Bacterial Sigma Factors: A Historical, Structural, and Genomic Perspective

Andrey Feklístov,[1] Brian D. Sharon,[2,3,4] Seth A. Darst,[1] and Carol A. Gross[3,4,5]

[1]The Rockefeller University, New York, NY 10065; email: darst@rockefeller.edu

[2]Graduate Group in Biophysics, [3]Department of Microbiology and Immunology, [4]California Institute of Quantitative Biology, and [5]Department of Cell and Tissue Biology, University of California, San Francisco, California 94158; email: cgrossucsf@gmail.com

Annu. Rev. Microbiol. 2014. 68:357–76

First published online as a Review in Advance on June 18, 2014

The *Annual Review of Microbiology* is online at micro.annualreviews.org

This article's doi: 10.1146/annurev-micro-092412-155737

Keywords

sigma, RNA polymerase, transcription, gene expression

Abstract

Transcription initiation is the crucial focal point of gene expression in prokaryotes. The key players in this process, sigma factors (σs), associate with the catalytic core RNA polymerase to guide it through the essential steps of initiation: promoter recognition and opening, and synthesis of the first few nucleotides of the transcript. Here we recount the key advances in σ biology, from their discovery 45 years ago to the most recent progress in understanding their structure and function at the atomic level. Recent data provide important structural insights into the mechanisms whereby σs initiate promoter opening. We discuss both the housekeeping σs, which govern transcription of the majority of cellular genes, and the alternative σs, which direct RNA polymerase to specialized operons in response to environmental and physiological cues. The review concludes with a genome-scale view of the extracytoplasmic function σs, the most abundant group of alternative σs.

Contents

OVERVIEW

The discovery that bacterial RNA polymerase (RNAP) required a specialized subunit, sigma (σ, for selectivity), to recognize promoters ushered in a new era in regulatory biology of seeking to understand how the cell ensures fidelity of transcription initiation. This finding presaged the discovery that all multisubunit RNAPs require separate initiation functions. In retrospect, this division of tasks is not surprising. To perform the central cellular function of transcription, RNAP must efficiently traverse a large fraction of the genome without preference for DNA sequence during the elongation phase of transcription. Nevertheless, the beginning of transcription units must be precisely specified and their use tightly regulated. Cells have solved this problem by encoding transcription start sites with specific DNA sequences (promoters; **Figure 1*a***) and allowing the cellular multisubunit RNAP to locate these sequences through a dissociable, sequence-specific DNA-binding protein that functions primarily at initiation.

Whereas initiation functions are divided among many proteins in eukaryotes, the prokaryotic σ performs all initiation functions, making σ an ideal model for studying initiation. Bacteria have a single essential housekeeping σ that promotes transcription of thousands of genes, and many alternative σs, each promoting transcription of a specialized gene repertoire necessary for coping with stress or development. Given the critical mechanistic and regulatory roles of σs, it is not surprising that they have been intensely studied since their discovery. We begin this review with a series of historical vignettes recounting key discoveries about the σ^{70} family of proteins. The body of the review is dedicated to understanding the current status of important broad trends in σ biology: how σs perform their critical functions of directing RNAP to promoters and initiating DNA-strand opening; how σs further integrate regulatory events in transcription; and how alternative σs have modified the mechanism of initiation to enhance their specificity. We end by considering the breadth and reach of the most abundant, smallest, and most divergent

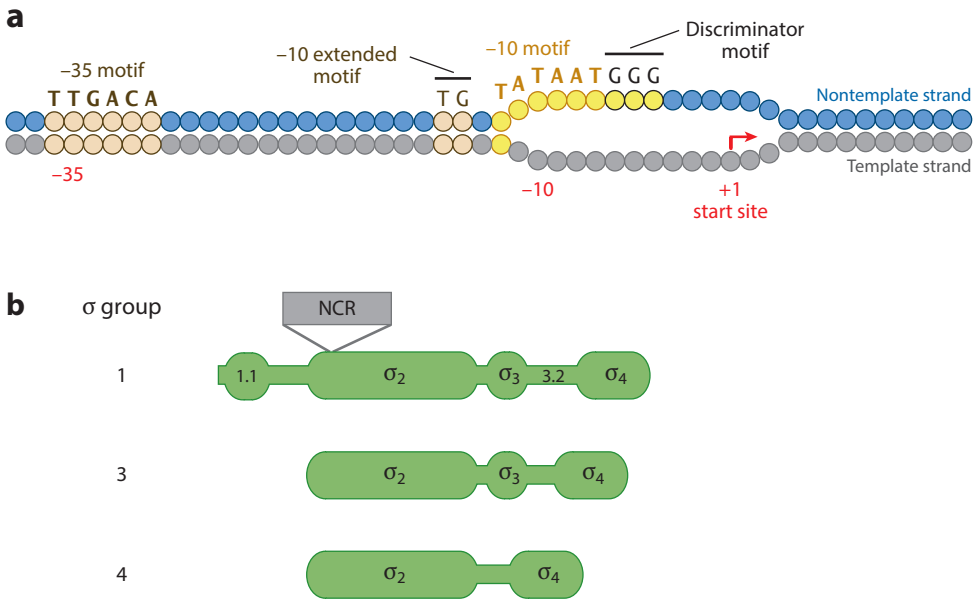


Figure 1

(*a*) Promoter motifs recognized by housekeeping σ s. Promoter DNA in RNA polymerase open complex is shown as circles. Blue circles represent nucleotides of the nontemplate DNA strand; light gray represents the template strand. Consensus sequences for each promoter motif are shown above. Nucleotide positions with respect to the transcription start site (+1, *red arrow*) are denoted below in red. (*b*) Domain architecture of σ s. Group 1 (housekeeping), 3, and 4 σs are represented by green bars. Domains and regions are labeled with their numerical names. Thicker regions denote structured domains. Abbreviation: NCR, nonconserved region.

alternative σs [called extracytoplasmic function (ECF) or Group 4 σs]. Overall, we highlight the mechanistic, genomic, and regulatory roles of the σ^{70} family of proteins.

KEY MOMENTS IN σ HISTORY

Discovery of σ^{70}

As a graduate student in the Watson-Gilbert group at Harvard University, Dick Burgess purified RNAP to homogeneity. There were few examples of multisubunit enzymes at the time, making it critical to show that all copurifying polypeptides were part of the enzyme itself. This led Dick to test additional purification steps, showing that purified RNAP retained the ability to initiate RNA chains on calf thymus DNA after passage through a phosphocellulose column. However, a fellow graduate student, Jeffrey Roberts, studying initiation on a bacteriophage λ template, found the enzyme to be inactive (R. Burgess and J. Roberts, personal communication). Reconstituting the RNAP with other fractions from the phosphocellulose chromatography identified σ^{70}, the *Escherichia coli* housekeeping σ. The σ worked during initiation, as it increased the number of chains (as judged by incorporation of ^{32}P-γ phosphate, present only at the 5′ end of the chain). Ekhardt Bautz and Joshua Dunn at Rutgers performed similar experiments, and the two groups published a single paper documenting these experiments (9); other groups also identified a dissociable subunit but did not identify its mode of action (45, 49). Additional experiments demonstrated that σ^{70} was catalytic: Once initiation occurred, σ^{70} could dissociate from one RNAP molecule and promote

initiation by a differentially marked (rifampicin-resistant) RNAP (45, 49, 107). Thus, RNAP has two distinct forms: holoenzyme ($\alpha_2\beta'\beta\omega+\sigma$), for initiation, and core ($\alpha_2\beta'\beta\omega$), for elongation.

The Bacterial Promoter Takes Shape

With σ in hand, the focus shifted to determining the DNA-binding regions that marked the promoter, a difficult feat in the absence of DNA sequencing. Undeterred, several groups initiated RNAP transcription from a strong promoter, isolated the short fragment protected from DNase digestion, and inferred the promoter sequence from that of complementary RNA using nearest-neighbor sequencing of oligos produced by digestion with various RNases. Based on the sequences of several strong promoters, Pribnow (74) and Schaller et al. (90) independently identified the -10 motif (**Figure 1*a***) (74, 90). However, RNAP could not rebind this minimal fragment, suggesting that there were missing recognition determinants. Digestion with restriction enzymes provided a longer fragment to sequence and identified the -35 motif as a second common promoter motif (**Figure 1*a***). The importance of the -10 and -35 motifs was confirmed by showing that previously isolated mutations that altered expression from the *lac* promoter had nucleotide changes in the -10 and -35 motifs of the promoter. The tour de force was the sequencing of the entire *lac* control region by Bill Reznikoff and colleagues, identifying binding sites for Lac repressor and cyclic-AMP binding protein, as well as mutations in the -35 motif of the promoter (23, 78).

σ Binds Promoter DNA

Although σ was required for initiation, σ^{70} alone was unable to recognize DNA (8, 117). This led to the suggestion that σ might alter RNAP conformation, thereby exposing latent DNA-binding determinants in polymerase. In fact, the opposite is true—latent DNA-binding domains of σ are exposed when σ binds to RNAP. The realization that alternative σs recruit RNAP to different -10 and -35 promoter sequences was difficult to reconcile with the idea that σ exposed latent DNA-binding domains in RNAP. It was extremely unlikely that RNAP carried the multiple DNA-binding determinants required for recognizing many alternative promoter sequences (57). Then, several groups found mutations in σ that either compensated for poor promoters or altered promoter preference (99, 112, 123). The finding that the N-terminus of the housekeeping σs was autoinhibitory for DNA binding began to address why σs might not bind DNA even though they carried the DNA-recognition determinants (24). Ultimately, structural data suggested that isolated σs likely have intramolecular interactions that occlude their DNA-binding determinants (101, 102). Binding to RNAP splays these domains apart and displays them in the correct orientation and spacing for promoter binding (67–69). Requiring σ to bind RNAP prior to DNA recognition ensures that the binding affinity of each σ for RNAP influences the relative abundance of different holoenzymes and thus the spectrum of promoters recognized.

