 579)	Encodes residues 1-248 of α fused by 3 alanine residues to *Sau* σ^A_4 bearing the corresponding *Eco* σ^{70} residues at positions 309 and 335 under the control of tandem *lpp* and *lacUV5* promoters; confers Carb^R	This work

Plasmid	Relevant details	Reference
pBRα-*Sau*(*Eco* 553, 556, 579)	Encodes residues 1-248 of α fused by 3 alanine residues to *Sau* σ^A_4 bearing the corresponding *Eco* σ^{70} residues at positions 309, 312, and 335 under the control of tandem *lpp* and *lacUV5* promoters; confers Carb^R	This work
pBRα-*Eco*(*Sau* 309)	Encodes residues 1-248 of α fused by 3 alanine residues to *Eco* σ^{70}_4 bearing the corresponding *Sau* σ^A residue at position 553 under the control of tandem *lpp* and *lacUV5* promoters; bears a mutation in the σ^{70} moiety specifying the D581G substitution; confers Carb^R	This work
pBRα-*Eco*(*Sau* 309, 312)	Encodes residues 1-248 of α fused by 3 alanine residues to *Eco* σ^{70}_4 bearing the corresponding *Sau* σ^A residues at positions 553 and 556 under the control of tandem *lpp* and *lacUV5* promoters; bears a mutation in the σ^{70} moiety specifying the D581G substitution; confers Carb^R	This work
pBRα-*Eco*(*Sau* 309, 312, 313)	Encodes residues 1-248 of α fused by 3 alanine residues to *Eco* σ^{70}_4 bearing the corresponding *Sau* σ^A residues at positions 553, 556, and 557 under the control of tandem *lpp* and *lacUV5* promoters; bears a mutation in the σ^{70} moiety specifying the D581G substitution; confers Carb^R	This work
pBRα-*Eco*(*Sau* 309, 312, 313, 335)	Encodes residues 1-248 of α fused by 3 alanine residues to *Eco* σ^{70}_4 bearing the corresponding *Sau* σ^A residues at positions 553, 556, 557, and 579 under the control of tandem *lpp* and *lacUV5* promoters; bears a mutation in the σ^{70} moiety specifying the D581G substitution; confers Carb^R	This work
pBRα-*Eco*(*Sau* 335)	Encodes residues 1-248 of α fused by 3 alanine residues to *Eco* σ^{70}_4 bearing the corresponding *Sau* σ^A residue at position 579 under the control of tandem *lpp* and *lacUV5* promoters; bears a mutation in the σ^{70} moiety specifying the D581G substitution; confers Carb^R	This work
pBRα-*Eco*(*Sau* 309, 335)	Encodes residues 1-248 of α fused by 3 alanine residues to *Eco* σ^{70}_4 bearing the corresponding *Sau* σ^A residues at positions 553 and 579 under the control of tandem *lpp* and *lacUV5* promoters; bears a mutation in the σ^{70} moiety specifying the D581G substitution; confers Carb^R	This work
pBRα-*Eco*(*Sau* 309, 312 335)	Encodes residues 1-248 of α fused by 3 alanine residues to *Eco* σ^{70}_4 bearing the corresponding *Sau* σ^A residues at positions 553, 556, and 579 under the control of tandem *lpp* and *lacUV5* promoters; bears a mutation in the σ^{70} moiety specifying the D581G substitution; confers Carb^R	This work

Table A1.2 (continued)

Plasmid	Relevant details	Reference
pBRα-*Bsu* σ$^A_4$	Encodes residues 1-248 of α fused by 3 alanine residues to residues 300-371 of *B. subtilis* σ^A under the control of tandem *lpp* and *lacUV5* promoters; confers Carb^R	This work
pBRα-*Bsu*(*Eco* 553, 556, 557, 579)	Encodes residues 1-248 of α fused by 3 alanine residues to *Bsu* σ$^A_4$ bearing the corresponding *Eco* σ^{70} residues at positions 312, 315, 316, and 338 under the control of tandem *lpp* and *lacUV5* promoters; confers Carb^R	This work
pBRα-*Taq* σ$^A_4$	Encodes residues 1-248 of α fused by 3 alanine residues to residues 353-438 of *T. aquaticus* σ^A under the control of tandem *lpp* and *lacUV5* promoters; confers Carb^R	This work
pBRα-*Taq*(*Sau* 309, 312, 313, 335)	Encodes residues 1-248 of α fused by 3 alanine residues to *Taq* σ$^A_4$ bearing the corresponding *Sau* σ^A residues at positions 378, 381, 382, and 404 under the control of tandem *lpp* and *lacUV5* promoters; confers Carb^R	This work
pBRα-*Sau* σ$^A_4$ E312K	Encodes residues 1-248 of α fused by 3 alanine residues to *Sau* σ$^A_4$ bearing the E312K substitution under the control of tandem *lpp* and *lacUV5* promoters; confers Carb^R	This work
pBRα-*Eco* σ$^{70}_4$ R603A	Encodes residues 1-248 of α fused by 3 alanine residues to *Eco* σ$^{70}_4$ bearing R603A and D581G substitutions under the control of tandem *lpp* and *lacUV5* promoters; confers Carb^R	This work
pBRα-*Sau* σ$^A_4$ R359A	Encodes residues 1-248 of α fused by 3 alanine residues to *Sau* σ$^A_4$ bearing the R359A substitution under the control of tandem *lpp* and *lacUV5* promoters; confers Carb^R	This work
pBRα-*Sau* σ$^A_4$ R310A	Encodes residues 1-248 of α fused by 3 alanine residues to *Sau* σ$^A_4$ bearing the R310A substitution under the control of tandem *lpp* and *lacUV5* promoters; confers Carb^R	This work
pBRα-*Sau* σ$^A_4$ R310D	Encodes residues 1-248 of α fused by 3 alanine residues to *Sau* σ$^A_4$ bearing the R310D substitution under the control of tandem *lpp* and *lacUV5* promoters; confers Carb^R	This work
pBRα-*Eco* σ$^{70}_4$ R554A	Encodes residues 1-248 of α fused by 3 alanine residues to *Eco* σ$^{70}_4$ bearing R554A and D581G substitutions under the control of tandem *lpp* and *lacUV5* promoters; confers Carb^R	This work
pBRα-*Eco* σ$^{70}_4$ R554D	Encodes residues 1-248 of α fused by 3 alanine residues to *Eco* σ$^{70}_4$ bearing R554D and D581G substitutions under the control of tandem *lpp* and *lacUV5* promoters; confers Carb^R	This work
pBRα-gp67	Encodes residues 1-248 of α fused by 3 alanine residues to gp67 under the control of tandem *lpp* and *lacUV5* promoters; confers Carb^R	This work

Plasmid	Relevant details	Reference
pACλCI	Encodes full-length λCI under the control of the *lacUV5* promoter; confers Cam^R	[2]
pACλCI-gp67	Encodes residues 1-236 of λCI fused via 3 alanine residues to gp67 under the control of the *lacUV5* promoter; confers Cam^R	[4]
pACλCI-*Sau* β-flap	Encodes residues 1-236 of λCI fused via 3 alanine residues to residues 789-917 of the β subunit of *Sau* RNAP under the control of the *lacUV5* promoter; confers Cam^R	[4]
pACλCI-gp67 K2D K195D	Encodes residues 1-236 of λCI fused via 3 alanine residues to gp67 bearing K2D and K195D substitutions under the control of the *lacUV5* promoter; confers Cam^R	This work
pACλCI-gp67 K2E K195E	Encodes residues 1-236 of λCI fused via 3 alanine residues to gp67 bearing K2E and K195E substitutions under the control of the *lacUV5* promoter; confers Cam^R	This work
pACλCI-gp67 Y37F N41R	Encodes residues 1-236 of λCI fused via 3 alanine residues to gp67 bearing Y37F and N41R substitutions under the control of the *lacUV5* promoter; confers Cam^R	This work
pACλCI-gp67 Y37R N41A	Encodes residues 1-236 of λCI fused via 3 alanine residues to gp67 bearing Y37R and N41A substitutions under the control of the *lacUV5* promoter; confers Cam^R	This work
pACλCI-gp67 Y37R N41R	Encodes residues 1-236 of λCI fused via 3 alanine residues to gp67 bearing Y37R and N41R substitutions under the control of the *lacUV5* promoter; confers Cam^R	This work
pACλCI-*Sau* β-flap ΔFTH	Encodes residues 1-236 of λCI fused via 3 alanine residues to *Sau* β-flap with residues 858-867 deleted under the control of the *lacUV5* promoter; confers Cam^R	This work
pCDF*lac*	High-copy plasmid with multiple cloning site; encodes no functional protein; confers Spec^R	Modified from pCDF-1b (Novagen)
pCDF*lac*-His$_6$-*Sau* σ$^A_4$	High-copy plasmid; encodes *Sau* σ$^A_4$ bearing an N-terminal His$_6$ tag under the control of the *lacUV5* promoter; confers Spec^R	This work
pCDF*lac*-gp67-His$_6$	High-copy plasmid; encodes gp67 bearing a C-terminal His$_6$ tag under the control of the *lacUV5* promoter; confers Spec^R	This work

References

1. **Deaconescu, A. M.*, et al.* 2006. Structural basis for bacterial transcription-coupled DNA repair. Cell **124**(3): 507-20.

2. **Dove, S. L., Joung, J. K., and Hochschild, A.** 1997. Activation of prokaryotic transcription through arbitrary protein-protein contacts. Nature **386**: 627-30.

3. **Nickels, B. E.*, et al.* 2002. Protein-protein and protein-DNA interactions of sigma70 region 4 involved in transcription activation by lambda cI. J Mol Biol **324**: 17-34.

4. **Osmundson, J.*, et al.* 2012. Promoter-specific transcription inhibition in *Staphylococcus aureus* by a phage protein. Cell **151**(5): 1005-16.

5. **Whipple, F. W.** 1998. Genetic analysis of prokaryotic and eukaryotic DNA-binding proteins in *Escherichia coli*. Nucleic Acids Res **26**(16): 3700-6.

Appendix 2:

Promoter-Specific Transcription Inhibition in *Staphylococcus aureus* by a Phage Protein

Promoter-Specific Transcription Inhibition in *Staphylococcus aureus* by a Phage Protein

Joseph Osmundson,[1] Cristina Montero-Diez,[2] Lars F. Westblade,[1,3] Ann Hochschild,[2] and Seth A. Darst[1,*]

[1]The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA
[2]Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA 02115, USA
[3]Present address: Department of Pathology and Laboratory Medicine, Hofstra North Shore-LIJ School of Medicine, Hempstead, NY 11549, USA
*Correspondence: darst@rockefeller.edu
http://dx.doi.org/10.1016/j.cell.2012.10.034

SUMMARY

Phage G1 gp67 is a 23 kDa protein that binds to the *Staphylococcus aureus* (*Sau*) RNA polymerase (RNAP) σ^A subunit and blocks cell growth by inhibiting transcription. We show that gp67 has little to no effect on transcription from most promoters but is a potent inhibitor of ribosomal RNA transcription. A 2.0-Å-resolution crystal structure of the complex between gp67 and *Sau* σ^A domain 4 (σ^A_4) explains how gp67 joins the RNAP promoter complex through σ^A_4 without significantly affecting σ^A_4 function. Our results indicate that gp67 forms a complex with RNAP at most, if not all, σ^A-dependent promoters, but selectively inhibits promoters that depend on an interaction between upstream DNA and the RNAP α-subunit C-terminal domain (αCTD). Thus, we reveal a promoter-specific transcription inhibition mechanism by which gp67 interacts with the RNAP promoter complex through one subunit (σ^A), and selectively affects the function of another subunit (αCTD) depending on promoter usage.

INTRODUCTION

Staphylococcus aureus (*Sau*) is a pathogenic Gram-positive bacterium that causes infections of the skin as well as pneumonia, meningitis, endocarditis, and sepsis (Lowy, 1998). Methicillin-resistant *Sau* (MRSA) infections have become increasingly common and lead to significant morbidity and mortality (Nordmann et al., 2007). MRSA strains were recently reported to also be resistant to vancomycin, the current antibiotic of last resort (Howden et al., 2010), highlighting the importance of discovering novel therapeutics that inhibit *Sau* growth.

Before the advent of modern antibiotics, bacteriophages (phages) were considered potential therapeutic agents because of their ability to kill specific bacterial species rapidly. As microbes become increasingly resistant to antibiotics, phages are once again being exploited for their ability to directly elimi-

nate bacterial infections and provide novel targets for the drug discovery process (Fischetti, 2008).

Recent studies have mined phage genomes for proteins and peptides that inhibit *Sau* cell growth (Kwan et al., 2005; Liu et al., 2004). One such protein, phage G1 gp67 (encoded by *orf67*), inhibits cell growth when produced in *Sau*, but not in *Escherichia coli* (*Eco*), and was shown to significantly decrease RNA production in *Sau* (Liu et al., 2004). Subsequent work demonstrated a direct interaction between gp67 and *Sau* RNA polymerase (RNAP; Dehbi et al., 2009).

All transcription in prokaryotes is performed by the ∼400 kDa core RNAP (subunit composition $\alpha_2\beta\beta'\omega$; Darst, 2001). Promoter recognition and initiation require an additional subunit, σ, which binds to the core RNAP to form the holoenzyme (Murakami and Darst, 2003). The group 1, or primary, σ factors (σ^{70} in *Eco*, σ^A in *Sau*) are responsible for the bulk of transcription during log-phase growth and are essential for viability (Gruber and Gross, 2003). In the context of the RNAP holoenzyme, the structural domains of σ^A directly recognize conserved core promoter elements: domain 2 of σ^A (σ^A_2) recognizes the −10 element, and σ^A_4 recognizes the −35 element (Campbell et al., 2002; Feklistov and Darst, 2011; Murakami et al., 2002a; Shultzaberger et al., 2007). Promoters may contain additional elements recognized by σ, such as the extended −10 (Keilty and Rosenberg, 1987) or discriminator element (Feklistov et al., 2006; Haugen et al., 2008b; Travers, 1984). In addition, certain promoters are dependent on an A/T-rich DNA sequence upstream of the −35 element, known as the UP element (Ross et al., 1993). The UP element is recognized by the 9 kDa C-terminal domains of the RNAP α subunits (αCTDs), which are attached to the α-N-terminal domains by ∼15-residue-long, flexible unstructured linkers (Blatter et al., 1994; Gourse et al., 2000).

Previous biochemical analyses established that gp67I (1) interacts directly with *Sau* σ^A but not *Eco* σ^{70}, (2) uses σ^A_4 as its primary interaction determinant, and (3) forms a stable, ternary gp67/*Sau*-σ^A/*Sau*-core-RNAP complex (Dehbi et al., 2009). Using the λP_L promoter with a hybrid system comprising *Sau* σ^A with *Eco* core RNAP, as well as *Sau*-σ^A/*Sau*-core-RNAP, Dehbi et al. (2009) concluded that gp67 is a *Sau* σ^A-specific anti-σ factor that blocks σ^A_4 recognition of the −35 element

194
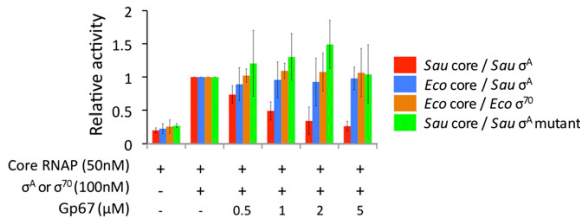
Figure 1. Gp67 Inhibits Transcription by the *Sau* RNAP Holoenzyme but Not the *Eco* RNAP Holoenzyme or the Hybrid Holoenzyme *Sau*-σ^A/*Eco*-core RNAP

Values represent the mean of three independent experiments, and error bars denote the mean ± 1 SD.

