Protein family review

The σ^{70} family of sigma factors

Mark SB Paget* and John D Helmann†

Addresses: *School of Biological Sciences, University of Sussex, Brighton BN1 9QG, UK. †Department of Microbiology, Cornell University, Ithaca, NY 14853-8101, USA.

Correspondence: Mark SB Paget. E-mail: M.Paget@sussex.ac.uk

Published: 3 January 2003

Genome ***Biology*** 2003, **4**:203

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2003/4/1/203

Summary

Members of the σ^{70} family of sigma factors are components of the RNA polymerase holoenzyme that direct bacterial or plastid core RNA polymerase to specific promoter elements that are situated 10 and 35 base-pairs upstream of transcription-initiation points. Members of the σ^{70} family also function as contact points for some activator proteins, such as PhoB and λcI, and play a role in the initiation process itself. The primary σ factor, which is essential for general transcription in exponentially growing cells, is reversibly associated with RNA polymerase and can be replaced by alternative σ factors that co-ordinately express genes involved in diverse functions, such as stress responses, morphological development and iron uptake. On the basis of gene structure and function, members of the σ^{70} family can broadly be divided into four main groups. Sequence alignments of the σ^{70} family members reveal that they have four conserved regions, although the highest conservation is found in regions 2 and 4, which are involved in binding to RNA polymerase, recognizing promoters and separating DNA strands (so-called 'DNA melting'). The division of the linear sequence of σ^{70} factors into four regions is largely supported by recent structural data indicating that primary σ factors have three stable domains that incorporate regions 2, 3 and 4. Furthermore, structures of the RNA polymerase holoenzyme have revealed that these domains of σ^{70} are spread out across one face of RNA polymerase. These structural data are starting to illuminate the mechanistic role of σ factors in transcription initiation.

Gene organization and evolutionary history

The bacterial core RNA polymerase complex, which consists of five subunits ($\beta\beta'\alpha_2\omega$), is sufficient for transcription elongation and termination but is unable to initiate transcription. Transcription initiation from promoter elements requires a sixth, dissociable subunit called a σ factor, which reversibly associates with the core RNA polymerase complex to form a holoenzyme. The vast majority of σ factors belong to the so-called σ^{70} family, reflecting their relationship to the principal σ factor of *Escherichia coli*, σ^{70}. A second family of σ factors, the σ^{54} family, comprises proteins that are functionally similar to, but structurally distinct from, σ^{70} of *E. coli*. Here, we limit ourselves to the σ^{70} family.

Members of the σ^{70} family direct RNA polymerase to specific promoter elements that are usually 5-6 base-pairs (bp) in length and are centred 10 and 35 bp upstream (positions -10 and -35) of the transcription initiation site. They also function in the melting of promoter DNA and the early stages of elongation of transcripts. The discovery of the σ factor as a dissociable RNA polymerase subunit [1] heralded the subsequent finding that RNA polymerase recruits alternative σ factors as a means of switching on specific regulons [2]. Multiple members of the σ^{70} family have since been discovered in most bacteria, with up to 63 encoded by a single genome, in the case of the antibiotic-producing bacterium *Streptomyces coelicolor* [3]. Furthermore, σ^{70}-related factors have

been discovered in higher plants, in which they act together with a bacterial-type RNA polymerase to direct transcription in the plastid [4].

The σ^{70} family has been divided broadly into four phylogenetic groups on the basis of gene structure and function [5,6]. Group 1 consists of the essential primary σ factors, each of which is closely related to σ^{70} of *E. coli*. Group 2 proteins are closely related to the primary σ factors but are dispensable for bacterial cell growth. Group 3 σ factors are more distantly related to σ^{70} and usually activate regulons in response to a specific signal, such as a developmental checkpoint or heat shock. The group 3 σ factors can be further divided into several clusters of functionally related proteins with roles in sporulation, flagella biosynthesis, or the heat-shock response, for example, (Figure 1). Finally, group 4 accommodates the numerically largest, but highly diverged extracytoplasmic function (ECF) subfamily, most members of which respond to signals from the extracytoplasmic environment, such as the presence of misfolded proteins in the periplasmic space. Whereas most bacteria have a single group 1 primary σ factor, the number of other group members varies widely, reflecting in part the different physiological and developmental characteristics of the various organisms. For example, whereas *E. coli* has two members in each of group 3 and group 4, the physiologically and developmentally complex *S. coelicolor* has 10 group 3 members and 49 group 4 members [3]. A phylogenetic tree that illustrates the relationships between σ^{70}-family members from four different bacteria is shown in Figure 1. (For in-depth phylogenetic analyses of the σ^{70} family see, for example, [7,8].)