The σ^{70} Family of Proteins

The initial report describing σ suggested that the cell might have additional σ-like positive regulators with different specificity, possibly explaining changes in transcription patterns, such as those observed late in the growth cycle of lytic phages (9). Indeed, new RNAP-binding proteins appeared during *Bacillus subtilis* sporulation (31, 32) (e.g., SpoIIG, now called SigE) or after infection by the lytic phages SPO1 (105) or T4 (43, 44, 114, 115), and these altered the in vitro transcription preferences of RNAP, suggesting a role for σs in development. Then, an *E. coli* σ-like positive regulator (HtpR, now called σ^{32}) that directed RNAP to heat shock gene promoters was

identified, indicating that σ-like proteins were general across bacteria and controlled both stress and developmental responses (30). A true understanding of the σ universe awaited the onset of DNA sequencing. First came the realization that the essential housekeeping σs of *E. coli* and *B. subtilis* were highly conserved and had strong sequence similarity to SigE and σ^{32}, hinting that they were a protein family (29, 34, 104). As additional σs were sequenced, this idea was confirmed, and alternative σs were subdivided into Group 2 σs (most similar to housekeeping Group 1 σs) and the more divergent Group 3 σs (55). The most divergent group of alternative σs, the ECF σs (Group 4), was not identified until 1994 (56). These small σs initially escaped detection via homology searching because they lacked one conserved σ domain (**Figure 1*b***). With the onset of large-scale sequencing of microbial genomes, thousands of σs have been identified, with ECF σs the most abundant. Except for *Mycoplasma*, all bacteria have multiple alternative σs.

HOW DOES σ PERFORM ITS FUNCTIONS?

The σs perform two main functions: (*a*) directing the catalytic core of the RNAP to appropriate transcription start sites, and (*b*) initiating strand separation of the double-helical DNA as the first step in the formation of the transcription bubble. Indeed, to be classified as a σ factor, a protein must perform both tasks. In addition to these central functions, a number of additional σ activities have been identified. A dozen years of elegant structural studies combined with incisive biochemical and biophysical approaches have improved our understanding of the structural basis of all of the functions of σ.

Directing the Catalytic Core of the RNAP to Transcription Start Sites

Prokaryotic promoters are marked by five motifs: the -35, extended -10, -10, and discriminator motifs recognized by σ; and the UP element recognized by the C-terminal domain (CTD) of the RNAP α-subunit. A great deal is known about how each of these motifs is recognized, and here we outline the σ recognition strategy. The σ proteins are composed of a variable number of structured domains connected by flexible linkers. The simplest σs have two domains (Group 4 or ECF σs: σ_2, σ_4), some have three domains (Group 3 σs: σ_2, σ_3, σ_4), and the housekeeping σs have four domains ($\sigma_{1.1}$, σ_2, σ_3, σ_4) (**Figure 1*b***). Except for $\sigma_{1.1}$, each domain has DNA-binding determinants: σ_4, -35 motif; σ_3, extended -10 motif; σ_2, -10 and discriminator motifs (**Figure 2**). The -35 motif is recognized by a canonical helix-turn-helix (HTH) DNA-binding domain within σ_4 and has been visualized with high-resolution X-ray crystal structures of -35 motif/σ_4 interactions for housekeeping (14) and ECF (51) σs (**Figure 2**). A long α-helix of σ_3 recognizes the DNA major groove of the extended -10 motif (5, 88). The -35 and extended -10 motifs are recognized as double-stranded DNA and remain double stranded throughout the initiation process. The -10 and discriminator motifs are recognized upon strand separation (**Figure 2**) as nontemplate-strand, single-stranded DNA (ssDNA, see below). Normally, one of the double-stranded motifs (-35 or extended -10) is required for initiation (although see 27), and these are thought to anchor the holoenzyme in the appropriate position and orientation to then engage the -10 motif upon strand separation (26, 28).

Although all specific recognition of promoter sequence is mediated by σ, σ does not specifically bind promoters on its own. At least two mechanisms (that are not mutually exclusive) contribute to suppressing the DNA-binding capacity of free σs.

Conformational restriction. The structure of a full-length, isolated σ has not been determined, but it is thought that the σ domains interact with each other to maintain a relatively compact conformation that buries the DNA-binding determinants, or at least does not properly present

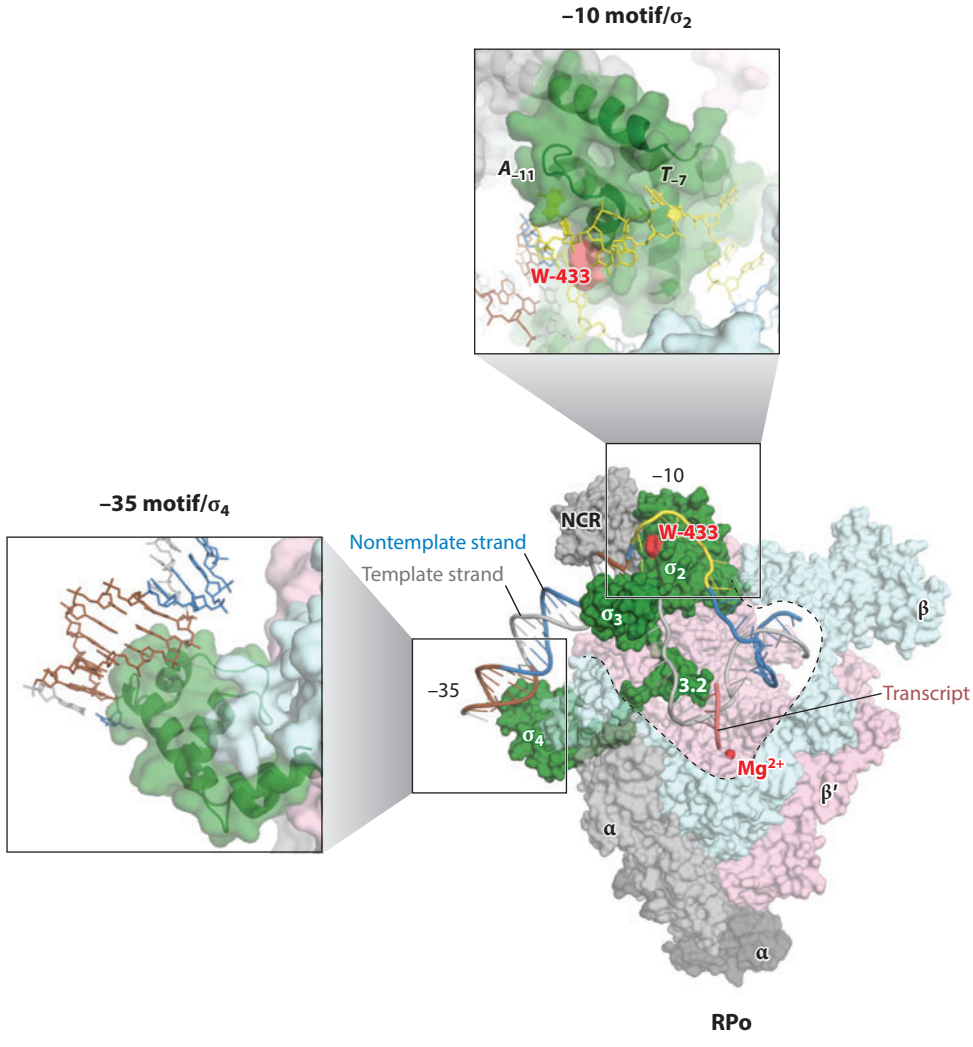


Figure 2

The lower-right image shows an *Escherichia coli* RNA polymerase (RNAP) open complex (RPo) model (2). The molecular surface of RNAP holoenzyme is shown, color-coded as follows: gray, α-subunits; light cyan, β; light pink, β′; green, σ [but with the nonconserved region (NCR) shown in gray]. A segment of the β-subunit is cut away to expose the active site channel. The DNA, shown as a phosphate-backbone worm, is color-coded: The nontemplate DNA strand is blue, the template strand is red, the -35 motif is brown, and the -10 motif is yellow. A nascent 5mer RNA transcript is shown in pink. The catalytic Mg^{2+} ion is shown as a red sphere. The boxed regions are magnified. The close-up view on the left shows the -35 motif/σ_4 interaction. The -10 motif/σ_2 interaction is shown in detail on the top.

them to interact with the widely spaced promoter motifs (**Figure 2**). This idea is supported by three key pieces of information: (*a*) Structural studies of a Group 3 σ bound to its inhibitory anti-σ factor reveal a compact σ conformation with buried DNA-binding determinants. Cross-linking results suggest that the isolated σ assumes a similar conformation as that bound to the anti-σ (101, 102). (*b*) FRET/LRET analyses show that the σ domains are relatively compact in free σ and move to a spread-out conformation upon formation of holoenzyme (10, 11). (*c*) Structural results show

that the core RNAP is responsible for positioning the σ domains at the appropriate orientation and spacing to interact with the promoter motifs (**Figure 2**) (69).

Autoinhibitory restriction. The N-terminal $\sigma_{1.1}$, found only in housekeeping σs (**Figure 1*b***), suppresses recognition of promoter motifs by the isolated σ. Specific promoter recognition by σ alone is weak and difficult to detect even in the absence of $\sigma_{1.1}$; the presence of $\sigma_{1.1}$ suppresses that activity even further (12, 24, 91). We note that autoinhibition and conformational restriction may be linked; one way $\sigma_{1.1}$ might inhibit σ promoter-binding activity is by stabilizing a compact conformation of the σ that would be recalcitrant to promoter recognition.