See also Figure S1.

and is therefore a general inhibitor of −35 element-dependent promoters.

Here, we demonstrate that gp67 is a promoter-specific transcription inhibitor. Gp67 does not inhibit transcription from the majority of *Sau* promoters, but is a potent inhibitor of ribosomal RNA (rRNA) transcription in vitro and in vivo. Using X-ray crystal structures of gp67 bound to *Sau* σ$^A_4$, in vivo RNA sequencing (RNA-seq) analysis, and in vitro biochemical analyses, we show that gp67 interferes with binding of the αCTDs to DNA upstream of the promoter −35 element, and therefore inhibits promoters that depend on the αCTD/DNA interaction. Gp67 is a phage-encoded transcription factor that acts through a novel mechanism in which it joins the RNAP holoenzyme through σ$^A_4$ but modulates the host transcription program by interfering with promoter DNA interactions of the αCTDs.

RESULTS

Gp67 Function Requires a Native *Sau* Transcription System

In contrast to *Sau* RNAP, *Eco* RNAP is well characterized biochemically, and numerous tools such as mutant and modified *Eco* RNAPs are available for probing mechanistic questions. Therefore, as a first step toward understanding the molecular mechanism through which gp67 inhibits RNAP, we sought to reproduce the results of Dehbi et al. (2009) using the hybrid transcription system of *Sau* σ^A with *Eco* core RNAP on the λP$_L$ promoter. In contrast to Dehbi et al. (2009), we did not observe any effect of gp67 on λP$_L$ transcription, even when gp67 was at a 100-fold molar excess over the hybrid holoenzyme (Figure S1A available online). Moreover, a native gel-shift analysis revealed that gp67 formed a ternary gp67/*Sau*-σ^A/*Sau*-core-RNAP complex (Figure S1B, lane 3) but did not interact with the hybrid *Sau*-σ^A/*Eco*-core-RNAP (Figure S1C, lane 1).

Because gp67 failed to inhibit or even interact with the hybrid holoenzyme, we purified *Sau* core RNAP (Deora and Misra, 1996) and optimized the in vitro transcription conditions (see Experimental Procedures). Few *Sau* promoters have been characterized in vitro, so we first investigated *Sau* RNAP-holoenzyme activity in vitro using *Sau* genomic DNA as a template. Gp67 inhibited promoter-specific *Sau* RNAP-holoenzyme transcription from *Sau* genomic DNA (Figure 1, red bars), but did not inhibit

transcription by the hybrid holoenzyme *Sau*-σ^A/*Eco*-RNAP (Figure 1, blue bars) or by *Eco* RNAP holoenzyme (Figure 1, orange bars), consistent with our in vitro results on the λP$_L$ promoter (Figure S1A). To determine whether inhibition of *Sau* RNAP holoenzyme by gp67 was dependent on the *Sau*-σ^A/gp67 interaction, we used a bacterial two-hybrid assay (Dove and Hochschild, 2004) to generate an otherwise functional σ^A mutant that could no longer interact with gp67 (Figure S1D). We found that gp67 was unable to inhibit transcription by the corresponding mutant *Sau* holoenzyme (Figure 1, green bars). Gp67 inhibition therefore requires a native *Sau* transcription system.
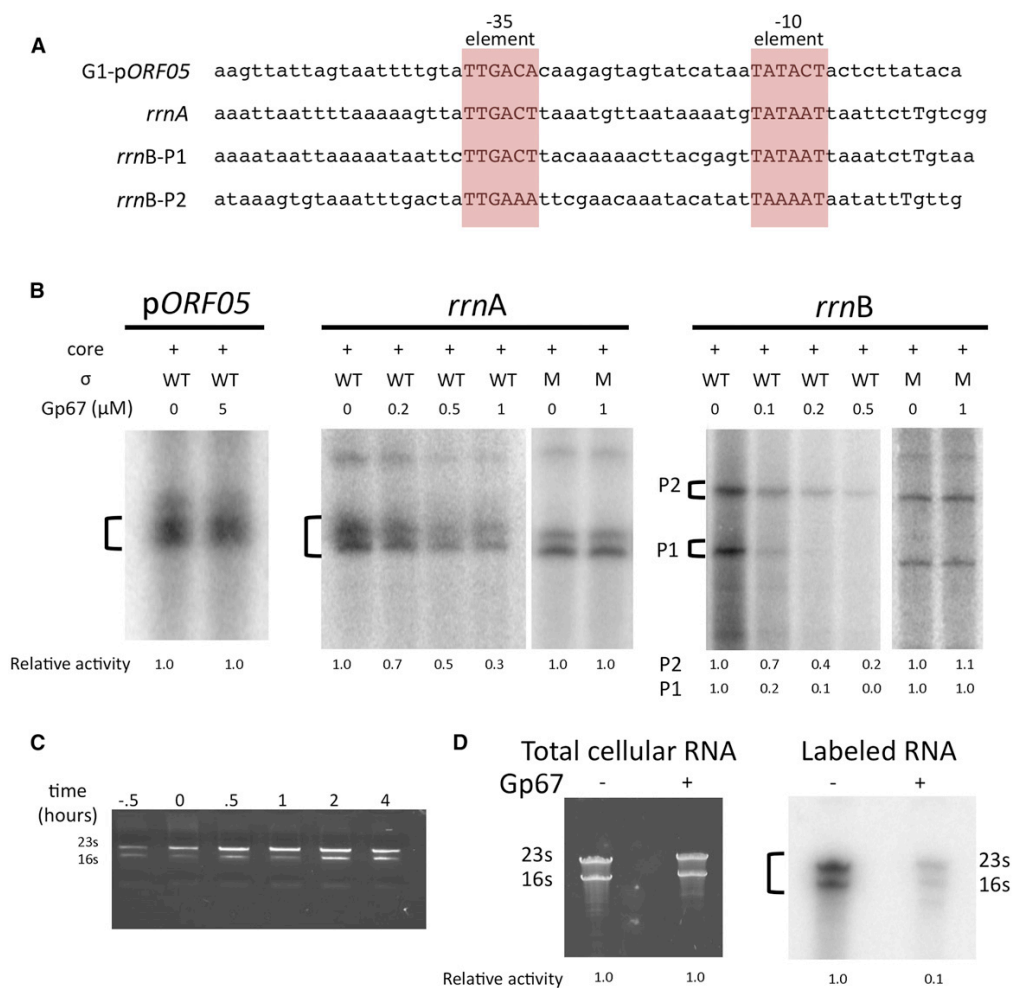
Gp67 Inhibits rRNA Synthesis In Vitro and In Vivo

Phage promoters have been important tools for studying RNAP function (Kadesch et al., 1982; Rosenberg et al., 1982; Stevens, 1977). The genome of the *Sau*-specific G1 phage has many easily identifiable promoters that closely match the −10/−35 consensuses with the optimal 17 bp spacing between them (Shultzaberger et al., 2007). One such promoter is G1-pORF05 (Figure 2A). *Sau* RNAP-holoenzyme has robust activity from this promoter when tested in vitro, but gp67 has no effect on this activity even at very high concentrations (Figure 2B). Because gp67 does not inhibit this −10/−35 promoter, it is unlikely that gp67 blocks recognition of the −35 element as previously proposed (Dehbi et al., 2009).

Gp67 inhibits bulk RNA synthesis in vivo (Liu et al., 2004) and in vitro (Figure 1). Because the major fraction of transcription in growing bacterial cells is dedicated to ribosome synthesis (Gourse et al., 1996), we examined the effect of gp67 on transcription from three *Sau* rRNA promoters: *rrnA*, *rrnB*-P1, and *rrnB*-P2 (Figure 2A). Gp67 is a potent inhibitor of in vitro transcription from these *Sau* rRNA promoters, and the inhibition depends on the interaction between gp67 and RNAP (Figure 2B).

To test whether gp67 inhibits rRNA synthesis in vivo, we expressed gp67 in *Sau* cells using an inducible expression vector (Corrigan and Foster, 2009). Upon addition of inducer, cell growth was slowed but not halted by gp67 expression (Figure S2) and rRNA levels remained unchanged (Figure 2C). Whereas messenger RNAs (mRNAs) generally have short half-lives in vivo, rRNAs are stable in prokaryotic cells (Deutscher, 2003). Therefore, to assess the effect of gp67 expression on rRNA synthesis, we used metabolic labeling to observe newly transcribed RNAs. Gp67 was expressed in *Sau* cells, and then newly synthesized RNAs were labeled by addition of radiolabeled phosphate to the growth medium (Wade et al., 1964). Whereas gp67 had little to no effect on the overall rRNA levels (Figures 2C and 2D, left panel), gp67 production significantly reduced new rRNA synthesis (Figure 2D, right panel). These data indicate that gp67 inhibits RNAP activity at rRNA promoters, but the rRNAs remain stable in the cells for hours after gp67 production. We conclude that gp67 is not a general inhibitor of *Sau* σ^A holoenzyme, but is a potent and direct inhibitor of rRNA transcription in vitro (Figure 2B) and in vivo (Figure 2D).

RNA-seq Reveals Promoters Inhibited by gp67 on a Genome-wide Scale

Because gp67 appears to be a promoter-specific inhibitor of *Sau* transcription, we wanted to determine the effect of gp67

195

Figure 2. Gp67 Inhibits rRNA Synthesis In Vitro and In Vivo

(A) Sequences of promoters used in subsequent experiments.

(B) Gp67 inhibits RNAP activity at *Sau rrn* promoters but not at a G1 phage −10/−35 promoter (pORF05). The indicated bands were quantified by phosphorimagery; relative activity (versus holoenzyme alone) is denoted below each lane.

(C) Gp67 does not significantly decrease rRNA abundance in vivo. Cellular RNA was purified from 2×10^8 *Sau* RN4220 cells at different times after the induction of gp67 synthesis. The *Sau* σ is either wild-type (WT) or the mutant (M) unable to interact with gp67 (Figure S1D).

(D) Gp67 inhibits synthesis of rRNA in vivo. Inducer was added to *Sau* RN4220 cells containing pRMC2 or pRMC2-gp67, and newly synthesized RNA was labeled by the addition of [^{32}P]-orthophosphoric acid to the growth medium. RNA was purified from 2×10^8 pRMC2 and pRMC2-gp67 cells, run on a 6% urea-PAGE gel, stained with GelRed to visualize all RNAs (left), and visualized by phosphorimagery (right).

See also Figure S2.

on a wide variety of promoters in order to understand the characteristics that lead to gp67 sensitivity. For this purpose, we used RNA-seq to quantitatively compare RNA transcripts from *Sau* cells grown in the absence or presence of gp67. RNA was purified from *Sau* RN4220 cells producing gp67 as well as from control cells (Figure S2C), rRNAs were depleted, and libraries were prepared by standard procedures. The RNA was then sequenced by Illumina technology. We found that gp67 had no effect on transcription

from the majority of promoters in the *Sau* genome (Figure 3A). Less than 4% of all transcripts were downregulated, and ~5% were upregulated by gp67 production during the growth phase.

RNA-seq only reveals RNA transcript levels; it does not differentiate between direct and indirect regulation of gene expression, nor does it reveal whether differential gene expression is due to changes in promoter binding and initiation or mRNA stability. We used the *Sau* in vitro transcription system to test

196
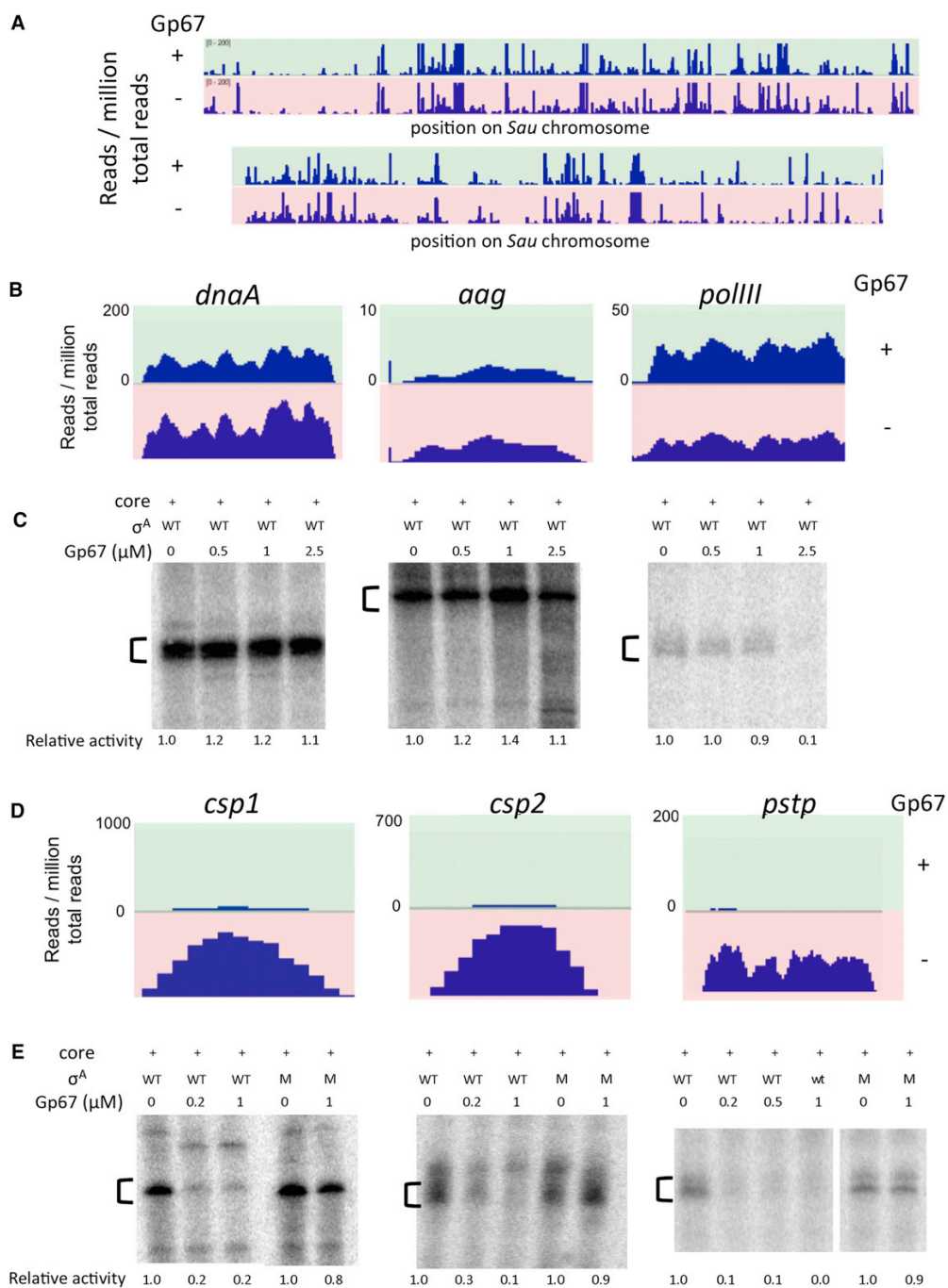
Figure 3. RNA-seq Reveals Promoters that Are Sensitive to gp67 Inhibition In Vivo

(A) Gp67 production in *Sau* RN4220 cells does not significantly inhibit RNA levels from ~91% of promoters. RNAs were sequenced directly and visualized with the Integrated Genomics Viewer. The position on the genome is shown on the horizontal axis, and the number of RNA reads per million total reads is shown on the vertical axis. Upper panel (green) represents RNA-seq data from pRMC2-gp67 cells; the lower panel (red) represents RNA-seq data from control pRMC2 cells.