For historical reasons the nomenclature of the σ superfamily is complex. In *E. coli* and several other Gram-negative bacteria, σ factor genes are designated *rpo* (for RNA polymerase subunit), whereas in most Gram-positive bacteria the genes are designated *sig*. The proteins may be designated σ with a superscript reflecting the molecular weight or gene name, or may have an arbitrary single-letter designation. In the post-genomic era, the situation has naturally become further complicated with the realization that some organisms have more σ factors than there are letters in the alphabet.

Characteristic structural features

Sequence alignments of the σ^{70} family members reveal four conserved regions that can be further divided into subregions (Figure 2) [5]. Only regions 2 and 4 are well conserved in all members of the σ^{70} family, and include subregions involved in binding to the core RNA polymerase complex, recognition of the -10 and -35 promoter (regions 2.4 and 4.2, respectively), and promoter melting (region 2.3). Much of region 1 is conserved only between the primary and closely related σ factors (groups 1 and 2), and region 1 appears to function in antagonizing the DNA-binding activity of the σ factor. Region 3, which is virtually absent from ECF σ factors, includes a subregion 3.0 (previously named 2.5) that interacts with DNA upstream of the -10 element in certain 'extended-10' promoters that lack the -35 element [9,10]. The linear division of σ^{70} factors into functionally distinct regions is largely confirmed by recent structural data, which revealed that primary σ factors have three flexibly linked compact domains, σ_2, σ_3, and σ_4, which incorporate regions 2, 3 and 4, respectively [10].

The crystal structure of the σ_2 domain has been solved for two primary σ factors (σ^{70} of *E. coli* and σ^{A} of *Thermus aquaticus*) [10,11], and one ECF σ factor (σ^{R} of *S. coelicolor*) [12]. Discounting a non-conserved region that occurs between subregions 1.2 and 2.4 in some primary σ factors, each σ_2 domain is composed of a bundle of three α helices that is virtually identical in all three structures analyzed. The second helix of this bundle is a major point for contact with a coiled-coil domain in the β′ subunit of the core RNA polymerase complex [13]. The third helix of the bundle includes conserved residues along one face that are involved in DNA melting and in recognition of the -10 promoter element (Figure 2b).

The σ_3 domain, which is less conserved between members of the σ^{70} family, and is absent from ECF σ factors, is also a three-helix domain, the first helix of which contains the residues implicated in contacting DNA upstream of extended -10 promoters [10]. The σ_4 domain has two pairs of α helices; the carboxy-terminal pair forms a helix-turn-helix motif that contacts the promoter DNA in the region from -30 to -38 [10,14]. The spectacular crystallographic views obtained recently of the σ^{70} factor in the holoenzyme complex [14,15] revealed that the σ_2, σ_3 and σ_4 domains extend across a wide area of the RNA polymerase, with an interface between the core complex and σ^{70} of more than 8000 Å^2. Also revealed was a gap of approximately 45 Å between the σ_3 and σ_4 domains, taken up by a 33-residue linker (the 'σ_3-σ_4 linker') that, strikingly, travels close to the active site of RNA polymerase and through the channel from which the growing transcript exits the RNA polymerase complex before connecting with the σ_4 domain.

Localization and function

Whereas the function of the essential group 1 σ^{70} factors is to direct general transcription, the accessory σ factors of groups 2-4 usually function to turn on specific gene sets in response to an appropriate signal. Their functions can be divided into three very broad categories: stress responses, development and ancillary metabolism. The wide variety of stress responses controlled by members of the σ^{70} family includes the stationary phase and general stress responses (mediated by, for example, σ^{S} in *E. coli* and σ^{B} in *Bacillus subtilis*), intracellular and extracytoplasmic protein misfolding (regulated in *E. coli* by σ^{32} and σ^{E}, respectively), oxidative stress (e.g. σ^{R} in *S. coelicolor*), osmotic stress (σ^{M}