Strand Separation of the DNA: The First Step in Formation of the Transcription Bubble

Once the RNAP holoenzyme locates promoters through recognition of double-stranded, upstream DNA motifs, RNAP unwinds 13 base pairs of the duplex DNA to form the transcription bubble. Remarkably, the bacterial enzyme, containing either a housekeeping or alternative σ, accomplishes this feat without the assistance of ATP hydrolysis. Current thinking suggests that σ_2-mediated capture of a flipped base (or bases) of the nontemplate-strand -10 motif underlies the strand separation process of all σs.

Housekeeping *σ*s. Helmann & Chamberlin (34) noted a number of invariant basic and aromatic residues in σ_2 and proposed, by analogy with ssDNA- and RNA-binding proteins, that this region binds one of the DNA strands in the nascent bubble to stabilize the strand-separated state. Later studies showed that these invariant residues of σ are indeed critical for transcription-bubble formation (21, 40, 99, 112, 123), functioning as a sequence-specific, ssDNA-binding determinant specific for the nontemplate strand sequence of the -10 motif ($T_{-12}A_{-11}T_{-10}A_{-9}A_{-8}T_{-7}$) (27, 58, 84, 89, 94, 95, 119–121).

The atomic details of the nontemplate-strand -10 motif/σ_2 recognition (**Figure 2**) were revealed by high-resolution (2.1-Å) X-ray crystal structures (28). Extensive interactions occur between σ_2 and the phosphate backbone of every -10 motif nucleotide, but only two bases are recognized sequence specifically: A_{-11} and T_{-7} are flipped out of the ssDNA base-stack and buried in complementary protein pockets in σ_2 (**Figure 2**). This explains the remarkably high conservation of these two -10 motif positions and the extreme sensitivity of promoter activity to mutations away from the consensus A_{-11} (52, 54) and, to a slightly lesser degree, T_{-7} (65).

Many studies (although not all, see 86) suggest that transcription-bubble formation initiates at A_{-11} and then propagates downstream to the transcription start site (+1; 16, 36, 54), further highlighting the importance of this base in the -10 motif. Initiation of bubble formation in duplex DNA is more difficult than subsequent propagation of the bubble (1), but stacking interactions dictate that the A nucleotide of an adjacent TA step is the easiest base to flip (75). This may contribute to the relatively high conservation of T_{-12} (to provide a $T_{-12}A_{-11}$ step to facilitate flipping of the all-important A_{-11}). All of the invariant σ_2 aromatic residues important for transcription-bubble formation (F-419, Y-430, W-433, W-434 in *E. coli* σ^{70}) are dedicated to forming the A_{-11} pocket and/or stabilizing the double-strand/single-strand (ds/ss) junction formed by the flipping of A_{-11} (28). The prominent position of W-433 at the ds/ss junction precludes binding to an intact B-form DNA helix. This residue may be used as a wedge to disrupt the -11 base pair and fill the space vacated by the flipped A_{-11}, suggesting a functional analogy with other DNA-binding proteins that recognize flipped-out bases using bulky side chains for helix invasion (53, 118). Binding of the nontemplate-strand -10 motif to σ_2 introduces a 90° bend in the DNA backbone that directs

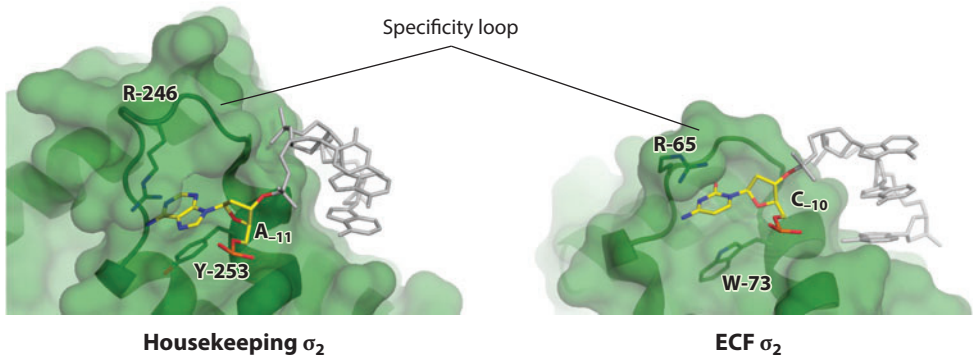


Figure 3

Nucleation of promoter melting by housekeeping (*left*) and extracytoplasmic function (ECF) (*right*) σs. In both cases, promoter melting is triggered by the capture of a single base in a complementary pocket of σ_2. The pocket is formed by a specificity loop and additional invariant residues. (*left*) A_{-11} is stacked between R-246 and Y-253 of *Thermus aquaticus* $\sigma^A{}_2$ [Protein Data Bank (PDB) ID 3UGO; 28]. (*right*) C_{-10} is stacked between W-73 and R-65 of *Escherichia coli* σ^E (PDB ID 4LUP; 13).

the downstream DNA toward the RNAP active site cleft, a prerequisite for subsequent steps of promoter opening and loading of the template strand into the active site (26, 28).

The prevailing view was that the -10 motif must be recognized sequence specifically first as duplex DNA by conserved region 2.4 in σ_2, and that the nontemplate-strand bases of the -10 motif are recognized after strand separation by the critical melting residues mostly in conserved region 2.3 of σ_2 (22, 35, 37). However, structural modeling combined with biochemical results suggests that specific recognition of the -10 motif in duplex form is unlikely (28). Thus, -10 motif sequence recognition and initiation of strand separation are coupled. The molecular details of -10 motif/σ_2 interaction were fully corroborated by a subsequent structure in the context of the entire RNAP holoenzyme and a partial promoter construct (122). This study also delineated contacts between exposed bases of the discriminator (just downstream of the -10 motif) and σ_2 that stabilize the RNAP open complex (RPo; **Figure 1*c***, 27, 33).

Alternative σs. The biochemical, biophysical, and structural details of the promoter-opening process for alternative σs have not been studied to nearly the same extent as those for housekeeping σs, but available data suggest that the overall process is similar (17, 18, 60). Recent work structurally characterizing an ECF σ (*E. coli* σ^E) in complex with its cognate nontemplate-strand -10 motif ($T_{-13}G_{-12}T_{-11}C_{-10}A_{-9}A_{-8}A_{-7}$; 13) provides important insights into promoter melting by this highly divergent family. Like housekeeping σs, a nearly absolutely conserved base (analogous to housekeeping σ -10 motif A_{-11}) is flipped out of the single-strand base stack and buried in the σ_2 pocket corresponding to the A_{-11} pocket in housekeeping σ_2 (**Figure 3**). However, in the case of σ^E, the flipped-out base is a C (C_{-10}). There is no structural equivalent of the T_{-7} pocket of housekeeping σs; only one base is flipped and specifically captured in a protein pocket. Remarkably, the recognition of the identity of this all-important flipped base in the -10 motif is governed by a modular specificity loop (**Figure 3**). For instance, the -10 motif consensus for promoters recognized by *B. subtilis* ECF σ^W suggests that A is the flipped base. Swapping the σ^W specificity-loop into σ^E gave rise to a hybrid σ that recognizes A as the flipped base rather than C (13). This modularity allows for the rapid evolutionary diversification of ECF σs and their cognate -10 motifs.

Suppressing Nonspecific Transcription Initiation

In addition to having the critical function of imparting sequence-specific DNA-binding capacity to the RNAP holoenzyme, σ has been reported to suppress the capacity of the core RNAP to initiate transcription at nonspecific (nonpromoter) DNA sites (97). The large active site channel for binding DNA, located between the pincers of the crab-claw-like structure of core RNAP, is basic (positively charged; 19) and is expected to promote nonspecific interactions between nucleic acids and core RNAP, increasing the chances for nonspecific initiation. In addition to suppressing promoter recognition by free σ, the N-terminal $\sigma_{1.1}$ domain of housekeeping σs may also suppress nonspecific DNA binding in the RNAP active site cleft. In the RNAP holoenzyme, the highly acidic $\sigma_{1.1}$ situates itself in the RNAP active site cleft (2, 61; **Figure 4*a***). Remarkably, upon formation of RPo, loading of promoter DNA into the RNAP active site cleft displaces $\sigma_{1.1}$ out of the channel (61) (**Figure 4*a***). A long, unstructured linker connects $\sigma_{1.1}$ to the rest of σ (this linker is 37 residues in *E. coli* σ^{70}), facilitating the large movements of $\sigma_{1.1}$ (2). The highly dynamic nature of $\sigma_{1.1}$ in the promoter opening process explains why mutations and truncations of $\sigma_{1.1}$ can have significant effects on the kinetics of RPo formation (110, 116). These findings suggest that $\sigma_{1.1}$ acts as a gatekeeper for the RNAP active site, sitting in the active site cleft and blocking access to the active site for random DNA sequences. Displacement of $\sigma_{1.1}$ requires promoter sequences that interact specifically with σ to initiate the proper pathway for RPo formation. The mechanism of displacement is unknown. Only housekeeping σs harbor $\sigma_{1.1}$ (**Figure 1*b***): Suppression of nonspecific DNA binding may be especially important for housekeeping σs given their increased DNA melting capacity (see below).

Interaction with Activators

Activation of basal transcription is a major regulatory strategy in bacteria. Most bacterial transcription activators bind to DNA operators upstream of the promoter -35 motif and activate transcription through direct contacts with a target subunit of the RNAP, creating new protein-protein and protein-DNA contacts to help recruit RNAP to promoters (7). When the activator binds very near the upstream part of the -35 motif, σ_4 is frequently its interaction target (100, 106; reviewed in 25).