(B) Gp67 production does not inhibit transcription from selected DNA replication promoters in vivo. RNA-seq data were visualized as above from the *dnaA*, *aag*, and polIII loci.

197
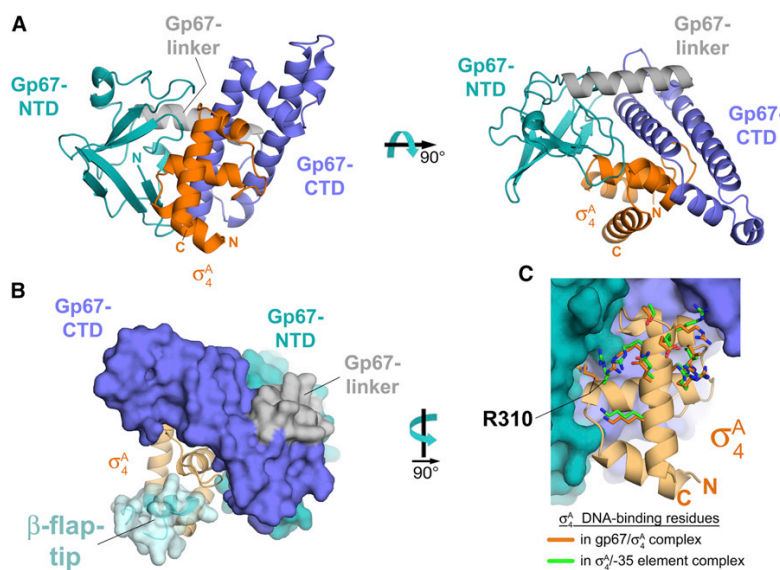
(A) Cocrystal structure of gp67 bound to σ$^A_4$ (orange). Two orthogonal views are shown. Proteins are shown in ribbon format.

(B) Structural modeling indicates that the RNAP β-flap-tip/σ$^A_4$ interaction is compatible with the presence of gp67. Gp67 is shown as a molecular surface, color-coded as in (A). The σ$^A_4$ is shown as a light orange ribbon. The RNAP β-flap-tip, modeled by superimposing σ$^A_4$ from the *Thermus aquaticus* RNAP holoenzyme (Murakami et al., 2002b), is shown as a light blue ribbon with a transparent molecular surface (cyan).

(C) Gp67 does not contact most residues of σ$^A_4$ that interact with −35-element DNA. Shown is a close-up view of *Sau* σ$^A_4$ (light orange ribbon) in the complex with gp67 (molecular surface, color-coded as in [A]). Side chains of residues that directly interact with −35-element DNA are shown from the gp67/*Sau* σ$^A_4$ structure (orange) and from the superimposed *Taq* σ$^A_4$/-35 element DNA complex (green side chains; PDB ID code 1KU7; Campbell et al., 2002). Only the side chain of *Sau* σ$^A_4$ R310 (corresponding to *Taq* σ$^A_4$ R379/*Eco* σ^{70} R554) is altered through an interaction with gp67. See also Figure S4 and Tables S1 and S2.

whether gp67 directly affects transcription at promoters identified by RNA-seq.

DNA replication is often a target of early phage proteins (Datta et al., 2005; Yano and Rothman-Denes, 2011). However, expression of genes involved in replication (Xu et al., 2010) was not significantly altered in the presence of gp67 in vivo (Figure 3B). Gp67 also did not inhibit the activity of RNAP at the *dnaA* or *aag* (a putative DNA repair protein) promoters in vitro, and had an effect on already weak RNAP activity at the *pol*III promoter only at a very high gp67 concentration (Figure 3C). In contrast, when tested on promoters that show significant levels of inhibition by gp67 in vivo (Figure 3D), gp67 was found to be a potent inhibitor of transcription in vitro (Figure 3E), indicating that, at least at these promoters, the in vivo effects are direct. Gp67 had no effect on transcript levels from *Sau* promoters that were previously studied using purified in vitro transcription systems (Figure S3; Rao et al., 1995; Reyes et al., 2011). Because gp67 had no effect on >90% of *Sau* promoters in vivo, or on the G1-pORF05, *dnaA*, and *aag* promoters in vitro, gp67 must not target general σ functions, such as −10 or −35 promoter element recognition or core RNAP interaction.

Crystal Structure of the gp67/*Sau* σ$^A_4$ Complex

To shed light on the mechanism by which gp67 modulates *Sau* RNAP function, we determined X-ray crystal structures of the gp67/*Sau*-σ$^A_4$ complex. We solved the structure from two crystal forms: form I (2.0 Å resolution) and form II (3.0 Å resolution; Tables S1 and S2; Figure S4C). Three crystallographically independent structures did not show any significant differences (root-mean-square deviation [rmsd] of α-carbon positions = ≤ 0.78 Å). Therefore, we derived all of the structural analyses using the high-resolution form I structure. Although the structure of σ$_4$ has been well described (Campbell et al., 2002), gp67 has no sequence or structural homology to any previously described fold.

The 2.0 Å crystal structure reveals the nature of the interactions between gp67 and σ$^A_4$ (Figures 4A, S4D, and S4E). The σ$^A_4$ has been studied structurally in many contexts, including bound to DNA (Campbell et al., 2002), RNAP (Murakami et al., 2002b; Vassylyev et al., 2002), RNAP and DNA (Murakami et al., 2002b), and anti-σ factors (Campbell et al., 2008). Gp67 does not reorganize the conformation of σ$^A_4$ (Figures 4A and S4F), as occurs in some anti-σ/σ complexes (Campbell et al., 2007; Lambert et al., 2004).

Gp67 itself comprises two domains: an N-terminal β-sheet-rich domain (gp67-NTD; Figure 4A, teal) and a C-terminal α-helical domain (gp67-CTD; Figure 4A, blue). The two domains are connected by a linker-α-helix on the surface of gp67 opposite the σ$^A_4$ interaction surface (gp67-linker; Figure 4A, gray). Gp67 forms an extensive molecular interface with σ$^A_4$ through both the gp67-NTD (1,032 Å^2 buried surface area) and the gp67-CTD (1,757 Å^2 buried surface area), with a total buried

---

(C) Gp67 does not inhibit transcription from selected DNA replication promoters in vitro. The indicated bands were quantified by phosphorimagery; relative activity (versus holoenzyme alone) is denoted below each lane.

(D) Selected genes inhibited by gp67 in vivo. RNA-seq data were visualized as above from the *csp1*, *csp2*, and *pstp* loci.

(E) Gp67 inhibits transcription in vitro from the selected promoters that are susceptible in vivo when wild-type (WT) σ^A is used, but not the quadruple mutant σ^A that does not interact with gp67 (M). The indicated bands were quantified by phosphorimagery; relative activity (versus holoenzyme alone) is denoted below each lane.

See also Figures S2 and S3, and Table S3.

198
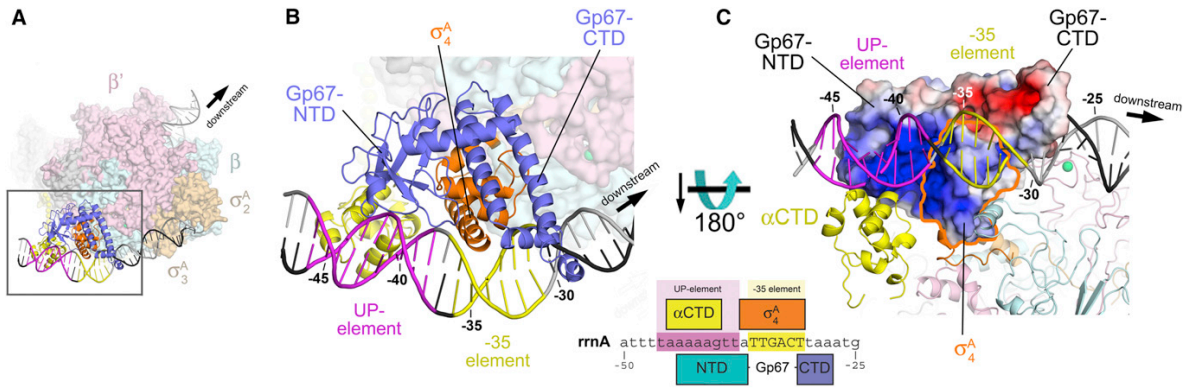
Figure 5. Structural Model of the gp67/RNAP Holoenzyme/Open Promoter Complex

(A) Overall view. The core RNAP is shown as a molecular surface (β', pale pink; β, pale cyan). The σ^A subunit is shown as a molecular surface (σ^A_2 and σ^A_3, pale orange) or a ribbon (σ^A_4, orange). Gp67 is shown as a blue ribbon. One αCTD is shown (yellow ribbon) bound to the proximal UP-element subsite, modeled as described in Jain et al. (2005). The DNA is shown as a phosphate-backbone worm, with the promoter −35 element in yellow and the proximal UP-element subsite (Estrem et al., 1998, 1999) in magenta. The boxed region is magnified in (B).

(B) Magnified view of the −35-element/proximal UP-element region. Shown as in (A). The green sphere denotes a Zn^{2+} ion. The DNA is numbered with respect to the *Sau rrnA* promoter start site (see Figure 6A). The schematic at the lower right shows the −35-element/UP-element region of the *Sau rrnA* promoter. The regions of close protein/DNA interaction for σ^A_4 (orange), gp67 (blue), and αCTD (yellow) are denoted by the span of the boxes.

(C) View toward the gp67/σ^A_4 DNA interaction surface. The RNAP is shown in ribbon format. The molecular surface of the gp67/σ^A_4 complex is colored according to the electrostatic surface distribution (red, −5 kT; white, neutral; blue, +5 kT). The region corresponding to σ^A_4 is outlined in orange.

See also Figure S4 and Tables S1 and S2.

surface area of 2,789 Å^2. Based on the results of our two-hybrid analysis, we identified a limited number of amino acid differences between *Sau* σ^A_4 and *Eco* σ^{70}_4 that determine the specificity of gp67 binding (Figure S1D). In accord with this genetic analysis, the identified residues (*Sau* σ^A D309, E312, N313, and V335) all participate in the gp67/σ^A interface (Figures S4D and S4E).

Most anti-σ factors bind their cognate σ factor and block at least one of the main σ functions: core RNAP binding and/or core promoter element recognition (Campbell et al., 2008; Lambert et al., 2004; Patikoglou et al., 2007). Gp67 does not prevent core RNAP binding of *Sau* σ^A (Figure S1B), and our finding that gp67 does not inhibit transcription from the vast majority of *Sau* −10/−35 promoters in vivo (Figure 3A) is inconsistent with a model in which gp67 blocks recognition of the −35 element by σ^A_4. Indeed, structural modeling, as described below, suggests that core RNAP binding and −35 element DNA binding by σ^A_4 would be allowed in the complex with gp67 (Figures 4B and 4C).

The σ^A_4 binds to core RNAP primarily through an interaction with the β-flap-tip-helix, a structural element of the RNAP β subunit. The σ^A_4/β-flap-tip-helix interaction is essential to position σ^A_4 in the proper orientation and spacing for −35 element recognition (Kuznedelov et al., 2002; Murakami et al., 2002b). One can superimpose the β-flap-tip from the RNAP holoenzyme (Murakami et al., 2002b) on the gp67/σ^A_4 structure (by superimposing the σ^A_4 structural core) without introducing any steric clashes with gp67 (Figure 4B).

Ten highly conserved residues of σ^A_4 make direct contact with −35 element DNA (Campbell et al., 2002; Jain et al., 2004).

All but one of these DNA-contacting residues of σ^A_4 are unperturbed by gp67 and do not make any contacts with gp67 (Figure 4C). The one exception is *Sau* σ^A_4 Arg310 (which corresponds to *Taq* σ^A Arg379/*Eco* σ^{70} Arg554). In the −35 element-bound complex, this Arg residue interacts with the DNA phosphate backbone at the −36 position just upstream of the −35 element (Campbell et al., 2002; Jain et al., 2004). In the gp67 complex, *Sau* σ^A_4 Arg310 is redirected away from the DNA-binding interface and is buried in a deep pocket of gp67 (Figure 4C) where it makes extensive interactions with gp67 residues.

Using either the *Taq* σ^A_4/-35 element DNA complex structure (Protein Data Bank [PDB] ID code 1KU7; Campbell et al., 2002) or an RNAP-holoenzyme open promoter complex (RPo) model to superimpose −35 element DNA on the gp67/σ^A_4 structure reveals minor steric clashes (Figure 5). Nevertheless, based on the following points, we argue that gp67 binding to σ^A_4 does not significantly affect recognition of the −35 promoter element by σ^A_4, and may introduce additional protein/DNA interactions:

1. Our in vitro and in vivo results indicate that gp67 is not a general inhibitor of −10/−35 promoters (Figures 2B and 3A–3C).
2. The minor steric clash with the DNA is at the distal end of the gp67-CTD helical tower, which has relatively high B-factors and is likely to be conformationally flexible. A small rearrangement of the end of the gp67-CTD helical tower and/or the DNA could relieve the clash and appears to facilitate an interaction in the DNA major groove.

199

3. The molecular surface of gp67 that faces the DNA in our model is very basic, particularly upstream of the −35 element (Figure 5C).

In summary, the 2-Å-resolution X-ray crystal structure of the gp67/*Sau*-σ$^A_4$ complex, combined with structural modeling, reveals that gp67 bound to σ$^A_4$ would not interfere with σ$^A_4$ binding to RNAP (Figure 4B), and would be unlikely to completely block σ$^A_4$ interactions with −35-element DNA (Figures 4C and 5). These observations are consistent with our biochemical findings that gp67 forms a ternary gp67/σ^A/RNAP complex (Figure S1B) and does not disrupt −35-element recognition by σ$^A_4$ at most −35-element-dependent promoters (Figure 3A).

### Gp67 Inhibits Transcription from Promoters Containing an A/T-Rich Sequence Upstream of the −35 Element

Because gp67 is unlikely to block −35-element recognition in the context of the RNAP holoenzyme, we hypothesized that other promoter DNA elements may confer susceptibility to gp67 inhibition. We aligned all promoters tested in vitro for direct gp67 activity by their −10 and −35 elements (Figure S5A). The sensitive promoters do not share an obvious common sequence but tend to be A/T rich in the region upstream of the −35 element, where an UP element would be expected (Estrem et al., 1998, 1999), whereas this characteristic was less prominent in promoters resistant to gp67 inhibition (Figure S5A).

To test whether the region upstream of the −35 element was important for gp67 function, we constructed hybrid promoters by swapping the DNA immediately upstream of the −35 element between a gp67-sensitive promoter, *rrnA*, and a gp67-resistant promoter, *dnaA* (Figure 6A). As observed previously, gp67 inhibited transcription from *rrnA* (Figure 6B, lanes 1 and 2). In the absence of gp67, the hybrid *rrnA*(upstream *dnaA*) promoter showed decreased activity, similar to the gp67-inhibited *rrnA* (Figure 6B, lane 3), and gp67 had no additional effect on this hybrid promoter (Figure 6B, lane 4). Gp67 did not significantly affect transcription from the *dnaA* promoter (Figure 6B, lanes 5 and 6). However, when its upstream sequence was replaced by the *rrnA* A/T-rich sequence, the resulting *dnaA*(upstream *rrnA*) promoter became highly sensitive to gp67 (Figure 6B, lanes 7 and 8). The loss of activity between *rrnA* and *rrnA*(upstream *dnaA*) (Figure 6B, lanes 1 and 3) demonstrates that the A/T-rich sequence just upstream of the *rrnA* −35 element contributes significantly to *rrnA* activity. Moreover, these results argue that susceptibility to gp67 inhibition is mediated by a promoter feature upstream of the −35 element, and implicate an A/T-rich sequence in this susceptibility.

Gp67 Alters RNAP Interactions with Promoter DNA Upstream of the −35 Element

To test whether gp67 directly modulates RNAP binding to promoters with A/T-rich elements, we used DNase I footprinting. DNase I cleaves at exposed minor grooves of the DNA double helix. Cleavage is enhanced by deformations or bends in the DNA double helix that widen the minor groove (Fox, 1997).

Because *Sau rrn* promoters have not been tested biochemically, and UP-element binding has not been shown in this organism, we first examined the *Sau rrnA* promoter by
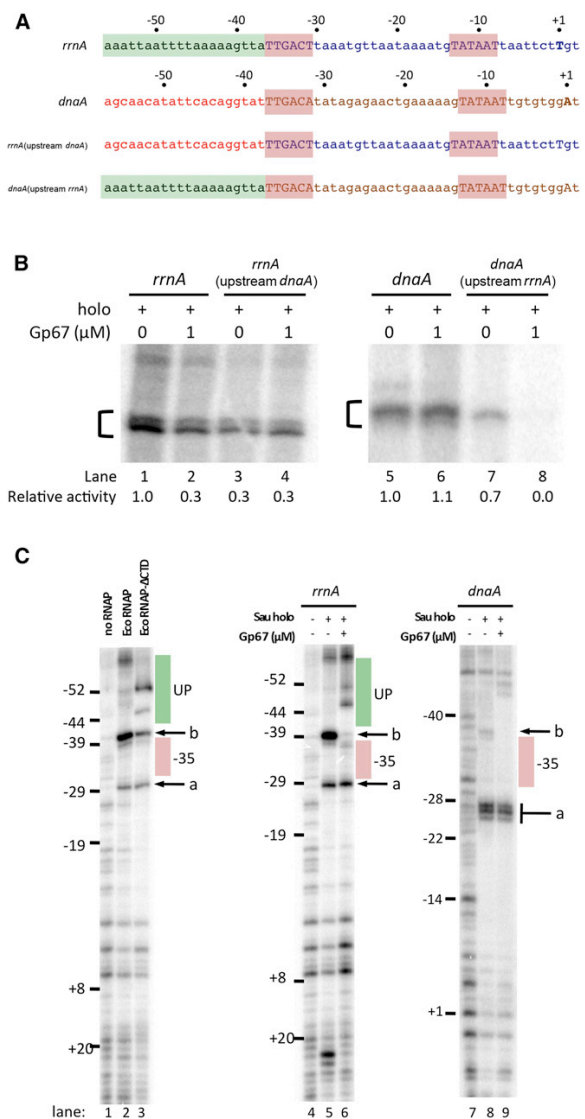
Figure 6. Gp67 Blocks UP-Element Utilization

(A) Schematic of promoters used in swapping and footprinting experiments. DNA positions for the *rrnA* and *dnaA* promoters are labeled relative to the start site (+1). The −10 and −35 elements are shaded red. The putative UP-element region of the *rrnA* promoter is shaded green. The corresponding region of the *dnaA* promoter is denoted by red text.

(B) The region upstream of the −35 element is required for gp67 inhibition. In vitro transcription assays were performed from hybrid promoters constructed by swapping the region upstream of the −35 element between the *rrnA* (gp67-sensitive) and *dnaA* (gp67-resistant) promoters.

(C) DNase I footprinting. Left panel: *Sau rrnA* promoter with *Eco* RNAP and *Eco* RNAP-ΔαCTD. Middle panel: *Sau rrnA* promoter with *Sau* RNAP holoenzyme (±gp67). Right panel: *Sau dnaA* promoter with *Sau* RNAP holoenzyme (±gp67). See also Figure S5.

DNase I footprinting using *Eco* RNAP (Figure 6C, left panel). Binding and distortion of the promoter DNA by *Eco* RNAP are indicated by strong DNase I hypersensitive sites observed on the template strand upon *Eco* RNAP binding (Figure 6C; compare no RNAP, lane 1, with *Eco* RNAP, lane 2) between the −10 and −35 elements (at ∼−29; Figure 6C, band a) and just upstream of the −35 element (at ∼−40; Figure 6C, band b). DNase I hypersensitivity in these regions is commonly observed in *Eco* RNAP/promoter complexes (Ozoline and Tsyganov, 1995). We also determined the DNase I footprint of a mutant *Eco* RNAP lacking the αCTDs (*Eco* RNAP-ΔCTD; Figure 6C, lane 3) to assess the effects of the αCTD/UP-element interaction on the cleavage pattern. In the absence of the αCTD/UP-element interaction, DNase I hypersensitive band a was unaffected, whereas band b was relatively much less intense. In addition, new cleavage sites appeared in the region upstream of the −35 element, at approximately −46 and −52, in the middle of the expected UP-element region (−37 to −56; Estrem et al., 1998; Gourse et al., 2000).

Thus, in terms of DNase I hypersensitivity, the RNAP/promoter complex with the αCTD/UP-element interaction is characterized by very strong hypersensitivity at band b, and an absence of hypersensitive sites in the expected region of the UP element. In the absence of the αCTD/UP-element interaction, band b is much reduced and cleavage sites appear within the region of the UP element.

The DNase I footprint of *Sau* RNAP on the gp67-sensitive *Sau rrnA* promoter is very similar to the *Eco* RNAP footprint (Figure 6D, compare lanes 4 and 5) and displays the characteristics of the αCTD/UP-element interaction (i.e., very strong hypersensitivity at band b and absence of cleavage upstream). The DNase I footprint in the presence of gp67 (Figure 6D, lane 6) is indicative of disrupted αCTD/UP-element interactions: band a and downstream regions of the footprint show few changes, but upstream of the −35 element, the hypersensitive site at −40 (band b) is completely eliminated and hypersensitive sites appear within the UP element (at approximately −47 and −51).

The disappearance of the hypersensitive band b in the presence of gp67 could be explained by protection of the site due to the physical presence of gp67 (since our structural modeling suggests that gp67 may interact with the DNA minor groove between −37 and −44; Figure 5) or by an alteration of the conformation of the DNA induced by gp67. In view of the results comparing *Eco* RNAP and RNAP-ΔCTD (Figure 6C, lanes 2 and 3), disruption of the αCTD/UP-element interaction by gp67 is the best explanation for the appearance of cleavage sites within the UP element with *Sau* RNAP in the presence of gp67 (Figure 6D, lane 6).

The DNase I footprint of *Sau* RNAP (±gp67) on the gp67-insensitive *dnaA* promoter shows qualitatively different features compared with the footprints on the *rrnA* promoter. As with *rrnA*, RNAP binding to the *dnaA* promoter induces DNase I hypersensitivity at sites between the −10 and −35 elements; at the *dnaA* promoter, these sites are at −27, −26, and to a lesser extent −25 (Figure 6C, lanes 7–9). RNAP binding (in the absence of gp67) induces DNase I cleavage at the upstream edge of the −35 element that is not present without RNAP (Figure 6C, compare lanes 7 and 8, band b). However, band b in this case

is weak (relative to band a) compared with band b of the *rrnA* footprints, which is very prominent (and stronger than band a). Moreover, a DNase I cleavage site at approximately −46 is not protected, suggesting that the αCTD binds weakly or not at all to this region. The weak features upstream of band a are altered in the presence of gp67; the weak band b cleavage is eliminated, and new weak cleavage sites appear, likely reflecting the disruption of weak αCTD-DNA interactions.

Thus, the DNase I footprint of *Sau* RNAP on the gp67-insensitive *dnaA* promoter does not show the distinctive features indicative of strong αCTD/UP-element interactions, suggesting that transcription from this promoter is not UP-element dependent. In the presence of gp67, alterations in the cleavage pattern just upstream of the −35 element indicate that gp67 is present in the RNAP-holoenzyme/*dnaA* promoter complex, even though gp67 has little to no effect on transcription output from this promoter (Figures 3C and 6B).

We also determined DNase I footprints for the *rrnA*(upstream *dnaA*) and *dnaA*(upstream *rrnA*) promoters (Figure S5B). On the gp67-insensitive *rrnA*(upstream *dnaA*) promoter, the DNase I cleavage pattern in the region upstream of the −35 element suggests weak αCTD/UP-element interactions (i.e., dramatically weakened hypersensitivity around −39, and distinct cleavage sites between −40 and −56, where one would expect protection due to αCTD binding). On the gp67-sensitive *dnaA*(upstream *rrnA*) promoter, the DNase I cleavage pattern suggests αCTD/UP-element interactions (increased hypersensitivity of band b and protection through the upstream region), which are altered in the presence of gp67.

DISCUSSION

Bacteriophages are the most abundant and diverse form of life on Earth, and exert a major influence over the biosphere. The complete genome sequences of >400 double-stranded DNA phages have been determined. Metagenomic analyses have shown that the phage population is dominated by genetic information that is not related to known sequences (Brüssow and Hendrix, 2002). Bacteriophages are dependent upon a host organism for propagation and have evolved ingenious mechanisms to subvert host cellular processes (such as transcription) for their own needs (Nechaev and Severinov, 2003). In addition to elucidating new mechanisms of bacterial RNAP regulation, the study of phage-encoded regulators can shed light on RNAP function. One can argue that most basic principles of cellular transcription regulation have been exploited by phages, and that the study of phages has revealed many of these mechanisms (Ptashne, 1992).

The use of bacteriophages, including *Sau*-specific phages, for the treatment or prophylaxis of bacterial infectious diseases is experiencing a resurgence, increasing the need to understand bacteriophage/host interactions. Identification and mechanistic analysis of phage proteins that bind and inhibit essential host enzymes can point to potential drug targets and mechanisms (Liu et al., 2004).

Our work on *Sau* phage G1 gp67 has uncovered a novel regulatory mechanism (Figure 7). Upon its expression in the cell, gp67 forms a tight complex with σ^A_4 (Figure 4A), but in a way that does

201
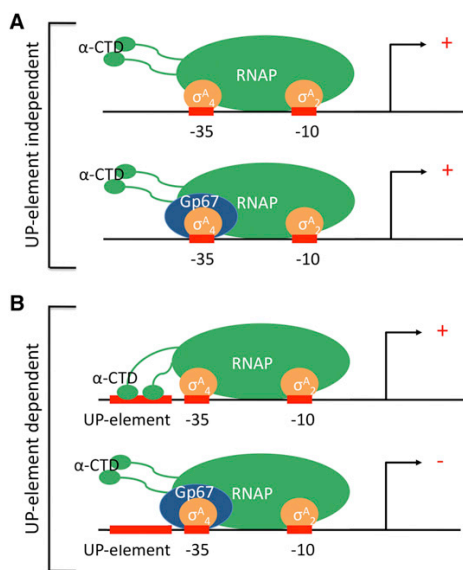
Figure 7. Model for gp67 Function
(A) At most −10/−35 promoters, which are not dependent on αCTD/UP-element interactions, gp67 joins the *Sau* RNAP holoenzyme through its interaction with σ^A_4. However, gp67 does not block any of the functions of σ^A_4 and does not inhibit RNAP.
(B) At promoters where activity depends on αCTD/UP-element interactions, gp67 joins the RNAP holoenzyme through its interaction with σ^A_4 and inhibits transcription initiation by preventing αCTD binding to the promoter-proximal UP-element subsite, or to both proximal and distal subsites (shown).

not block the two primary functions of σ^A_4: core RNAP binding (Figure 4B) and promoter −35 element recognition (Figure 4C). Therefore, gp67 becomes a stable component of the RNAP holoenzyme (Figure S1B) and even of RNAP-holoenzyme/promoter complexes. Nevertheless, the vast majority (∼91%) of promoters in *Sau* are unaffected by gp67 (Figure 3A). Rather than acting as a general anti-σ factor, gp67 inhibits *Sau* transcription in a promoter-specific fashion by selectively targeting promoters depending on the UP-element/RNAP αCTD interaction.

Approximately 90 *Sau* promoters are downregulated by gp67, and 116 promoters are upregulated (Figures 2D and 3A). Six of these in vivo downregulated promoters (*rrnA*, *rrnB*-P1, *rrnB*-P2, *csp1*, *csp2*, and *pstp*; Figure S5A) have been tested in vitro, and all six are directly downregulated by gp67 (Figures 2B and 3E). Based on this limited sample, we postulate that gp67 directly inhibits most, if not all, of the promoters that are downregulated by gp67 expression in vivo. We examined six in vivo upregulated promoters in vitro, but did not observe any evidence for direct, gp67-mediated stimulation. Therefore, we cannot rule out the possibility that the observed in vivo upregulation was an indirect effect of inhibition of rRNA transcription, which would significantly increase the cellular concentration of available RNAP (Barker et al., 2001).

Bacterial transcription initiation is often regulated by activator proteins that typically bind DNA operators upstream of the core promoter and contact the RNAP through the αCTD or through σ_4

(Browning and Busby, 2004). Thus, the mechanism of gp67-mediated inhibition of transcription initiation at UP-element-dependent promoters suggests the possibility that gp67 may interfere with factor-dependent transcription activation as well, a point that needs to be tested in future experiments.

Inhibition of *Sau* Cell Growth by gp67
To aid in the search for targets of small-molecule inhibitors, Xu et al., (2010) identified a set of 308 genes required for *Sau* growth. Among these were the rRNA genes that were shown here to be directly inhibited by gp67 (Figure 2). In addition, seven other genes inhibited by gp67 were required for cell growth in *Sau*. Three of these are also required for translation (Table S3).

Recent work has shown that ribosomal protein promoters and rRNA promoters are regulated by similar mechanisms (Lemke et al., 2011). Given that all but four of the required genes downregulated by gp67 are required for a functional translational machinery, and that cell growth is known to be limited by rRNA synthesis (Gourse et al., 1996), the direct effect of gp67 on rRNA expression specifically, and on expression of the translational machinery in general, is the likely mechanism through which gp67 inhibits cell growth. Further experiments will be required to determine whether gp67 inhibition of other genes that are not involved in translation is sufficient to affect cell growth.