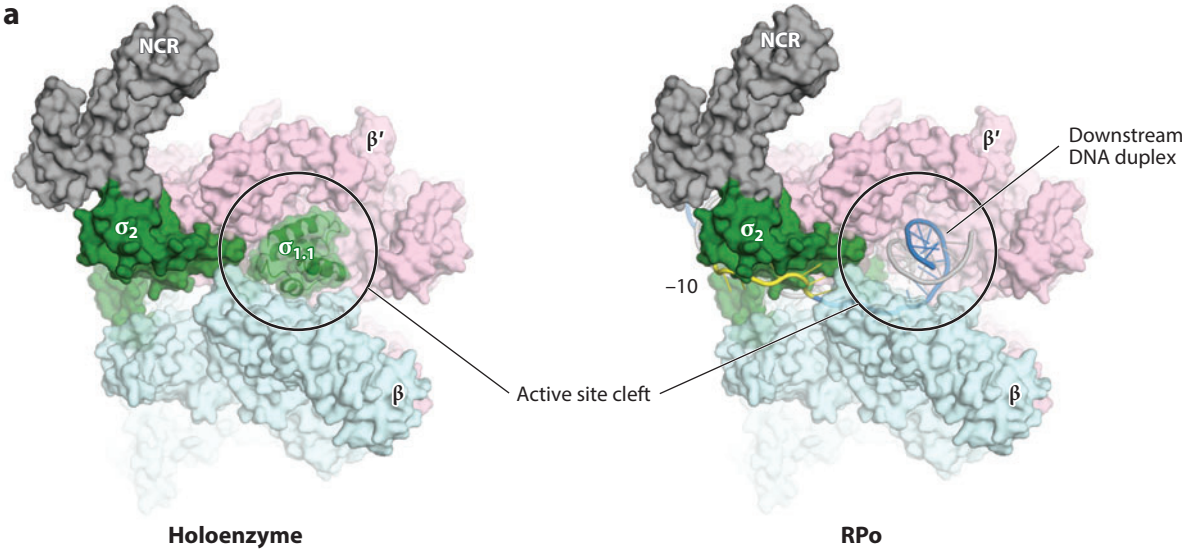
Mutations in σ_4 that selectively disrupt the function of an activator without affecting basal transcription (positive control mutants) cluster on the upstream face of σ_4 and are well positioned to interact with the downstream subunit of the dimeric activators (14, 39). The interaction of σ with some transcription regulators is likely to be more complex. For instance, GrgA in *Chlamydia trachomatis* interacts with the housekeeping σ through a nonconserved insertion in the σ between conserved regions 1.2 and 2.1 (3).

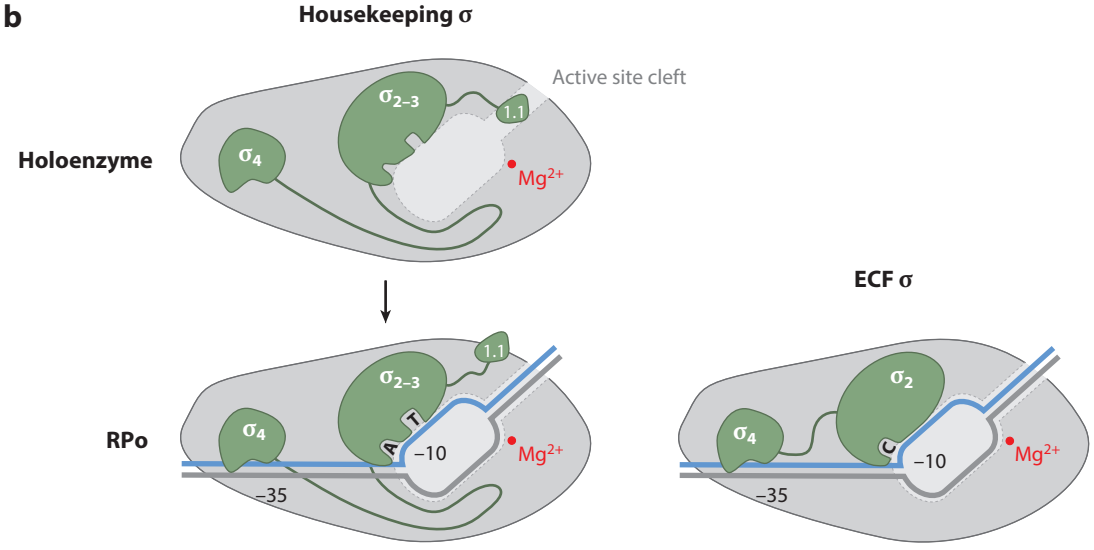
Abortive Initiation/Promoter Clearance

In RNAP holoenzyme, the σ_3 and σ_4 domains are linked by an approximately 50-residue segment that includes a loop-like feature comprising σ conserved region 3.2 (53; **Figure 1*b***). This segment winds through the RNAP active site channel and occupies the exit path for the nascent RNA (69, 109). Within this linker, a hairpin-like loop (dubbed the σ finger; 122) protrudes toward the RNAP active center (**Figure 2**). The σ finger plays an important role in initiation by helping the RNAP bind the initiating substrate. A 7-residue deletion of the σ finger caused a two-orders-of-magnitude increase of the K_m for the 3′-initiating NTP (50). Although this region of σ cross-links to initiating nucleotide analogs (50, 93), crystal structures of initiating complexes suggest that the effect of the

σ finger on initiating NTP binding is indirect through the preorganization of the DNA template strand (122). Notably, the 3.2 linker is much shorter in alternative σs (15) (**Figure 4*b***), pointing to a simplified initiation mechanism in this group of σs with reduced melting capacity.

Once RNAP catalyzes the first phosphodiester bond, the continued translocation and catalytic activity of the RNAP extends the 3′ end of the nascent transcript. Analysis of RNAP holoenzyme crystal structures indicates that a transcript of only four to five nucleotides in length will encounter the σ finger in its path (69; **Figure 2**). At each step, the elongating RNA chain must either displace σ out of its path or dissociate from the complex in a process termed abortive initiation. The structure-based proposal that σ plays a major role in abortive initiation is supported by extensive functional evidence:





1. Substitution of two conserved residues near the beginning of the σ_3-σ_4 linker (P-504 and S-506 in *E. coli* σ^{70}) led to a dramatic reduction in the ratio of abortive to full-length transcripts (92).
2. Deletion of a seven-residue segment at the tip of the σ finger (the part of the σ finger encountered first by the elongating transcript) led to a relative decrease in the ratio of short abortive transcripts (≤10 nucleotides) to longer ones, as longer transcripts are now required for an encounter with σ (50).
3. Complete deletion of the σ_3-σ_4 linker led to the nearly complete elimination of abortive transcripts (69).

Structural analysis indicates that the nascent RNA transcript emerges from the RNA exit channel when it reaches a length of ~16 nucleotides. This corresponds to the RNA length that results in displacement of both the σ_3-σ_4 linker and σ_4 from their normal positions on RNAP holoenzyme (70). Abortive transcripts beyond the length of 16 nucleotides are only rarely observed (38), supporting a model in which the transition from the initiation to the elongation phase of transcription involves a direct competition between the elongating RNA transcript and the σ_3-σ_4 linker and σ_4 for the same binding site within core RNAP. The competition gives rise to abortive initiation and also sets up the staged displacement of segments of σ from the RNAP, facilitating promoter escape (67, 69).

Promoter Proximal Pausing During Early Elongation

Even after displacement of the σ_3-σ_4 linker and σ_4 and promoter escape, σ can still be retained by core RNAP during the beginning of elongation through interactions between σ_2 and the β′ clamp helices. During elongation, σ undergoes stochastic release from RNAP, sometimes remaining dozens of base pairs downstream of the transcription start site (4, 6, 41, 66, 71, 83, 96). While bound to elongating RNAP, σ continues to scan the DNA for promoter-like elements and is able to recognize both -10- and -35-like sequences, thereby inducing transcriptional pausing (73, 83). The σ^{70}-dependent promoter-proximal pause observed at λ phage late gene promoters is prerequisite for the recruitment of the phage-encoded protein Q that converts RNAP into a terminator-resistant form (85). A number of *E. coli* promoters have been shown to have σ^{70}-dependent pauses, but their physiological significance is still unclear (6, 71, 73). Chromatin immunoprecipitation (ChIP) assays support stochastic release of σ from elongating RNAP in vivo (63, 64, 76, 77, 111).

A MODIFIED INITIATION STRATEGY ALLOWS THE DIVERGENT ALTERNATIVE σS TO PERFORM THEIR FUNCTIONS

The role of the essential housekeeping σ is to promote initiation at the vast bulk of bacterial promoters, regardless of their strength. Indeed, this hardwired promoter hierarchy is an essential level

Figure 4

(*a*) Views into the active site cleft of *Escherichia coli* RNAP holoenzyme (*left*; Protein Data Bank ID 4LK1; 2) and an RNAP open complex (RPo) model (*right*). In holoenzyme, $\sigma_{1.1}$ sits in the active site cleft but is displaced by the downstream DNA duplex in RPo. (*b*) Schematic views of RNAP holoenzymes. A version of this figure was first published in Reference 20 by Nature Publishing Group. The core RNAP is shown (*gray*) with active site channel (*light gray*). The disposition of σ domains (*green*) is also illustrated. The housekeeping σ (*left column*) triggers promoter melting by capturing two highly conserved bases of the -10 element, A_{-11} and T_{-7}, in RPo (*bottom row*). Nonspecific interactions of nucleic acids with the active site are prevented by $\sigma_{1.1}$ in the downstream duplex channel of holoenzyme (*top row*). Entry of promoter DNA in RPo displaces $\sigma_{1.1}$. Alternative σs such as *E. coli* σ^E (*right column*) lack $\sigma_{1.1}$ and employ a simplified strategy relying on a single pocket for capturing C_{-10}. Abbreviations: ECF, extracytoplasmic function; NCR, nonconserved region; RNAP, RNA polymerase.

of control built into the cell during evolution. The ability to recognize and initiate at promoters very divergent from the consensus promoter also enables the cell to respond to activators, which boost expression from such promoters, facilitating a dynamic response to changing conditions. The housekeeping σs can perform these roles because they are able to initiate at promoters having only a subset of recognition motifs (reviewed in 37). Of the five promoter sequence motifs, two suffice for transcription, as housekeeping σs initiate either at promoters with -10 and -35 motifs (with only loose spacing requirements) or at those with -10 and extended -10 motifs (**Figure 1*a***). The fourth motif, the discriminator, is modulatory. Furthermore, strong UP elements, recognized by the CTD of the α-subunit of RNAP (α-CTD), can substitute for the -35 promoter motif (62). What features of housekeeping σs enable recognition of such diverse promoters?

All housekeeping σs harbor two protein pockets that specifically capture two flipped-out bases of the nontemplate strand, complemented by the contacts with exposed bases of the discriminator (**Figure 4*b***). These structural features give housekeeping σs a powerful DNA-melting capacity, allowing them to function in the face of highly nonoptimal promoter sequences. As a consequence, as long as the -10 region is intact, housekeeping σs can use diverse upstream regions to anchor RNAP (UP element, -35, extended -10) and still accomplish promoter melting.

In stark contrast to the housekeeping σs with their generalized role, the divergent alternative σs [Group 3 and ECF (Group 4) σs] mount concerted, focused responses to specific environmental conditions. Two important characteristics of such responses are (*a*) their regulons are small, ranging from as few as two to three genes to as many as several hundred, thereby minimizing the metabolic cost of the response; and (*b*) the relative expression of regulon genes depends almost exclusively on a hardwired promoter hierarchy. How do these two σ groups achieve these critical goals? Based on the limited exploration performed thus far, it appears that a weakened melting capacity of these σs relative to the housekeeping σs plays a significant role in their increased specificity.

The Group 3 σs have the same three DNA-recognition domains (σ_2-σ_4) as the housekeeping σs (46, 47) but do not use the mix-and-match promoter strategy of the housekeeping σs. Instead, the two Group 3 σs studied thus far, *E. coli* σ^{32} (heat-shock σ) and σ^{28} (flagella σ), require all three recognition motifs (-35, extended -10, -10) as well as rigid spacing between them for promoter activity. As evidence, σ^{32} and σ^{28} promoters have a higher information content (18.3 and 21.3 bits, respectively) (72, 82) than σ^{70} promoters (6.5 bits) (98). These requirements sharply restrict the number of promoters per genome. Essentially all Group 3 σs lack one or more of the four critical melting residues, which form the A_{-11} (or equivalent) pocket in the housekeeping σs. We suggest that only promoters with a near-perfect match to the consensus can be melted. Consistent with this hypothesis, converting the missing melting residues in the Group 3 σs to the housekeeping consensus relaxed requirements for highly conserved sequence in the promoter-recognition motifs and for highly conserved spacing. Importantly, the melting-consensus alternative σs have improved ability to melt DNA, as measured by their ability to form a stable complex at 4C in vitro on a fork-junction template, an excellent mimic of open-complex formation. As the two Group 3 σs investigated represent very divergent branches of the family, reduced melting proficiency is likely to play a role in restricting the number of sites recognized by most, if not all, Group 3 σs (48).

The ECF σs are two-domain proteins (**Figure 1*b***) that recognize only the -10 and -35 motifs of their promoters, immediately limiting the number of potential promoters in the genome. Extensive studies indicate that both promoter motifs separated by rigid spacing are essential for promoter recognition by σ^E, an *E. coli* ECF σ (79). Moreover, unlike the housekeeping σs, recognition of the UP element by the α-subunit of RNAP cannot substitute for the -35 motif, although UP

elements do enhance expression from weak promoters (80). This observation has been validated for many ECF σs (81). Interestingly, at the time that this σ group was identified, it was noted that the regions of σ implicated in initiating melting and in recognizing the -10 motif (σ regions 2.3–2.4) were highly divergent from housekeeping σs. As we pointed out earlier, the ECF σ group diverges from the housekeeping paradigm by harboring only a single protein pocket to capture a flipped base of the -10 motif (rather than the two pockets of the housekeeping σs; **Figures 3** and **4*b***; 13, 28). Use of only a single flipped-out base, recognized by a single protein pocket formed by reordering a protein loop in the σ, is also likely to contribute to a weakened melting capacity.

In summary, weakened ability to initiate strand opening underlies the requirement for consensus promoter sequences in both Group 3 and ECF σs. The requirement for precise σ alignment on promoters may also contribute to the fact that these σs have very few activators.

A GENOME-SCALE VIEW OF THE ECF σS

ECF σs are the most abundant and diverse group of σs, but until recently the extent of this family was unclear. Thorsten Mascher and colleagues undertook the first comprehensive classification of ECF σs. They manually curated 2,700 ECF σs from hundreds of bacterial genomes and developed models to identify 43 phylogenetically distinct subgroups of ECF σs (103). These models were additionally used to develop a web-based ECF σ finder (**http://ecf.g2l.bio.uni-goettingen.de:8080/ECFfinder/**), enabling the inclusion of ECF σs in the MiST database (108) (**http://mistdb.com/**), so that as new genomes are sequenced, their ECF σs are added to the database. Genomes average 6 ECF σs; *Plesiocystis pacifica* SIR-1 holds the current record, encoding 118 ECF σs in its genome (**http://mistdb.com/**). This large number of ECF σs per genome raises the question of how transcriptional space is partitioned among ECF σs that co-occur in bacteria. Do these σs cross-talk with each other, sharing some fraction of their promoters? Or are they orthogonal, unable to recognize the promoters of co-occurring ECF σs? The co-occurrence of orthogonal ECF σs would enable insulated expression of their respective regulons, whereas the co-occurrence of cross-talking ECF σs would potentially enable σs to cooperate in coordinated responses.

To answer this question, it is necessary to identify promoters for the phylogenetically distinct ECF σ subgroups and then determine the cross-recognition capacity of these σs. Promoters were initially identified in silico by exploiting the fact that most ECF σs transcribe their own promoters. ECF σs in the database were binned according to subgroup, and the DNA sequences upstream of the σ genes in each subgroup were searched for overrepresented sequences to identify promoter motifs (81, 103). Currently, computational efforts have identified promoter models for 29 of 43 subgroups. Rhodius et al. (81) used these data to make in silico predictions of orthogonality; i.e., as to whether any two promoter models were sufficiently similar that it was likely that a single ECF σ could initiate from both promoter types. They found that most, but not all, promoters were orthogonal, suggesting that there is not much cross talk between ECF σs in different subgroups. Orthogonality depended on both the -35 and -10 motifs, as significantly more cross talk was predicted when a single motif was used to determine orthogonality. Importantly, Rhodius et al. also used high throughput methodologies to determine whether the predicted promoters were functional in vivo and whether their cognate σs were able to recognize promoters from other subgroups, largely validating the in silico predictions.

This large-scale experimental examination of ECF σ specificity also uncovered underlying trends for genome organization and the partitioning of transcriptional space (**Table 1**). Analysis

Table 1 Number of orthogonal and cross-talking ECF groups and σs in representative organisms

Organism	Orthogonal		Cross-talking		Unclassified	
	No. of groups	No. of σs	No. of groups	No. of σs	No. of groups	No. of σs
Bacillus subtilis	0	0	3	3	1	1
Caulobacter crescentus	6	9	1	2	1	1
Rhizobium leguminosarum	5	9	1	1	0	0
Streptomyces coelicolor	3	15	3	4	6	11

of ECF σ co-occurrence in genomes found that non-cross-talking σs tend to co-occur with more ECF σ subgroups than cross-talking σs do. For example, 6 of the 8 ECF σ subgroups found in *Caulobacter crescentus* and 5 out of 6 in *Rhizobium leguminosarum* are orthogonal subgroups. These findings suggest that insulation of transcriptional signaling pathways is typically beneficial for cells. Indeed, the most highly orthogonal ECF σs, subgroups 41 (113) and 42, not only are broadly distributed across bacteria but also often have multiple representative σs in a single genome. This could allow cells to develop complex cooperative responses within an overarching pathway that still remains insulated from the rest of the cell's transcriptional programs. *Streptomyces coelicolor* exhibits an impressive example of this strategy, with nearly half of its 30 ECF σs belonging to groups 41 and 42. In striking opposition to the broad distribution of highly orthogonal σs, the most promiscuous ECFs appear to have more restricted distributions. ECF σ subgroup 25 is a striking example, as it activates promoters from a wide range of other subgroups but co-occurs with very few of these and is only found in a small number of genomes overall.

Despite the apparent preference for insulated transcriptional signaling, cells can also evolve complex regulatory networks by utilizing cross-talking σs. The ECF σs in *B. subtilis*, which all respond to envelope stress, are one example. Although each is a member of a different ECF subgroup, all exhibit cross talk and thus have partially overlapping regulons. *B. subtilis* can therefore orchestrate a complex response in which a subset of core genes are induced by all envelope stresses, whereas additional genes are only induced by specific types of envelope stress (reviewed in 59). Experimental determinations of promoter specificity also revealed another role for cross talk. There are several examples of ECF σs that do not transcribe their own promoters. However, these ECF σs and their promoters co-occur with an ECF σ from a different subgroup that transcribes the promoter of the first ECF σ, raising the possibility of σ-σ cascades as an additional regulatory scheme (81). Such cascades can be found in a variety of divergent organisms, including *Flavobacterium* spp., *Mycobacterium* spp., and *Streptomyces* spp., and could control complex cellular decisions similar to the cascade of alternative σs controlling sporulation in *B. subtilis*.