Sau Transcription and Transcriptional Regulation
Previous studies examined transcription and its regulation in vitro in *Sau*, but these studies focused on atypical, highly regulated virulence promoters (Rao et al., 1995; Reyes et al., 2011). Our analysis of gp67 function enabled us to identify σ^A-dependent −10/−35 promoters in *Sau* and to characterize *Sau* rRNA promoters. The molecular tools developed here will be useful for further studies of gene expression and its regulation in this important pathogen. The native *Sau* transcription system (*Sau* RNAP holoenzyme and *Sau* promoters) used here was critical for elucidating the mechanism of gp67 inhibition, whereas the use of a hybrid transcription system (*Eco*-core-RNAP/*Sau*-σ^A/ *Eco*-RNAP promoters) in previous work led to misleading results (Dehbi et al., 2009).

Gp67 and rRNA Synthesis in _Sau_
The αCTD/UP-element interaction activates rRNA expression by >100-fold in *Eco* and by roughly 3-fold in the Gram-positive *Bacillus subtilis* (Krásný and Gourse, 2004). When gp67 is bound to RNAP (suppressing UP-element function), or when the UP element is removed from the *Sau* rRNA promoters in vitro, transcription decreases 2- to 3-fold (Figure 6B), arguing that the UP-element activation of rRNA transcription in *Sau* is more similar to that in *B. subtilis* than that in *Eco*. In vivo, gp67 decreases expression of rRNA promoters more dramatically (∼10-fold; Figure 2D), which may be explained by the competition for RNAP binding between rRNA promoters and other promoters in the genome.

rRNA synthesis is tightly regulated in prokaryotes and is quickly inhibited upon the entry of bacteria into stationary phase (growth-rate control; Haugen et al., 2008a). In *Eco*, growth-rate control of rRNA synthesis involves several mechanisms,

202

including direct modulation of the RNAP by the transcription factor DksA and the small molecule ppGpp (Paul et al., 2004; Perederina et al., 2004). Like DksA and ppGpp, gp67 is a transcription regulator that interacts with RNAP at many promoters, but affects transcriptional output from only a small subset of these promoters due to their unique features. Gp67 specifically takes advantage of the fact that full activity of the rRNA promoters depends on UP-element utilization.

Role of gp67 in the Phage Life Cycle

The genomes of many *Sau* phages have been sequenced (Kwan et al., 2005). The G1 phage is nearly identical to the *Sau* phage K (Kwan et al., 2005), and gp67 shares 100% sequence identity between these phages. Obvious gp67 homologs can be found in five *firmicute*-specific phages. The molecular underpinnings of these phages' life cycles are largely unknown.

A consensus −10/−35 promoter (with an extended −10 element) that is highly active with *Sau* RNAP in vitro (Figure S5C) directs the synthesis of gp67, which is likely produced at high levels early after initial injection of the double-stranded phage genome into the host cell. We hypothesize that gp67 would then engage with the host RNAP (Figure S1B) and suppress production of rRNAs (Figure 2) by selectively inhibiting UP-element-dependent promoters (Figure 6) while allowing transcription from the majority of −10/−35 promoters (Figure 3), including its own promoter and other potential early phage promoters that are resistant to gp67 inhibition (e.g., p*orf67* [Figure S5D] and p*orf05* [Figures 2A and 2B]). The phage ultimately will require the use of host ribosomes to translate phage gene products. Because rRNA is stable in prokaryotic cells (Deutscher, 2003), previously formed ribosomes remain abundant in *Sau* cells for hours after gp67 expression (Figures 2C and 2D).

During log-phase growth, the majority of RNAP in prokaryotic cells is occupied in actively transcribing rRNA. Inhibition of rRNA transcription not only leads to arrest of cell division (Gourse et al., 1996) but also frees a large pool of host RNAP that can then be recruited to the strong phage early promoters. The T4 phage anti-σ factor, AsiA, inhibits host RNAP transcription of the host genome by blocking −35 element recognition while an additional protein, MotA, recruits the AsiA-modified RNAP complex to phage promoters (Hinton et al., 2005). The T4 phage has many additional protein factors that suppress transcription of the host genome and/or favor transcription of the phage genome in a coordinated fashion, including enzymes that ADP-ribosylate the RNAP αCTD, leading to inhibition of UP-element-dependent promoters and recruitment of RNAP to phage promoters (Tiemann et al., 2004). The phage G1 genome does not encode its own RNAP, but relies on the host RNAP for transcription during its entire life cycle, which likely requires complex coordination akin to that observed for phage T4. Therefore, although gp67 expression alone is sufficient to inhibit cell growth (Figure S3; Dehbi et al., 2009; Liu et al., 2004), the existence of additional phage proteins that control the host RNAP throughout the phage life cycle seems likely; however, this remains to be determined. Particularly interesting is the possibility that other phage proteins directly cooperate with gp67 to promote regulated transcription of the phage genome.

Protein Expression and Purification

Gp67 or gp67/σ^A_4 encoding sequences were cloned into pET-based expression vectors, transformed into *Eco* BL21(DE3) cells, and overexpressed, and the proteins were purified using standard methods. The purified complex was dialyzed into crystallization buffer (10 mM Tris-HCl, pH 8.0, 0.5 M NaCl) and screened for crystallization conditions. Endogenous *Sau* RNAP was purified from NCTC8325 cells essentially as previously described (Deora and Misra, 1996).

In Vitro Transcription Assays

In vitro transcription assays were performed using standard methods.

Genetic Analysis of the gp67/*Sau*-σ^A_4 Interaction

A bacterial two-hybrid assay (Dove and Hochschild, 2004) was used to genetically dissect the interaction between gp67 and *Sau* σ^A_4. Taking advantage of the fact that gp67 binds to *Sau* σ^A_4 but not to *Eco* σ^{70}_4, we made a series of *Eco/Sau* σ_4 chimeras in order to define a minimal specificity-determining region (Figure S1D). Having identified four amino acid differences within this region that suffice to dictate whether or not gp67 can bind, we generated an otherwise functional *Sau* σ^A_4 mutant (bearing the corresponding *Eco* residues at *Sau* positions 309, 312, 313, and 335) that did not interact with gp67. We assessed the functional integrity of this mutant by using the two-hybrid assay to test its ability to interact with the β-flap (Figure S1D), and a one-hybrid assay to test its ability to bind a −35 element (data not shown).

Gp67 Expression In Vivo

Gp67 was cloned into the *Sau* expression vector pRMC2 (Corrigan and Foster, 2009). pRMC2-gp67 or empty pRMC2 was then transformed into *Sau* strain RN4220 by electroporation (Schenk and Laddaga, 1992). Cells were then grown in trypticase soy (TS) broth containing chloramphenicol, and transgene expression was induced with 100 ng/ml anhydrotetracycline.

RNA Purification and Metabolic Labeling

RNA was purified from cells at mid-log phase growth (OD$_{600}$ 0.3–0.4) using the RNeasy kit from QIAGEN. In vivo labeling of nascent RNAs was carried out as previously described (Wade et al., 1964).

RNA-seq: Sample Preparation, Sequencing, and Data Analysis

Samples were amplified onto flowcells using an Illumina cBot and sequenced on an Illumina HiSeq2000 for 51 cycles according to the manufacturer's protocols. Raw sequencing data were processed using the onboard SCS/RTA software, yielding 51 bp reads. Sequencing reads were processed using TopHat (Trapnell et al., 2009). Alignments reported from TopHat were processed by the Cufflinks software package (Trapnell et al., 2010) to determine differential expression of genes and transcripts between conditions. Expression values are reported as fragments per kilobase of gene per million mapped reads (FPKM). Data were visualized using the Integrated Genomics Viewer (Robinson et al., 2011).

Crystallization of the gp67/σ^A_4 Complex

Crystals of the gp67/σ^A_4 complex were grown under two different conditions at 22°C. Form I crystals grew from a crystallization solution of 0.1 M 2-ethanesulfonic acid (MES), pH 6.5, 10% (w/v) polyethylene glycol (PEG) 5,000 monomethyl-ether, 20% 1-propanol. The crystals were soaked briefly in crystallization solution supplemented with 15% glycerol and then flash-frozen in liquid nitrogen. Form II crystals grew from a crystallization solution of 0.16 M Ca-acetate, 0.08 M Na-cacodylate, 15% (w/v) PEG 8,000, 20% glycerol. The crystals were flash-frozen in liquid nitrogen directly from the mother liquor. Crystals were formed using a 1:1 ratio of gp67/σ^A_4 complex (10 mg/ml) and reservoir solution. Selenomethionyl-substituted protein was purified and crystallized under the same conditions.

203

Diffraction data were collected at the Advanced Photon Source (Argonne National Laboratory) beamline NE-CAT 24 ID and the National Synchrotron Light Source (Brookhaven National Laboratory) beamline X3A. Both structures were solved by single wavelength anomalous diffraction, and refined against the higher-resolution native data (Tables S1 and S2) to yield the final models.

DNase I Footprinting
DNase I footprinting was performed using standard methods.

ACCESSION NUMBERS

The X-ray crystallographic coordinates and structure factor files have been deposited in the Protein Data Bank under ID Codes 4G6D (form I) and 4G94 (form II).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, five figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2012.10.034.

REFERENCES

Barker, M.M., Gaal, T., and Gourse, R.L. (2001). Mechanism of regulation of transcription initiation by ppGpp. II. Models for positive control based on properties of RNAP mutants and competition for RNAP. J. Mol. Biol. *305*, 689–702.

Blatter, E.E., Ross, W., Tang, H., Gourse, R.L., and Ebright, R.H. (1994). Domain organization of RNA polymerase alpha subunit: C-terminal 85 amino acids constitute a domain capable of dimerization and DNA binding. Cell *78*, 889–896.

Browning, D.F., and Busby, S.J. (2004). The regulation of bacterial transcription initiation. Nat. Rev. Microbiol. *2*, 57–65.

Brüssow, H., and Hendrix, R.W. (2002). Phage genomics: small is beautiful. Cell *108*, 13–16.

Campbell, E.A., Muzzin, O., Chlenov, M., Sun, J.L., Olson, C.A., Weinman, O., Trester-Zedlitz, M.L., and Darst, S.A. (2002). Structure of the bacterial RNA polymerase promoter specificity sigma subunit. Mol. Cell *9*, 527–539.

Campbell, E.A., Greenwell, R., Anthony, J.R., Wang, S., Lim, L., Das, K., Sofia, H.J., Donohue, T.J., and Darst, S.A. (2007). A conserved structural module regulates transcriptional responses to diverse stress signals in bacteria. Mol. Cell *27*, 793–805.

Campbell, E.A., Westblade, L.F., and Darst, S.A. (2008). Regulation of bacterial RNA polymerase sigma factor activity: a structural perspective. Curr. Opin. Microbiol. *11*, 121–127.

Corrigan, R.M., and Foster, T.J. (2009). An improved tetracycline-inducible expression vector for *Staphylococcus aureus*. Plasmid *61*, 126–129.

Darst, S.A. (2001). Bacterial RNA polymerase. Curr. Opin. Struct. Biol. *11*, 155–162.

Datta, I., Sau, S., Sil, A.K., and Mandal, N.C. (2005). The bacteriophage lambda DNA replication protein P inhibits the oriC DNA- and ATP-binding functions of the DNA replication initiator protein DnaA of *Escherichia coli*. J. Biochem. Mol. Biol. *38*, 97–103.

Dehbi, M., Moeck, G., Arhin, F.F., Bauda, P., Bergeron, D., Kwan, T., Liu, J., McCarty, J., Dubow, M., and Pelletier, J. (2009). Inhibition of transcription in *Staphylococcus aureus* by a primary sigma factor-binding polypeptide from phage G1. J. Bacteriol. *191*, 3763–3771.

Deora, R., and Misra, T.K. (1996). Characterization of the primary sigma factor of *Staphylococcus aureus*. J. Biol. Chem. *271*, 21828–21834.

Deutscher, M.P. (2003). Degradation of stable RNA in bacteria. J. Biol. Chem. *278*, 45041–45044.

Dove, S.L., and Hochschild, A. (2004). A bacterial two-hybrid system based on transcription activation. Methods Mol. Biol. *261*, 231–246.

Estrem, S.T., Gaal, T., Ross, W., and Gourse, R.L. (1998). Identification of an UP element consensus sequence for bacterial promoters. Proc. Natl. Acad. Sci. USA *95*, 9761–9766.

Estrem, S.T., Ross, W., Gaal, T., Chen, Z.W.S., Niu, W., Ebright, R.H., and Gourse, R.L. (1999). Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase alpha subunit. Genes Dev. *13*, 2134–2147.

Feklistov, A., and Darst, S.A. (2011). Structural basis for promoter-10 element recognition by the bacterial RNA polymerase σ subunit. Cell *147*, 1257–1269.

Feklistov, A., Barinova, N., Sevostyanova, A., Heyduk, E., Bass, I., Vvedenskaya, I., Kuznedelov, K., Merkiene, E., Stavrovskaya, E., Klimasauskas, S., et al. (2006). A basal promoter element recognized by free RNA polymerase sigma subunit determines promoter recognition by RNA polymerase holoenzyme. Mol. Cell *23*, 97–107.

Fischetti, V.A. (2008). Bacteriophage lysins as effective antibacterials. Curr. Opin. Microbiol. *11*, 393–400.

Fox, K.R. (1997). DNase I footprinting. Methods Mol. Biol. *90*, 1–22.

Gourse, R.L., Gaal, T., Bartlett, M.S., Appleman, J.A., and Ross, W. (1996). rRNA transcription and growth rate-dependent regulation of ribosome synthesis in *Escherichia coli*. Annu. Rev. Microbiol. *50*, 645–677.

Gourse, R.L., Ross, W., and Gaal, T. (2000). UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. Mol. Microbiol. *37*, 687–695.

Gruber, T.M., and Gross, C.A. (2003). Multiple sigma subunits and the partitioning of bacterial transcription space. Annu. Rev. Microbiol. *57*, 441–466.

Haugen, S.P., Ross, W., and Gourse, R.L. (2008a). Advances in bacterial promoter recognition and its control by factors that do not bind DNA. Nat. Rev. Microbiol. *6*, 507–519.

Haugen, S.P., Ross, W., Manrique, M., and Gourse, R.L. (2008b). Fine structure of the promoter-sigma region 1.2 interaction. Proc. Natl. Acad. Sci. USA *105*, 3292–3297.

Hinton, D.M., Pande, S., Wais, N., Johnson, X.B., Vuthoori, M., Makela, A., and Hook-Barnard, I. (2005). Transcriptional takeover by sigma appropriation: remodeling of the sigma70 subunit of *Escherichia coli* RNA polymerase by the bacteriophage T4 activator MotA and co-activator AsiA. Microbiology *151*, 1729–1740.

Howden, B.P., Davies, J.K., Johnson, P.D., Stinear, T.P., and Grayson, M.L. (2010). Reduced vancomycin susceptibility in *Staphylococcus aureus*, including vancomycin-intermediate and heterogeneous vancomycin-intermediate strains: resistance mechanisms, laboratory detection, and clinical implications. Clin. Microbiol. Rev. *23*, 99–139.