Although these studies have begun to address general properties of the ECF σs, interesting questions remain unanswered. For highly orthogonal σs, it will be important to determine whether specific regulon functions are retained across organisms. Likewise, for cross-talking σs it will be interesting to address specific roles of σ-σ cascades, as well as whether cross-talking σs generally co-occur as part of cooperative responses or are in fact insulated by additional genome-level mechanisms.

PERSPECTIVE

In the 45 years since the discovery of σs, tremendous progress has been made in understanding their role in the cell, and consequently in understanding the process of transcription initiation.

The two fundamental properties of σs, identifying promoters and initiating strand separation of the DNA, are central to prokaryotic physiology. On the one hand, these properties enable housekeeping σs to carefully fine-tune transcription initiation at the majority of promoters, where they not only identify promoter sequence but also orchestrate several crucial checkpoints along the pathway of open-complex formation and monitor the abortive initiation process, promoter escape, and in some cases early elongation. On the other hand, the proliferation of alternative σs, with slightly divergent properties, has enabled the creation of critical stress and developmental responses; many of these responses have been identified and extensively studied.

What are the next big questions? We see two major areas that will be fleshed out in the coming years. First, increasingly sophisticated technologies for structural studies will enable us to follow the process of initiation through all of its complex steps at molecular resolution. Currently, we have single snapshots of parts of the process that are used to infer a causal set of events. However, in the next several years, we will have multiple structural snapshots complemented by results of state-of-the-art solution methods (42), which will allow us to follow the path of events, including understanding the complex dynamic rearrangements of σ during initiation. The mechanistic details of σ association with core RNAP and dissociation from transcription complexes entering the elongation phase, as well as the nature of transient intermediates of the promoter-opening pathway (87), are among the most important questions for future research.

Second, there will be extensive genome-scale and evolutionary investigations of the alternative σs. Genome-scale studies will address how transcriptional space is partitioned among the multitude of alternative σs encoded in most bacterial genomes, a study that is especially important for genomes with multiple ECF σs. The evolutionary studies will address the mechanisms for maintaining orthogonality/cross-talking relationships among σs and explore whether the trajectory of σ regulon evolution is sensitive either to the orthogonality of the parental σ or to the σs that commonly co-occur with the parental σ.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Brian Bae for sharing unpublished information. S.A.D. is supported by the National Institutes of Health. C.A.G. is supported by the National Institutes of Health. B.D.S is supported by National Institutes of Health training grant T32GM008284.

LITERATURE CITED

1. Altan-Bonnet G, Libchaber A, Krichevsky O. 2003. Bubble dynamics in double-stranded DNA. *Phys. Rev. Lett.* 90(13):138101
2. Bae B, Davis E, Brown D, Campbell EA, Wigneshweraraj SR, Darst SA. 2013. Phage T7 Gp2 inhibition of *Escherichia coli* RNA polymerase involves misappropriation of σ^{70} domain 1.1. *Proc. Natl. Acad. Sci. USA* 110(49):19772–77
3. Bao X, Nickels BE, Fan H. 2012. *Chlamydia trachomatis* protein GrgA activates transcription by contacting the nonconserved region of σ^{66}. *Proc. Natl. Acad. Sci. USA* 109(42):16870–75
4. Bar-Nahum G, Nudler E. 2001. Isolation and characterization of σ^{70}-retaining transcription elongation complexes from *Escherichia coli*. *Cell* 106(4):443–51

5. Barne KA, Bown JA, Busby SJW, Minchin SD. 1997. Region 2.5 of the *Escherichia coli* RNA polymerase σ^{70} subunit is responsible for the recognition of the "extended-10" motif at promoters. *EMBO J.* 16(13):4034–40
6. Brodolin K, Zenkin N, Mustaev A, Mamaeva D, Heumann H. 2004. The σ^{70} subunit of RNA polymerase induces *lac*UV5 promoter-proximal pausing of transcription. *Nat. Struct. Mol. Biol.* 11(6):551–57
7. Browning DF, Busby SJW. 2004. The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol.* 2(1):57–65
8. Burgess RR, Travers AA. 1970. *Escherichia coli* RNA polymerase: purification, subunit structure, and factor requirements. *Fed. Proc.* 29(3):1164–69
9. Burgess RR, Travers AA, Dunn JJ, Bautz EK. 1969. Factor stimulating transcription by RNA polymerase. *Nature* 221(5175):43–46
10. Callaci S, Heyduk E, Heyduk T. 1998. Conformational changes of *Escherichia coli* RNA polymerase σ^{70} factor induced by binding to the core enzyme. *J. Biol. Chem.* 273(49):32995–3001
11. Callaci S, Heyduk E, Heyduk T. 1999. Core RNA polymerase from *E. coli* induces a major change in the domain arrangement of the σ^{70} subunit. *Mol. Cell* 3(2):229–38
12. Camarero JA, Shekhtman A, Campbell EA, Chlenov M, Gruber TM, et al. 2002. Autoregulation of a bacterial σ factor explored by using segmental isotopic labeling and NMR. *Proc. Natl. Acad. Sci. USA* 99(13):8536–41
13. Campagne S, Marsh ME, Capitani G, Vorholt JA, Allain FH. 2014. Structural basis for -10 promoter element melting by environmentally induced sigma factors. *Nat. Struct. Mol. Biol.* 21(3):269–76. doi: 10.1038/nsmb.2777
14. Campbell EA, Muzzin O, Chlenov M, Sun JL, Olson CA, et al. 2002. Structure of the bacterial RNA polymerase promoter specificity σ subunit. *Mol. Cell* 9(3):527–39
15. Campbell EA, Tupy JL, Gruber TM, Wang S, Sharp MM, et al. 2003. Crystal structure of *Escherichia coli* σ^{E} with the cytoplasmic domain of its anti-σ RseA. *Mol. Cell* 11(4):1067–78
16. Chen J, Darst SA, Thirumalai D. 2010. Promoter melting triggered by bacterial RNA polymerase occurs in three steps. *Proc. Natl. Acad. Sci. USA* 107(28):12523–28
17. Cowing DW, Gross CA. 1989. Interaction of *Escherichia coli* RNA polymerase holoenzyme containing σ^{32} with heat shock promoters: DNase I footprinting and methylation protection. *J. Mol. Biol.* 210(3):513–20
18. Cowing DW, Mecsas J, Record MT Jr, Gross CA. 1989. Intermediates in the formation of the open complex by RNA polymerase holoenzyme containing the σ factor sigma 32 at the *groE* promoter. *J. Mol. Biol.* 210(3):521–30
19. Darst SA. 2001. Bacterial RNA polymerase. *Curr. Opin. Struct. Biol.* 11(2):155–62
20. Darst SA, Feklistov A, Gross CA. Promoter melting by an alternative σ, one base at a time. *Nat. Struct. Mol. Biol.* 21(4):350–51
21. deHaseth PL, Helmann JD. 2005. Open complex formation by *Escherichia coli* RNA polymerase: the mechanism of polymerase-induced strand separation of double helical DNA. *Mol. Microbiol.* 16(5):817–24
22. deHaseth PL, Zupancic ML, Record MT Jr. 1998. RNA polymerase-promoter interactions: the comings and goings of RNA polymerase. *J. Bacteriol.* 180(12):3019–25
23. Dickson RC, Abelson J, Barnes WM, Reznikoff WS. 1975. Genetic regulation: the Lac control region. *Science* 187(4171):27–35
24. Dombroski AJ, Walter WA, Record MT Jr, Siegele DA, Gross CA. 1992. Polypeptides containing highly conserved regions of transcription initiation factor σ^{70} exhibit specificity of binding to promoter DNA. *Cell* 70(3):501–12
25. Dove SL, Darst SA, Hochschild A. 2003. Region 4 of σ as a target for transcription regulation. *Mol. Microbiol.* 48(4):863–74
26. Feklistov A. 2013. RNA polymerase: in search of promoters. *Ann. N. Y. Acad. Sci.* 1293:25–32
27. Feklistov A, Barinova N, Sevostyanova A, Heyduk E, Bass I, et al. 2006. A basal promoter element recognized by free RNA polymerase σ subunit determines promoter recognition by RNA polymerase holoenzyme. *Mol. Cell* 23(1):97–107
28. Feklistov A, Darst SA. 2011. Structural basis for promoter -10 element recognition by the bacterial RNA polymerase σ subunit. *Cell* 147(6):1257–69

29. Gribskov M, Burgess RR. 1986. Sigma factors from *E. coli*, *B. subtilis*, phage SP01, and phage T4 are homologous proteins. *Nucleic Acids Res.* 14(16):6745–63
30. Grossman AD, Erickson JW, Gross CA. 1984. The *htpR* gene product of *E. coli* is a sigma factor for heat-shock promoters. *Cell* 38(2):383–90
31. Haldenwang WG, Lang N, Losick R. 1981. A sporulation-induced sigma-like regulatory protein from *B. subtilis*. *Cell* 23(2):615–24
32. Haldenwang WG, Losick R. 1980. Novel RNA polymerase sigma factor from *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA* 77(12):7000–4
33. Haugen SP, Berkmen MB, Ross W, Gaal T, Ward C, Gourse RL. 2006. rRNA Promoter regulation by nonoptimal binding of σ region 1.2: an additional recognition element for RNA polymerase. *Cell* 125(6):1069–82
34. Helmann JD, Chamberlin MJ. 1988. Structure and function of bacterial sigma factors. *Annu. Rev. Biochem.* 57:839–72
35. Helmann JD, deHaseth PL. 1999. Protein–nucleic acid interactions during open complex formation investigated by systematic alteration of the protein and DNA binding partners. *Biochemistry* 38(19):5959–67
36. Heyduk E. 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):12362–69
37. Hook-Barnard IG, Hinton DM. 2007. Transcription initiation by mix and match elements: flexibility for polymerase binding to bacterial promoters. *Gene Regul. Syst. Biol.* 1:275–93
38. Hsu LM. 2002. Promoter clearance and escape in prokaryotes. *Biochim. Biophys. Acta* 1577(2):191–207
39. Jain D, Nickels BE, Sun L, Hochschild A, Darst SA. 2004. Structure of a ternary transcription activation complex. *Mol. Cell* 13(1):45–53
40. Juang YL, Helmann JD. 1994. A promoter melting region in the primary σ factor of *Bacillus subtilis*: identification of functionally important aromatic amino acids. *J. Mol. Biol.* 235(5):1470–88
41. Kapanidis AN, Margeat E, Laurence TA, Doose S, Ho SO, et al. 2005. Retention of transcription initiation factor σ^{70} in transcription elongation: single-molecule analysis. *Mol. Cell* 20(3):347–56
42. Kapanidis AN, Strick T. 2009. Biology, one molecule at a time. *Trends Biochem. Sci.* 34(5):234–43
43. Kassavetis GA, Elliott T, Rabussay DP, Geiduschek EP. 1983. Initiation of transcription at phage T4 late promoters with purified RNA polymerase. *Cell* 33(3):887–97
44. Kassavetis GA, Geiduschek EP. 1984. Defining a bacteriophage T4 late promoter: bacteriophage T4 gene 55 protein suffices for directing late promoter recognition. *Proc. Natl. Acad. Sci. USA* 81(16):5101–5
45. Khesin RB, Shemyakin MF, Gorlenko AM, Mindlin SZ, Ilyina TS. 1969. Studies on the RNA polymerase in *Escherichia coli* K12 using the mutation affecting its activity. *J. Mol. Biol.* 42(3):401–11
46. Koo B-M, Rhodius VA, Campbell EA, Gross CA. 2009. Dissection of recognition determinants of *Escherichia coli* σ^{32} suggests a composite -10 region with an 'extended -10' motif and a core -10 element. *Mol. Microbiol.* 72(4):815–29
47. Koo B-M, Rhodius VA, Campbell EA, Gross CA. 2009. Mutational analysis of *Escherichia coli* σ^{28} and its target promoters reveals recognition of a composite -10 region, comprised of an 'extended -10' motif and a core -10 element. *Mol. Microbiol.* 72(4):830–43
48. Koo B-M, Rhodius VA, Nonaka G, deHaseth PL, Gross CA. 2009. Reduced capacity of alternative sigmas to melt promoters ensures stringent promoter recognition. *Genes Dev.* 23(20):2426–36
49. Krakow JS, Daley K, Karstadt M. 1969. *Azotobacter vinelandii* RNA polymerase. VII. Enzyme transitions during unprimed r[I-C] synthesis. *Proc. Natl. Acad. Sci. USA* 62(2):432–37
50. Kulbachinskiy A, Mustaev A. 2006. Region 3.2 of the subunit contributes to the binding of the 3′-initiating nucleotide in the RNA polymerase active center and facilitates promoter clearance during initiation. *J. Biol. Chem.* 281(27):18273–76
51. Lane WJ, Darst SA. 2006. The structural basis for promoter -35 element recognition by the group IV σ factors. *PLoS Biol.* 4(9):e269
52. Lee HJ. 2004. An unsubstituted C2 hydrogen of adenine is critical and sufficient at the -11 position of a promoter to signal base pair deformation. *J. Biol. Chem.* 279(17):16899–902
53. Lee S, Bowman BR, Ueno Y, Wang S, Verdine GL. 2008. Synthesis and structure of duplex DNA containing the genotoxic nucleobase lesion N7-methylguanine. *J. Am. Chem. Soc.* 130(35):11570–71

54. Lim HM, Lee HJ, Roy S, Adhya S. 2001. A "master" in base unpairing during isomerization of a promoter upon RNA polymerase binding. *Proc. Natl. Acad. Sci. USA* 98(26):14849–52
55. Lonetto M, Gribskov M, Gross CA. 1992. The σ^{70} family: sequence conservation and evolutionary relationships. *J. Bacteriol.* 174(12):3843–49
56. Lonetto MA, Brown KL, Rudd KE, Buttner MJ. 1994. Analysis of the *Streptomyces coelicolor sigE* gene reveals the existence of a subfamily of eubacterial RNA polymerase σ factors involved in the regulation of extracytoplasmic functions. *Proc. Natl. Acad. Sci. USA* 91(16):7573–77
57. Losick R, Pero J. 1981. Cascades of sigma factors. *Cell* 25(3):582–84
58. Marr MT. 1997. Promoter recognition as measured by binding of polymerase to nontemplate strand oligonucleotide. *Science* 276(5316):1258–60
59. Mascher T, Hachmann A-B, Helmann JD. 2007. Regulatory overlap and functional redundancy among *Bacillus subtilis* extracytoplasmic function σ factors. *J. Bacteriol.* 189(19):6919–27
60. Mecsas J, Cowing DW, Gross CA. 1991. Development of RNA polymerase-promoter contacts during open complex formation. *J. Mol. Biol.* 220(3):585–97
61. Mekler V, Kortkhonjia E, Mukhopadhyay J, Knight J, Revyakin A, et al. 2002. Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell* 108(5):599–614
62. Minakhin L, Severinov K. 2003. On the role of the *Escherichia coli* RNA polymerase σ^{70} region 4.2 and α-subunit C-terminal domains in promoter complex formation on the extended -10 *gal*P1 promoter. *J. Biol. Chem.* 278(32):29710–18
63. Mooney RA, Darst SA, Landick R. 2005. Sigma and RNA polymerase: an on-again, off- again relationship? *Mol. Cell* 20(3):335–45
64. Mooney RA, Davis SE, Peters JM, Rowland JL, Ansari AZ, Landick R. 2009. Regulator trafficking on bacterial transcription units in vivo. *Mol. Cell* 33(1):97–108
65. Moyle H, Waldburger C, Susskind MM. 1991. Hierarchies of base pair preferences in the P22 *ant* promoter. *J. Bacteriol.* 173(6):1944–50
66. Mukhopadhyay J, Kapanidis AN, Mekler V, Kortkhonjia E, Ebright YW, Ebright RH. 2001. Translocation of σ^{70} with RNA polymerase during transcription: fluorescence resonance energy transfer assay for movement relative to DNA. *Cell* 106(4):453–63
67. Murakami KS, Darst SA. 2003. Bacterial RNA polymerases: the wholo story. *Curr. Opin. Struct. Biol.* 13(1):31–39
68. Murakami KS, Masuda S, Campbell EA, Muzzin O, Darst SA. 2002. Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* 296(5571):1285–90
69. Murakami KS, Masuda S, Darst SA. 2002. Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science* 296(5571):1280–84
70. Nickels BE, Garrity SJ, Mekler V, Minakhin L, Severinov K, et al. 2005. The interaction between σ^{70} and the β-flap of *Escherichia coli* RNA polymerase inhibits extension of nascent RNA during early elongation. *Proc. Natl. Acad. Sci. USA* 102(12):4488–93
71. Nickels BE, Mukhopadhyay J, Garrity SJ, Ebright RH, Hochschild A. 2004. The σ^{70} subunit of RNA polymerase mediates a promoter-proximal pause at the *lac* promoter. *Nat. Struct. Mol. Biol.* 11(6):544–50
72. Nonaka G, Blankschien M, Herman C, Gross CA, Rhodius VA. 2006. Regulon and promoter analysis of the *E. coli* heat-shock factor, σ^{32}, reveals a multifaceted cellular response to heat stress. *Genes Dev.* 20(13):1776–89
73. Perdue SA, Roberts JW. 2011. σ^{70}-dependent transcription pausing in *Escherichia coli*. *J. Mol. Biol.* 412(5):782–92
74. Pribnow D. 1975. Bacteriophage T7 early promoters: nucleotide sequences of two RNA polymerase binding sites. *J. Mol. Biol.* 99(3):419–43
75. Protozanova E, Yakovchuk P, Frank-Kamenetskii MD. 2004. Stacked-unstacked equilibrium at the nick site of DNA. *J. Mol. Biol.* 342(3):775–85
76. Raffaelle M, Kanin EI, Vogt J, Burgess RR, Ansari AZ. 2005. Holoenzyme switching and stochastic release of sigma factors from RNA polymerase in vivo. *Mol. Cell* 20(3):357–66
77. Reppas NB, Wade JT, Church GM, Struhl K. 2006. The transition between transcriptional initiation and elongation in *E. coli* is highly variable and often rate limiting. *Mol. Cell* 24(5):747–57