Jain, D., Nickels, B.E., Sun, L., Hochschild, A., and Darst, S.A. (2004). Structure of a ternary transcription activation complex. Mol. Cell *13*, 45–53.

Jain, D., Kim, Y., Maxwell, K.L., Beasley, S., Zhang, R., Gussin, G.N., Edwards, A.M., and Darst, S.A. (2005). Crystal structure of bacteriophage λ cII and its DNA complex. Mol. Cell *19*, 259–269.

Kadesch, T.R., Rosenberg, S., and Chamberlin, M.J. (1982). Binding of *Escherichia coli* RNA polymerase holoenzyme to bacteriophage T7 DNA. Measurements of binding at bacteriophage T7 promoter A1 using a template competition assay. J. Mol. Biol. *155*, 1–29.

Keilty, S., and Rosenberg, M. (1987). Constitutive function of a positively regulated promoter reveals new sequences essential for activity. J. Biol. Chem. *262*, 6389–6395.

Krásný, L., and Gourse, R.L. (2004). An alternative strategy for bacterial ribosome synthesis: *Bacillus subtilis* rRNA transcription regulation. EMBO J. *23*, 4473–4483.

Kuznedelov, K., Minakhin, L., Niedziela-Majka, A., Dove, S.L., Rogulja, D., Nickels, B.E., Hochschild, A., Heyduk, T., and Severinov, K. (2002). A role for interaction of the RNA polymerase flap domain with the sigma subunit in promoter recognition. Science *295*, 855–857.

Kwan, T., Liu, J., DuBow, M., Gros, P., and Pelletier, J. (2005). The complete genomes and proteomes of 27 *Staphylococcus aureus* bacteriophages. Proc. Natl. Acad. Sci. USA *102*, 5174–5179.

Lambert, L.J., Wei, Y., Schirf, V., Demeler, B., and Werner, M.H. (2004). T4 AsiA blocks DNA recognition by remodeling sigma70 region 4. EMBO J. *23*, 2952–2962.

Lemke, J.J., Sanchez-Vazquez, P., Burgos, H.L., Hedberg, G., Ross, W., and Gourse, R.L. (2011). Direct regulation of *Escherichia coli* ribosomal protein promoters by the transcription factors ppGpp and DksA. Proc. Natl. Acad. Sci. USA *108*, 5712–5717.

Liu, J., Dehbi, M., Moeck, G., Arhin, F., Bauda, P., Bergeron, D., Callejo, M., Ferretti, V., Ha, N., Kwan, T., et al. (2004). Antimicrobial drug discovery through bacteriophage genomics. Nat. Biotechnol. *22*, 185–191.

Lowy, F.D. (1998). *Staphylococcus aureus* infections. N. Engl. J. Med. *339*, 520–532.

Murakami, K.S., and Darst, S.A. (2003). Bacterial RNA polymerases: the wholo story. Curr. Opin. Struct. Biol. *13*, 31–39.

Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O., and Darst, S.A. (2002a). Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. Science *296*, 1285–1290.

Murakami, K.S., Masuda, S., and Darst, S.A. (2002b). Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 A resolution. Science *296*, 1280–1284.

Nechaev, S., and Severinov, K. (2003). Bacteriophage-induced modifications of host RNA polymerase. Annu. Rev. Microbiol. *57*, 301–322.

Nordmann, P., Naas, T., Fortineau, N., and Poirel, L. (2007). Superbugs in the coming new decade; multidrug resistance and prospects for treatment of *Staphylococcus aureus*, Enterococcus spp. and Pseudomonas aeruginosa in 2010. Curr. Opin. Microbiol. *10*, 436–440.

Ozoline, O.N., and Tsyganov, M.A. (1995). Structure of open promoter complexes with *Escherichia coli* RNA polymerase as revealed by the DNase I footprinting technique: compilation analysis. Nucleic Acids Res. *23*, 4533–4541.

Patikoglou, G.A., Westblade, L.F., Campbell, E.A., Lamour, V., Lane, W.J., and Darst, S.A. (2007). Crystal structure of the *Escherichia coli* regulator of sigma70, Rsd, in complex with sigma70 domain 4. J. Mol. Biol. *372*, 649–659.

Paul, B.J., Barker, M.M., Ross, W., Schneider, D.A., Webb, C., Foster, J.W., and Gourse, R.L. (2004). DksA: a critical component of the transcription initiation machinery that potentiates the regulation of rRNA promoters by ppGpp and the initiating NTP. Cell *118*, 311–322.

Perederina, A., Svetlov, V., Vassylyeva, M.N., Tahirov, T.H., Yokoyama, S., Artsimovitch, I., and Vassylyev, D.G. (2004). Regulation through the secondary channel—structural framework for ppGpp-DksA synergism during transcription. Cell *118*, 297–309.

Ptashne, M. (1992). A Genetic Switch, Phage Lambda and Higher Organisms (Cambridge: Cell Press).

Rao, L., Karls, R.K., and Betley, M.J. (1995). In vitro transcription of pathogenesis-related genes by purified RNA polymerase from *Staphylococcus aureus*. J. Bacteriol. *177*, 2609–2614.

Reyes, D., Andrey, D.O., Monod, A., Kelley, W.L., Zhang, G., and Cheung, A.L. (2011). Coordinated regulation by AgrA, SarA, and SarR to control agr expression in *Staphylococcus aureus*. J. Bacteriol. *193*, 6020–6031.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.

Rosenberg, S., Kadesch, T.R., and Chamberlin, M.J. (1982). Binding of *Escherichia coli* RNA polymerase holoenzyme to bacteriophage T7 DNA. Measurements of the rate of open complex formation at T7 promoter A. J. Mol. Biol. *155*, 31–51.

Ross, W., Gosink, K.K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K., and Gourse, R.L. (1993). A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. Science *262*, 1407–1413.

Schenk, S., and Laddaga, R.A. (1992). Improved method for electroporation of *Staphylococcus aureus*. FEMS Microbiol. Lett. *73*, 133–138.

Shultzaberger, R.K., Chen, Z., Lewis, K.A., and Schneider, T.D. (2007). Anatomy of *Escherichia coli* sigma70 promoters. Nucleic Acids Res. *35*, 771–788.

Stevens, A. (1977). Effect of salt on the transcription of T7 DNA by RNA polymerase from T4 phage-infected *E.coli*. Nucleic Acids Res. *4*, 877–882.

Tiemann, B., Depping, R., Gineikiene, E., Kaliniene, L., Nivinskas, R., and Rüger, W. (2004). ModA and ModB, two ADP-ribosyltransferases encoded by bacteriophage T4: catalytic properties and mutation analysis. J. Bacteriol. *186*, 7262–7272.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

Travers, A.A. (1984). Conserved features of coordinately regulated *E. coli* promoters. Nucleic Acids Res. *12*, 2605–2618.

Vassylyev, D.G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M.N., Borukhov, S., and Yokoyama, S. (2002). Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. Nature *417*, 712–719.

Wade, H.E., Lovett, S., and Robinson, H.K. (1964). The autodegradation of 32-P-labelled ribosomes from *Escherichia coli*. Biochem. J. *93*, 121–128.

Xu, H.H., Trawick, J.D., Haselbeck, R.J., Forsyth, R.A., Yamamoto, R.T., Archer, R., Patterson, J., Allen, M., Froelich, J.M., Taylor, I., et al. (2010). *Staphylococcus aureus* TargetArray: comprehensive differential essential gene expression as a mechanistic tool to profile antibacterials. Antimicrob. Agents Chemother. *54*, 3659–3670.

Yano, S.T., and Rothman-Denes, L.B. (2011). A phage-encoded inhibitor of *Escherichia coli* DNA replication targets the DNA polymerase clamp loader. Mol. Microbiol. *79*, 1325–1338.

205

Supplemental Information

EXTENDED EXPERIMENTAL PROCEDURES

Protein Expression and Purification

The gp67/σ^A_4 complex was cloned into a single operon as described (Campbell and Darst, 2000). The complex was expressed in *Eco* BL21(DE3) cells with 1 mM IPTG at 37°C for 3 hr and purified using Ni-affinity chromatography. The His$_6$-tag was subsequently cleaved from σ^A_4 by incubation overnight with Precision protease (GE Healthcare) and the complex was further purified by subtractive Ni-affinity chromatography, and finally size exclusion chromatography (Superdex 75, GE Healthcare). Gp67 and σ^A_4 formed a stable, stoichiometric complex throughout the purification. We dialyzed the complex into crystallization buffer (10 mM Tris-HCl, pH 8.0, 0.5 M NaCl) and screened for crystallization conditions.

Gp67 was cloned into pET29a (Novagen) using NdeI and XhoI, removing the native stop codon, to produce a C-terminally His-tagged protein. Gp67 was expressed in BL21(DE3) cells with 0.5 mM IPTG overnight at 18°C and purified using standard Ni-affinity chromatography.

Sau σ^A was cloned into pSKB2, a modified pET vector, using NheI and HindIII. The His-tagged protein was expressed in BL21(DE3) plysS cells at 25°C for 5 hr. The protein was purified using Ni-affinity chromatography, dialysed into low salt buffer, and subsequently purified on a Q-sepharose column (GE Healthcare). Contaminating *Eco* core RNAP was removed by a final purification step using size-exclusion chromatography (Superdex 75, GE Healthcare). The σ^A_4 mutant was made by megaprimed-PCR and cloned into the same expression vector.

Purification of Sau RNAP

Sau RNAP was purified natively from *Sau* cells essentially as described (Deora and Misra, 1996). Briefly, *Sau* NCTC8325 cells were grown to an O.D. of 1.0, collected by centrifugation, washed in a high salt buffer, and resuspended in grinding buffer [(TGED: 10 mM Tris HCl, pH 8.0, 5% glycerol, 0.1 mM EDTA, 1 mM DTT) + 0.2 M NaCl]. Cells were lysed by French-Press (Avestin Emusiflex C-50) and the cleared lysate was precipitated with polethyleneimine at 0.6% (v/v). The pellets were washed with TGED + 0.45 M NaCl, RNAP was eluted with TGED + 1.0 M NaCl and precipitated with 35% (w/v) ammonium sulfate. The pelleted protein was resuspended in TGED and diluted to a conductivity equal to TGED + 0.1 M NaCl. Protein was purified sequentially by a heparin column (GE Healthcare) and gel filtration column (Superdex 200, GE Healthcare). An S-sepharose column (GE Healthcare) was required as a final step to remove natively formed holoenzyme. RNAP was stored at −20°C in *Sau* protein storage buffer (10 mM Tris-HCl, pH 8.0, 0.15 M Na-glutamate, 15% glycerol, 1 mM DTT).

In Vitro Transcription Assays

Proteins used in in vitro transcription assays were diluted into 1x protein storage buffer. *Sau* σ^A (100 nM) was preincubated with gp67 for 10 min on ice, followed by the addition of *Sau* core RNAP (50 nM). After 10 min, DNA (50 ng of genomic DNA or 50 nM of purified promoter DNA) was added and the reaction brought to 20 μl in 1x *Sau* transcription buffer (40 mM Tris-acetate, pH 7.9, 10 mM MgCl$_2$, 1 mM EDTA, 50 ug/ml BSA, 100 mM NaCl, 1 mM DTT). Reactions were incubated for 10 min at 37°C to form open promoter complexes and then transcription was initiated with NTPs (200 uM each GTP, CTP, UTP, 50 uM ATP, 0.1 ul α-[P^{32}]ATP). After 5 min, reactions were stopped with 2x formamide buffer (98% formamide, 5 mM EDTA) and run on a 12% Urea-PAGE gel. Products were visualized by phosphorimagery and, where applicable, quantified using ImageQuant. For the assays using genomic DNA as the transcription template, reactions were stopped with 20 mM EDTA, 0.5% SDS, pipetted onto Whatman DE81 paper, washed 5 times with Na-phosphate (50 g/l), rinsed, dried and visualized by phosphorimagery. Promoter fragments were prepared by PCR from *Sau* genomic DNA, purified on a 1.5% agarose gel, and electroeluted into dialysis tubing. Following phenol/chloroform extraction, DNA was ethanol precipitated and resuspended in 10 mM Tris-HCl, pH 8.0, at 1 μM final concentration.

Genetic Analysis of the gp2/Sau-σ^A_4 Interaction

A bacterial two-hybrid assay (Dove and Hochschild, 2004) was used to genetically dissect the interaction between gp67 and *Sau* σ^A_4. In this assay contact between a protein domain (here σ_4) fused to the RNAP α subunit N-terminal domain (αNTD) and a partner protein (here gp67) fused to the bacteriophage λCI protein activates transcription from a test promoter bearing an upstream λ operator. We first used the assay to confirm that gp67 interacts with *Sau* σ^A_4, but not with *Eco* σ^{70}_4. We then sought to identify the relevant amino acid differences. An alignment of the *Eco* σ^{70}_4 and *Sau* σ^A_4 amino acid sequences (see Figure S1D) indicated that most of the amino acid differences were concentrated in the N-terminal half of the domain. We constructed a series of reciprocal *Eco*/*Sau* σ_4 chimeras to define a minimal specificity-determining region. We tested all chimeras for their abilities to interact with the *Eco* β-flap to distinguish between properly folded variants and misfolded or destabilized variants. Additionally, a one-hybrid assay (Nickels, 2009) enabled us to test the chimeras for their abilities to interact with a −35 element; in all cases, chimeras that were competent to interact with the β-flap were also competent to interact with the −35 element (data not shown). An initial set of constructs revealed that all of the relevant amino acid differences were located on the N-terminal side of *Sau* σ^A residue 338 (*Eco* σ^{70} residue 582), but that the *Eco* σ^{70}_4 N-terminal extension (residues 528-540) was not responsible for the inability of gp67 to bind to *Eco* σ^{70}_4. Additional constructs enabled us to define the minimal specificity-determining region comprising *Sau* σ^A residues 309-335 (*Eco* σ^{70} residues 553-579). Focusing on the amino acid differences within this region, we then tested the effects of reciprocal amino acid substitutions in various

206

combinations. The results of this analysis identified 4 amino acid differences (*Sau* positions 309, 312, 313 and 335) that fully account for the ability and inability of *Sau* σ^A_4 and *Eco* σ^{70}_4, respectively, to interact with gp67.

β-Galactosidase Assays

For the bacterial two-hybrid assays, reporter strain FW102 O_L2-62 (Nickels, 2009) was cotransformed with two compatible plasmids: one encoding either α or the indicated α-σ_4 fusion protein, and the other encoding either a λCI-gp67 fusion protein or a λCI-β flap fusion protein. These plasmids direct the synthesis of the fusion proteins (or α) under the control of IPTG-inducible promoters. Individual transformants were selected and three independent isolates of each were grown overnight in LB supplemented with kanamycin (50 μg/ml), carbenicillin (50 μg/ml), and chloramphenicol (25 μg/ml) in the absence of IPTG. Overnight cultures were diluted 1:100 and grown in LB supplemented with kanamycin, carbenicillin, chloramphenicol and 20 μM IPTG to mid-exponential phase (O.D.$_{600}$ 0.3–0.4). β-galactosidase assays were performed as described (Nickels, 2009) using microtiter plates and a microtiter plate reader. Miller units were calculated as described (Nickels, 2009). The results shown in Figure S1D are the averages of the 3 independent measurements with standard deviations.

Gp67 Expression In Vivo

Gp67 was cloned into the *Sau* expression vector pRMC2 (Corrigan and Foster, 2009) using primers containing a consensus Shine-Delgarno sequence, a BglII site upstream of the start codon, and a stop codon and EcoRI site downstream. pRMC2-gp67, pRMC2-Twort-gp65 and empty pRMC2 were then transformed into *Sau* strain RN4220 by standard electroporation (Schenk and Laddaga, 1992) and transformants selected on trypticase soy (TS) plates containing chloramphenicol (10 μg/ml). RN4220 containing empty pRMC2 and pRMC2-gp67 were then grown in TS broth containing chloramphenicol and transgene expression was induced with 100 ng/ml anhydrotetracycline, which was the minimum required concentration for maximal cell growth inhibition by gp67.

RNA Purification and Metabolic Labeling

RNA was purified from cells at mid-log phase growth (O.D.$_{600}$ 0.3–0.4) using the RNeasy kit (QIAGEN). Briefly, 2×10^8 cells were removed from growing cultures, immediately added to 2 volumes of BioStabilize solution, and incubated for 5 min at room temperature. Cells were then collected by centrifugation, resuspended in TE buffer containing 1 mg/ml lysostaphin, and 200 μg proteinase K, and incubated for 15 min at room temperature. Zirconia beads (100 μl, 0.1 mm) were added and the cells lysed for 3 × 2 min, with a 1 min rest on ice, in a bead-beater at top speed. The lysate was centrifuged briefly to remove the beads and the remaining procedure was carried out to the manufacturer's specifications. Purified RNA was quantified using a NanoDrop spectrometer. For RNA-seq analysis, rRNAs were removed to enrich the relative mRNA signal using a Ribo-Zero rRNA removal kit designed for gram-positive organisms (Epicenter) using the manufacturer's protocol. cDNA libraries were prepared using the mRNA-seq sample prep kit (Illumina), eliminating the step for amplification of poly-A tailed mRNA.

In vivo labeling of nascent RNAs was carried out as described (Wade et al., 1964). Briefly, cells containing pRMC2-gp67 or empty pRMC2 were induced with 100 ng/ml anhydrotetracycline at O.D.$_{600}$ 0.2, allowed to grow for 40 min (1 normal doubling time), after which 200 μCi orthophosphoric acid was added directly to the growth medium. 2×10^8 cells were collected after 20 min and RNA purified as described above. To visualize rRNA, total RNA was run on a 6% polyacrylamide gel, which was stained with 1x GelRed (Biotium). The same gel was then exposed to a phosphoimaging cassette and labeled rRNA quantified using ImageQuant.

RNA-seq: Sample Preparation and Sequencing

Samples were amplified onto flowcells using an Illumina cBot and sequenced on an Illumina HiSeq2000 for 51 cycles per manufacturer protocols. Raw sequencing data were processed using the onboard SCS/RTA software yielding 51-bp reads.

RNA-seq: Data Analysis

Sequencing reads were processed using TopHat (Trapnell et al., 2009), an alignment package designed to align sequencing reads derived from transcribed RNA. Briefly, the program aligns reads to a reference genome, identifying regions of coverage that correspond to transcribed RNA. These regions are joined and queried for potential junctions by attempting alignment of reads that did not initially align. Reads aligning to multiple locations are kept (to a maximum of 20 potential positions) to assist constructing gene models for genes with repetitive or low complexity features. When aligning reads, two mismatches to the reference (Ensembl S_aureus_nctc_8325.EB1.fa) were allowed.

Alignments reported from TopHat were processed by the Cufflinks software package (Trapnell et al., 2010) to determine differential expression of genes and transcripts between conditions. Alignments were quantified against the Ensembl annotation: (S_aureus_nctc_8325.EB1_s_aureus_nctc_8325.gtf). Expression values are reported as fragments-per-kilobase-of-gene-per-million-mapped reads (FPKM). Data were visualized using the Integrated Genomics Viewer (Robinson et al., 2011).

Crystallization of the gp67/Domain 4 Complex

Crystals of the gp67/σ^A_4 complex were grown under two different conditions at 22°C. Form I crystals grew from a crystallization solution of 0.1 M MES, pH 6.5, 10% (w/v) PEG 5000 monomethyl-ether, 20% 1-propanol. The crystals were soaked briefly in crystallization solution supplemented with 15% glycerol before being flash-frozen in liquid nitrogen. Form II crystals grew from

a crystallization solution of 0.16 M Ca-acetate, 0.08 M Na-cacodylate, 15% (w/v) PEG 8000, 20% glycerol. The crystals were flash-frozen in liquid nitrogen directly from the mother liquor. Sitting and hanging drops were used and crystals were formed using a 1:1 ratio of gp67/σ^A_4 complex (10 mg/ml) and reservoir solution. Selenomethionyl-substituted protein was purified and crystallized under the same conditions.

Diffraction data were collected at the Advanced Photon Source (APS, Argonne National Laboratory) beamline NE-CAT 24 ID and at the National Synchrotron Light Source (NSLS, Brookhaven National Laboratory) beamline X3A. The data were processed using HKL2000 (Otwinowski and Minor, 1997), selenomethioninyl sites were located using SnB (Weeks et al., 2002), single wavelength anomalous diffraction phases calculated using SHARP, and density modification using SOLOMON (de La Fortelle et al., 1997). Coot (Emsley and Cowtan, 2004) was used for model building and the structures were refined using Phenix (Adams et al., 2010). The high-resolution Form I model was used late in the refinement process for the Form II structure to aid in the modeling of poorly defined density.
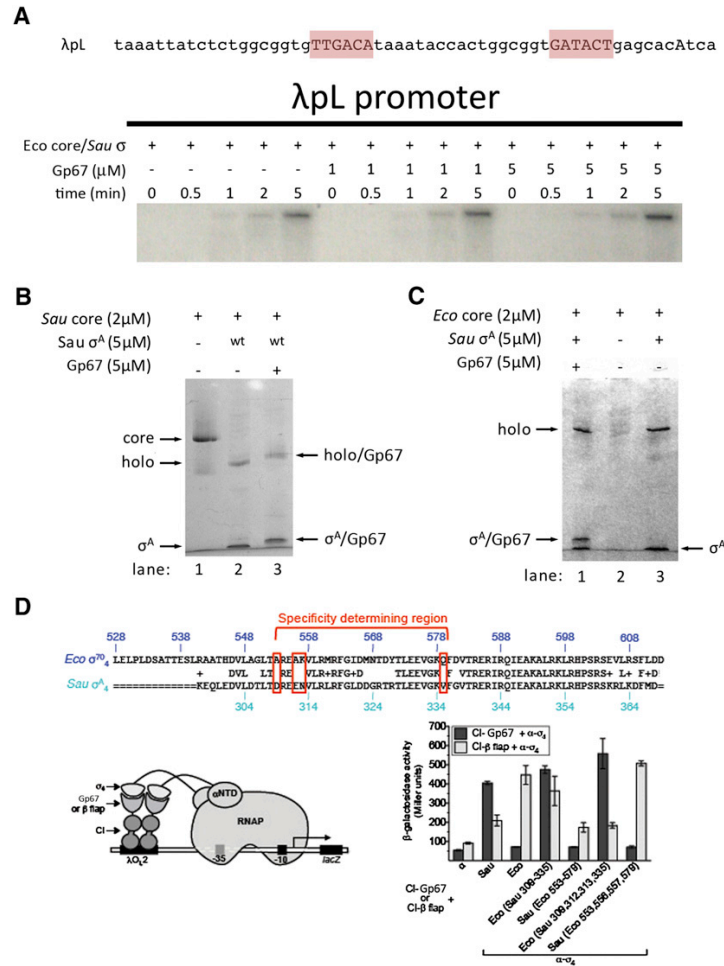
DNase I Footprinting

DNA fragments were made by PCR and end-labeled on the template strand. The reverse PCR oligo (100 pmol) was gel purified, P^{32} end labeled with polynucleotide kinase, and unincorporated nucleotide was removed on a sephadex G-50 spin column. The radio-labeled primer was then added to a standard PCR reaction and the PCR product was gel purified.

Reactions were performed by forming a complex between σ^A (5 μM), gp67 (5 μM), and core RNAP (2 μM) on ice. Labeled promoter DNA (0.1 μl) and DNase I reaction buffer (5 mM Tris-acetate, pH 7.9, 5 mM KCl, 1 mM MgCl$_2$) were added and incubated at 37°C for 10 min. DNase I (0.1 μg/ml) was added for 1 min and reactions stopped with 15 mM EDTA. Samples were boiled for 5 min, separated by 6% UREA PAGE and visualized on a phosphoimager screen.

SUPPLEMENTAL REFERENCES

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213–221.

Campbell, E.A., and Darst, S.A. (2000). The anti-σ factor SpoIIAB forms a 2:1 complex with σ(F), contacting multiple conserved regions of the σ factor. J. Mol. Biol. *300*, 17–28.

Corrigan, R.M., and Foster, T.J. (2009). An improved tetracycline-inducible expression vector for *Staphylococcus aureus*. Plasmid *61*, 126–129.

de La Fortelle, E., Irwin, J.J., and Bricogne, G. (1997). SHARP: A maximum-likelihood heavy-atom parameter refinement and phasing program for the MIR and MAD methods. In Crystallographic Computing, P. Bourne and K. Watenpaugh, eds. (Boston: Kluwer Academic Publishers), pp. 1–9.

Deora, R., and Misra, T.K. (1996). Characterization of the primary sigma factor of *Staphylococcus aureus*. J. Biol. Chem. *271*, 21828–21834.

Dove, S.L., and Hochschild, A. (2004). A bacterial two-hybrid system based on transcription activation. Methods Mol. Biol. *261*, 231–246.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Kuznedelov, K., Minakhin, L., Niedziela-Majka, A., Dove, S.L., Rogulja, D., Nickels, B.E., Hochschild, A., Heyduk, T., and Severinov, K. (2002). A role for interaction of the RNA polymerase flap domain with the sigma subunit in promoter recognition. Science *295*, 855–857.

Nickels, B.E. (2009). Genetic assays to define and characterize protein-protein interactions involved in gene regulation. Methods *47*, 53–62.

Nickels, B.E., Garrity, S.J., Mekler, V., Minakhin, L., Severinov, K., Ebright, R.H., and Hochschild, A. (2005). The interaction between sigma70 and the beta-flap of *Escherichia coli* RNA polymerase inhibits extension of nascent RNA during early elongation. Proc. Natl. Acad. Sci. USA *102*, 4488–4493.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.

Schenk, S., and Laddaga, R.A. (1992). Improved method for electroporation of *Staphylococcus aureus*. FEMS Microbiol. Lett. *73*, 133–138.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

Wade, H.E., Lovett, S., and Robinson, H.K. (1964). The autodegradation of 32-P-labelled ribosomes from *Escherichia coli*. Biochem. J. *93*, 121–128.

Weeks, C.M., Blessing, R.H., Miller, R., Mungee, R., Potter, S.A., Rappleye, J., Smith, G.D., Xu, H., and Furey, W. (2002). Toward automated protein structure determination: BnP, the SnB-PHASES interface. Z. Kristallogr. *217*, 686–693.

A

λpL   taaattatctctggcggtgTTGACAtaaataccactggcggtGATACTgagcacAtca

λpL promoter

Eco core/*Sau* σ	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Gp67 (μM)	-	-	-	-	-	1	1	1	1	1	5	5	5	5	5
time (min)	0	0.5	1	2	5	0	0.5	1	2	5	0	0.5	1	2	5

B

Sau core (2μM)	+	+	+
Sau σ^A (5μM)	-	wt	wt
Gp67 (5μM)	-	-	+

core →
holo →   ← holo/Gp67
σ^A →   ← σ^A/Gp67
lane: 1   2   3

C

Eco core (2μM)	+	+	+
Sau σ^A (5μM)	+	-	+
Gp67 (5μM)	+	-	-

holo →
σ^A/Gp67 →   ← σ^A
lane: 1   2   3

D

Specificity determining region

Eco σ$^{70}_4$ LELPLDSATTESLRAATHDVLAGLTAREAEVLRMRFGIDMNTDYTLEEVGKQFDVTRERIRQIEAKALRKLRHPSRSEVLRSFLDD

   +   DVL   LT RE VLR+RFG+D   TLEEVGK F VTRERIRQIEAKALRKLRHPSRS+ L+ F+D

Sau σ$^A_4$ ============KEQLEDVLDTLTDRFENVLRLRFGLDDGRTRTLEEVGKVFGVTRERIRQIEAKALRKLRHPSRSKRLKDFMD=

Figure S1. Gp67 Binds Specifically to *Sau* σ$^A_4$ but Does Not Bind to or Inhibit a *Sau*-σ^A/*Eco*-core Hybrid Holoenzyme, Related to Figure 1

(A) Gp67 does not inhibit transcription by a *Sau*-σ^A/*Eco*-core hybrid holoenzyme on the λpL promoter. Gp67, at 1 μM or 5 μM, was bound to *Sau* σ^A (100 nM) and *Eco*-core RNAP (50 nM). Promoter DNA (λpL, 50 nM) was added and the reaction was incubated in 1× transcription buffer at 37°C. Reactions were initiated with NTPs (200 μM each GTP, CTP, and UTP, 50 μM ATP, and 0.1 μl α-[P^{32}]ATP), stopped with 2× stop buffer at the times indicated, and visualized by 12% urea-PAGE and autoradiography.

(B) Gp67 binds to *Sau* holoenzyme. A native gel-shift analysis shows the interaction between gp67 and *Sau* RNAP. *Sau*-core RNAP (2 μM) was incubated with *Sau* σ^A (5 μM) and gp67 (5 μM). Complexes were visualized by Coomassie staining on a 4%–12% native PhAST gel.

(C) Gp67 does not interact with the *Sau*-σ^A/*Eco*-core hybrid RNAP. *Eco* core was incubated with *Sau* σ^A and gp67 at the indicated concentrations and visualized by Coomassie staining on a 4%–12% native PhAST gel.