78. Reznikoff WS, Siegele DA, Cowing DW, Gross CA. 1985. The regulation of transcription initiation in bacteria. *Annu. Rev. Genet.* 19:355–87
79. Rhodius VA, Mutalik VK. 2010. Predicting strength and function for promoters of the *Escherichia coli* alternative sigma factor, σ^E. *Proc. Natl. Acad. Sci. USA* 107(7):2854–59
80. Rhodius VA, Mutalik VK, Gross CA. 2012. Predicting the strength of UP-elements and full-length *E. coli* σ^E promoters. *Nucleic Acids Res.* 40(7):2907–24
81. Rhodius VA, Segall-Shapiro TH, Sharon BD, Ghodasara A, Orlova E, et al. 2013. Design of orthogonal genetic switches based on a crosstalk map of σs, anti-σs, and promoters. *Mol. Syst. Biol.* 9:702
82. Rhodius VA, Suh WC, Nonaka G, West J, Gross CA. 2006. Conserved and variable functions of the σ^E stress response in related genomes. *PLoS Biol.* 4(1):e2
83. Ring BZ, Yarnell WS, Roberts JW. 1996. Function of *E. coli* RNA polymerase σ factor σ^{70} in promoter-proximal pausing. *Cell* 86(3):485–93
84. Roberts CW, Roberts JW. 1996. Base-specific recognition of the nontemplate strand of promoter DNA by *E. coli* RNA polymerase. *Cell* 86(3):495–501
85. Roberts JW, Yarnell W, Bartlett E, Guo J, Marr M, et al. 1998. Antitermination by bacteriophage λ Q protein. *Cold Spring Harb. Symp. Quant. Biol.* 63:319–25
86. Roy S, Lim HM, Liu M, Adhya S. 2004. Asynchronous basepair openings in transcription initiation: CRP enhances the rate-limiting step. *EMBO J.* 23(4):869–75
87. Saecker RM, Record MT Jr, deHaseth PL. 2011. Mechanism of bacterial transcription initiation: RNA polymerase—promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *J. Mol. Biol.* 412(5):754–71
88. Sanderson A, Mitchell JE, Minchin SD, Busby SJW. 2003. Substitutions in the *Escherichia coli* RNA polymerase σ^{70} factor that affect recognition of extended -10 elements at promoters. *FEBS Lett.* 544(1–3):199–205
89. Savinkova LK, Baranova LV, Knorre VL, Salganik RI. 1988. [Binding of RNA-polymerase from *Escherichia coli* with oligodeoxyribonucleotides homologous to transcribed and non-transcribed DNA stands in the "-10"-promoter region of bacterial genes]. *Mol. Biol. (Mosk.).* 22(3):807–12
90. Schaller H, Gray C, Herrmann K. 1975. Nucleotide sequence of an RNA polymerase binding site from the DNA of bacteriophage fd. *Proc. Natl. Acad. Sci. USA* 72(2):737–41
91. Schwartz EC, Shekhtman A, Dutta K, Pratt MR, Cowburn D, et al. 2008. A full-length group 1 bacterial sigma factor adopts a compact structure incompatible with DNA binding. *Chem. Biol.* 15(10):1091–103
92. Sen R, Nagai H, Hernandez VJ, Shimamoto N. 1998. Reduction in abortive transcription from the λP_R promoter by mutations in region 3 of the σ^{70} subunit of *Escherichia coli* RNA polymerase. *J. Biol. Chem.* 273(16):9872–77
93. Severinov K, Fenyö D, Severinova E, Mustaev A, Chait BT, et al. 1994. The sigma subunit conserved region 3 is part of "5′-face" of active center of *Escherichia coli* RNA polymerase. *J. Biol. Chem.* 269(33):20826–28
94. Severinova E, Severinov K, Fenyö D, Marr M, Brody EN, et al. 1996. Domain organization of the *Escherichia coli* RNA polymerase σ^{70} subunit. *J. Mol. Biol.* 263(5):637–47
95. Sevostyanova A, Djordjevic M, Kuznedelov K, Naryshkina T, Gelfand MS, et al. 2007. Temporal regulation of viral transcription during development of *Thermus thermophilus* bacteriophage φYS40. *J. Mol. Biol.* 366(2):420–35
96. Shimamoto N, Kamigochi T, Utiyama H. 1986. Release of the σ subunit of *Escherichia coli* DNA-dependent RNA polymerase depends mainly on time elapsed after the start of initiation, not on length of product RNA. *J. Biol. Chem.* 261(25):11859–65
97. Shorenstein RG, Losick R. 1973. Purification and properties of the σ subunit of ribonucleic acid polymerase from vegetative *Bacillus subtilis*. *J. Biol. Chem.* 248(17):6163–69
98. Shultzaberger RK, Chen Z, Lewis KA, Schneider TD. 2007. Anatomy of *Escherichia coli* 70 promoters. *Nucleic Acids Res.* 35(3):771–88
99. Siegele DA, Hu JC, Gross CA. 1988. Mutations in *rpoD*, the gene encoding the σ^{70} subunit of *Escherichia coli* RNA polymerase, that increase expression of the *lac* operon in the absence of CAP-cAMP. *J. Mol. Biol.* 203(1):29–37

100. Silverstone AE, Goman M, Scaife JG. 1972. ALT: a new factor involved in the synthesis of RNA by *Escherichia coli*. *Mol. Gen. Genet.* 118(3):223–34
101. Sorenson MK, Darst SA. 2006. Disulfide cross-linking indicates that FlgM-bound and free σ^{28} adopt similar conformations. *Proc. Natl. Acad. Sci. USA* 103(45):16722–27
102. Sorenson MK, Ray SS, Darst SA. 2004. Crystal structure of the flagellar σ/anti-σ complex σ^{28}/FlgM reveals an intact σ factor in an inactive conformation. *Mol. Cell* 14(1):127–38
103. Staro A, Sofia HJ, Dietrich S, Ulrich LE, Liesegang H, Mascher T. 2009. The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF) σ factor protein family. *Mol. Microbiol.* 74(3):557–81
104. Stragier P, Parsot C, Bouvier J. 1985. Two functional domains conserved in major and alternate bacterial sigma factors. *FEBS Lett.* 187(1):11–15
105. Talkington C, Pero J. 1978. Promoter recognition by phage SP01-modified RNA polymerase. *Proc. Natl. Acad. Sci.* 75(3):1185–89
106. Travers AA, Buckland R, Goman M, Le Grice SS, Scaife JG. 1978. A mutation affecting the sigma subunit of RNA polymerase changes transcriptional specificity. *Nature* 273(5661):354–58
107. Travers AA, Burgess RR. 1969. Cyclic re-use of the RNA polymerase sigma factor. *Nature* 222(5193):537–40
108. Ulrich LE, Zhulin IB. 2010. The MiST2 database: a comprehensive genomics resource on microbial signal transduction. *Nucleic Acids Res.* 38(Suppl. 1):D401–7
109. Vassylyev DG, Sekine S-I, Laptenko O, Lee J, Vassylyeva MN, et al. 2002. Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* 417(6890):712–19
110. Vuthoori S, Bowers CW, McCracken A, Dombroski AJ, Hinton DM. 2001. Domain 1.1 of the σ^{70} subunit of *Escherichia coli* RNA polymerase modulates the formation of stable polymerase/promoter complexes. *J. Mol. Biol.* 309(3):561–72
111. Wade JT, Struhl K. 2004. Association of RNA polymerase with transcribed regions in *Escherichia coli*. *Proc. Natl. Acad. Sci.* 101(51):17777–82
112. Waldburger C, Gardella T, Wong R, Susskind MM. 1990. Changes in conserved region 2 of *Escherichia coli* σ^{70} affecting promoter recognition. *J. Mol. Biol.* 215(2):267–76
113. Wecke T, Halang P, Staroń A, Dufour YS, Donohue TJ, Mascher T. 2012. Extracytoplasmic function σ factors of the widely distributed group ECF41 contain a fused regulatory domain. *MicrobiologyOpen* 1(2):194–213
114. Williams KP, Kassavetis GA, Esch FS, Geiduschek EP. 1987. Identification of the gene encoding an RNA polymerase-binding protein of bacteriophage T4. *J. Virol.* 61(2):597–99
115. Williams KP, Kassavetis GA, Geiduschek EP. 1987. Interactions of the bacteriophage T4 gene 55 product with *Escherichia coli* RNA polymerase: competition with *Escherichia coli* σ^{70} and release from late T4 transcription complexes following initiation. *J. Biol. Chem.* 262(25):12365–71
116. Wilson C, Dombroski AJ. 1997. Region 1 of σ^{70} is required for efficient isomerization and initiation of transcription by *Escherichia coli* RNA polymerase. *J. Mol. Biol.* 267(1):60–74
117. Wu CW, Yarbrough LR, Hillel Z, Wu FY. 1975. Sigma cycle during in vitro transcription: demonstration by nanosecond fluorescence depolarization spectroscopy. *Proc. Natl. Acad. Sci. USA* 72(8):3019–23
118. Yi C, Chen B, Qi B, Zhang W, Jia G, et al. 2012. Duplex interrogation by a direct DNA repair protein in search of base damage. *Nat. Struct. Mol. Biol.* 19(7):671–76
119. Young BA. 2004. Minimal machinery of RNA polymerase holoenzyme sufficient for promoter melting. *Science* 303(5662):1382–84
120. Young BA, Anthony LC, Gruber TM, Arthur TM, Heyduk E, et al. 2001. A coiled-coil from the RNA polymerase β′ subunit allosterically induces selective nontemplate strand binding by σ^{70}. *Cell* 105(7):935–44
121. Zenkin N, Kulbachinskiy A, Yuzenkova Y, Mustaev A, Bass I, et al. 2007. Region 1.2 of the RNA polymerase σ subunit controls recognition of the -10 promoter element. *EMBO J.* 26(4):955–64
122. Zhang Y, Feng Y, Chatterjee S, Tuske S, Ho MX, et al. 2012. Structural basis of transcription initiation. *Science* 338(6110):1076–80
123. Zuber P, Healy J, Carter HL, Cutting S, Moran CP, Losick R. 1989. Mutation changing the specificity of an RNA polymerase sigma factor. *J. Mol. Biol.* 206(4):605–14