(D) Use of the bacterial two-hybrid assay to genetically dissect the gp67/*Sau*-σ$^A_4$ interaction. Shown at the top is an alignment of the *Eco* σ$^{70}_4$ and *Sau* σ$^A_4$ amino acid sequences, with identical and similar residues indicated. Shown below (left side) is a cartoon depicting how the interaction between *Sau* σ$^A_4$ (fused to the αNTD) and gp67 (fused to the bacteriophage λCI protein) activates transcription from the test promoter placO$_L$2–62, which bears λ operator O$_L$2 centered 62 bp upstream of the start site of the *lac* core promoter. In strain FW102, O$_L$2–62, the test promoter, which directs transcription of a *lacZ* reporter gene, is located on an F' episome. The graph (right side) shows the results of β-galactosidase assays done to assess the ability of a panel of σ$_4$ variants to interact with gp67 (black bars) or with the *Eco* β-flap (residues 831–1057, gray bars). The following native and chimeric σ$_4$ variants were tested: *Sau* σ$^A_4$ (*Sau*); *Eco* σ$^{70}_4$ (*Eco*); *Eco* σ$^{70}_4$ bearing *Sau* σ^A residues 309–335 in place of the corresponding *Eco* residues (*Eco* [*Sau* 309–335]); *Sau* σ$^A_4$ bearing *Eco* σ^{70} residues 553–579 in place of the corresponding *Sau* residues (*Sau* [*Eco* 553–579]); *Eco* σ$^{70}_4$ bearing *Sau* σ^A residues 309, 312, 313, and 335 in place of the corresponding *Eco* residues (*Eco* [*Sau* 309, 312, 313, 335]); and *Sau* σ$^A_4$ bearing *Eco* σ^{70} residues 553, 556, 557 and 579 in place of the corresponding *Sau* residues (*Sau* [*Eco* 553, 556, 557, 579]). In all cases, the *Eco* σ$^{70}_4$ moiety bore substitution D581G, which stabilizes the folded structure of *Eco* σ$^{70}_4$ and facilitates the detection of two-hybrid interactions (Kuznedelov et al., 2002; Nickels et al., 2005). *Sau* σ$^A_4$ bears a glycine at the corresponding position. The results indicate that *Eco* σ$^{70}_4$ bearing just four *Sau* residues is fully competent to interact with gp67 and, conversely, that *Sau* σ$^A_4$ bearing just four *Eco* residues can no longer interact detectably with gp67. Importantly, the ability of this quadruply substituted *Sau* σ$^A_4$ variant to interact with the β-flap indicates that it is properly folded and that the fusion protein is stably produced. Error bars are ± SD of three independent assays.

209
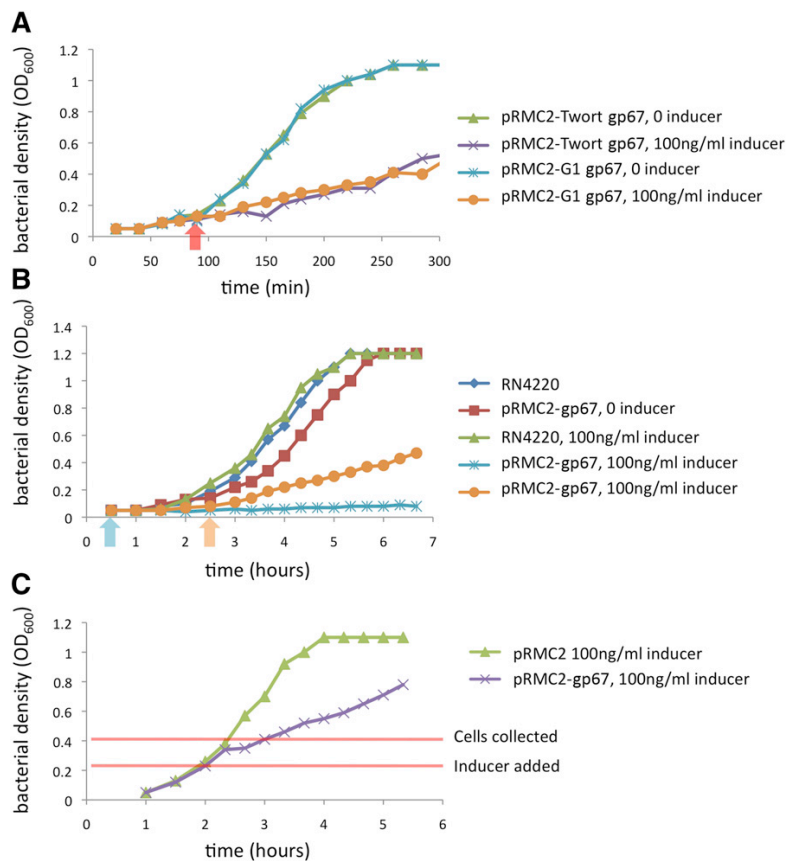
Figure S2. Gp67 Expression in *Sau* Cells Inhibits Cell Growth, Related to Figures 2 and 3

(A) Gp67 expression in *Sau* cells inhibits cell growth. Gp67 and the Phage Twort gp67 homolog, gp65, were cloned into the *Sau* expression vector pRMC2 and transformed into *Sau* RN4220 cells by electroporation. Cells containing pRMC2-Phage G1 gp67, pRMC2-Phage Twort gp65, or empty pRMC2 were grown in TS media containing chloramphenicol (25 μg/ml). G1-gp67 or Twort-gp65 expression was induced by adding 100 ng/ml anhydrotetracycline at the time indicated by the arrow, and bacterial growth was monitored by the OD_{600}.

(B) Gp67 does not fully inhibit cell growth. When gp67 expression is induced at the beginning of the culture (as indicated by the blue arrow and blue curve), no cell growth is evident. However, when gp67 expression is induced after cells have entered the exponential growth phase (as indicated by the orange arrow and orange curve), cell growth is inhibited compared with normally growing cells (green, blue, and red curves), but not completely.

(C) Cells grown for subsequent RNA-seq experiments. To purify cellular RNA for genomic analysis, RN4220 cells containing empty pRMC2 or pRMC2-gp67 were grown and inducer was added at an OD_{600} of 0.2 (lower red line). Cells were collected at OD_{600} of 0.4 (upper red line) for RNA purification.
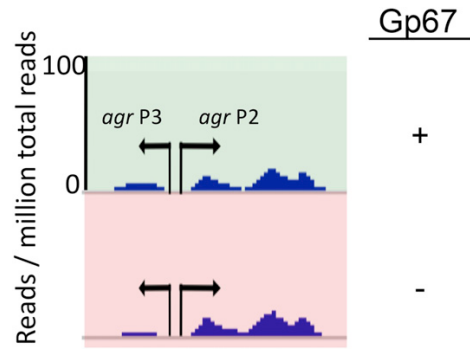
Figure S3. Gp67 Does Not Affect RNA Levels at the *agr* Promoters In Vivo, Related to Figure 3

RNA-seq data showing RNA levels at the *agr* operons in the absence (lower panel, red) or presence (upper panel, green) of gp67.

211
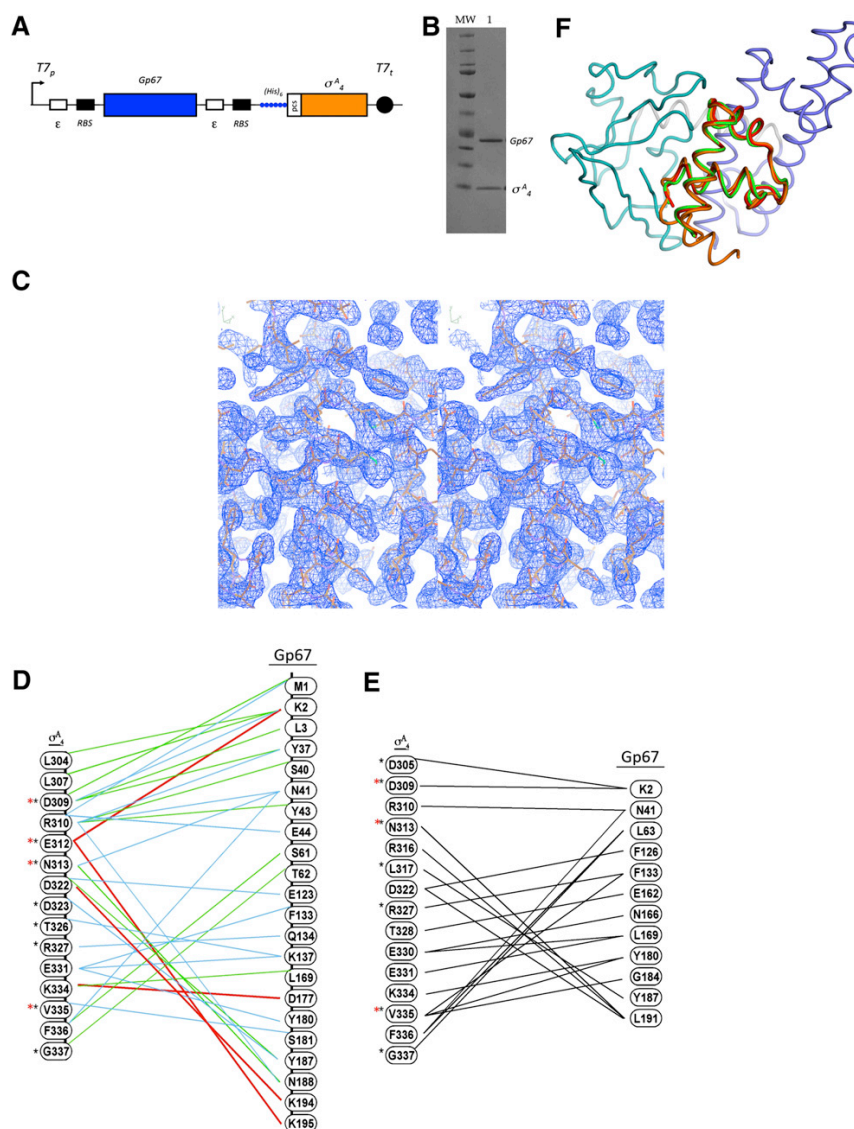
Figure S4. Purification, Crystallization, and Structural Analysis of the gp67/σ$^A_4$ Complex, Related to Figures 4 and 5

(A) Gp67/σ$^A_4$ coexpression cassette. Gp67 and σ$^A_4$ were cloned into a pET-based vector for coexpression in *Eco* cells (Campbell and Darst, 2000). σ$^A_4$ was tagged with a Precision protease cleavable His$_6$-tag that was cleaved during the purification of the complex. The gp67/σ$^A_4$ complex was stable over the course of the purification and ran as a heterodimer on a gel filtration column.

(B) SDS-PAGE of the purified gp67/σ$^A_4$ complex. Lane 1: The purified complex (5 μg) was run on a 4%–12% NuPAGE gel. MW, Precision Plus Protein Dual Color molecular weight markers (BioRad).

(C) Experimental electron density map (shown in stereo). Phases were calculated by SHARP using the anomalous signal from the selenomethioninyl-substituted gp67/σ$^A_4$ complex, with density modification using SOLOMON (de La Fortelle et al., 1997). The final model of the gp67/σ$^A_4$ complex is shown in yellow.

(D) Polar and ionic interactions between σ$^A_4$ and gp67. Residues that differ between *Eco* and *Sau* are indicated with a black asterisk. Green lines indicate hydrogen-bonding interactions, blue lines indicate hydrogen bonds mediated by water molecules, and red lines indicate ionic (salt-bridge) interactions. Residues that eliminated the σ$^A_4$/gp67 interaction when they were switched to the *Eco* sequence (see Figure S2D) are indicated with a red asterisk.

(E) Hydrophobic (van der Waals) interactions (≤4 Å) between σ$^A_4$ and gp67. Residues that differ between *Eco* and *Sau* are indicated with a black asterisk. Residues that eliminated the σ$^A_4$/gp67 interaction when they were switched to the *Eco* sequence (see Figure S2D) are indicated with a red asterisk.

(F) Gp67 does not conformationally rearrange σ$^A_4$. Gp67 (teal/blue) is shown in complex with Sau σ$^A_4$ (orange). The structural core of σ$^A_4$ (*Sau* residues 307–355) was aligned with the structures of *Taq* σ$^A_4$ alone (red, PDB ID code 1KU3; rmsd = 1.15) and *Taq* σ$^A_4$ in complex with a promoter DNA element (green, PDB ID code 1KU7; rmsd = 1.21).

212
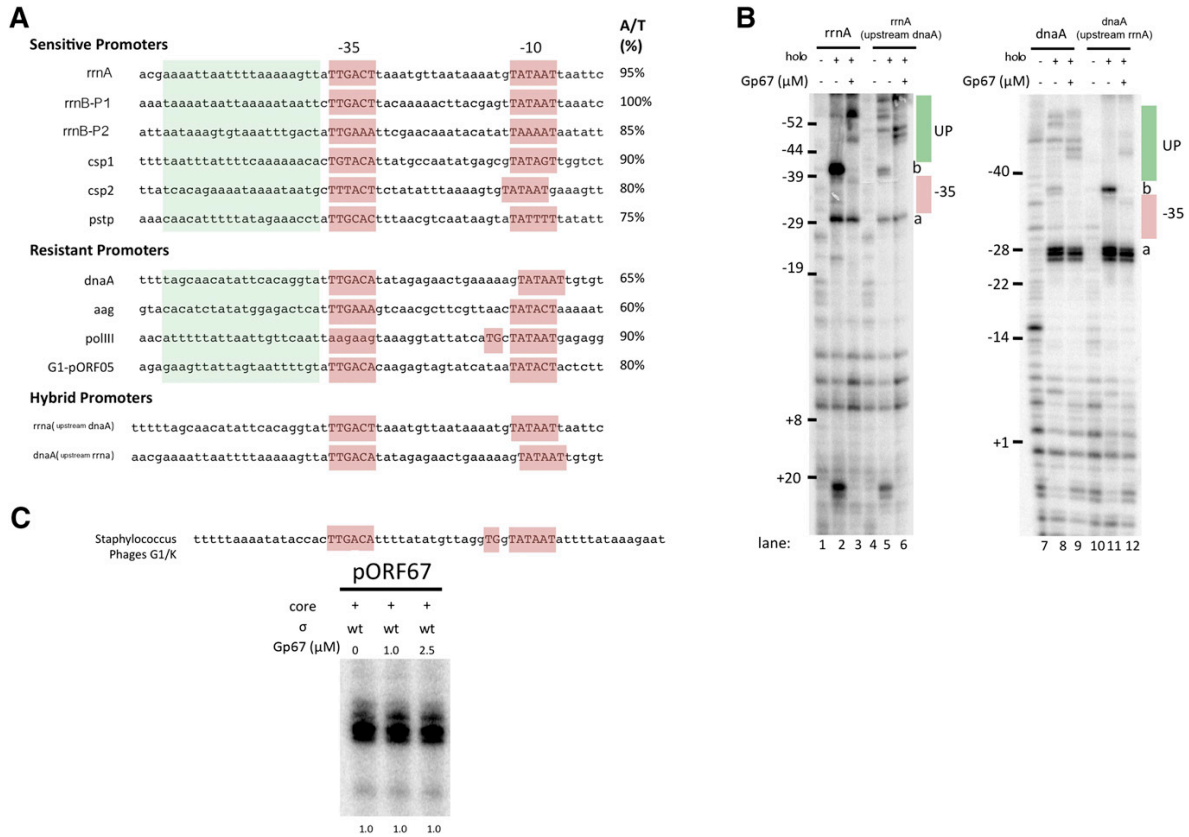
Figure S5. List of Promoters Used in this Study, Related to Figure 6

(A) *Sau* promoters that were identified and used to probe gp67 activity. The −10, extended −10, and −35 elements are capitalized and highlighted in red. The putative UP elements used to calculate the A/T content (shown to the right of the promoter sequence) are highlighted in green.

(B) DNase I footprinting of the hybrid promoters described in Figure 6.

(C) Gp67 does not inhibit transcription from its own promoter. The promoter upstream of the gene for gp67 was identified (upper panel) and shows robust activity in vitro (lower panel). However, gp67 is unable to inhibit RNAP activity at this promoter.