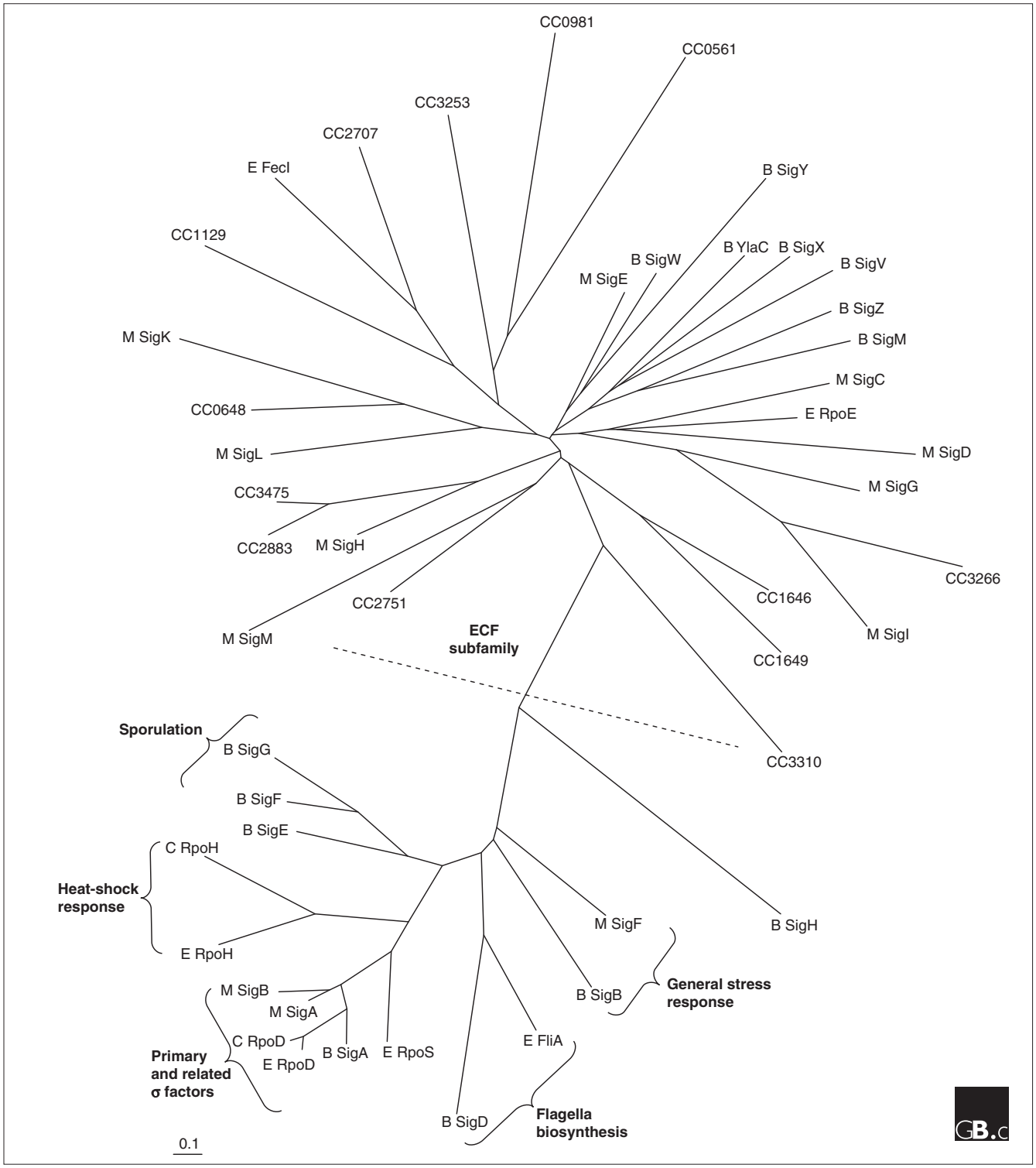

Figure 1
Phylogenetic relationships between members of the σ^{70} family of sigma factors from four diverse bacteria: *E. coli* (E); *Caulobacter cresentus* (CC or C); *B. subtilis* (B); and *Mycobacterium tuberculosis* (M). For *C. cresentus*, gene numbers from the annotated genome sequence are given. Note that the primary and related σ factors comprise groups 1 and 2, the group 3 σ factors are divided into functionally related groups (sporulation, flagella biosynthesis, general stress response and heat-shock response), and the divergent ECF σ factors comprise group 4. Particularly striking is the distance between the ECF subfamily and other members of the σ^{70} family, as well as divergence within the ECF subfamily itself. The unrooted tree was constructed using the Phylip programs PROTDIST and NEIGHBOUR from a multiple amino-acid sequence alignment made by the program ClustalW. Only conserved regions 2 and 4 were included in the alignments.

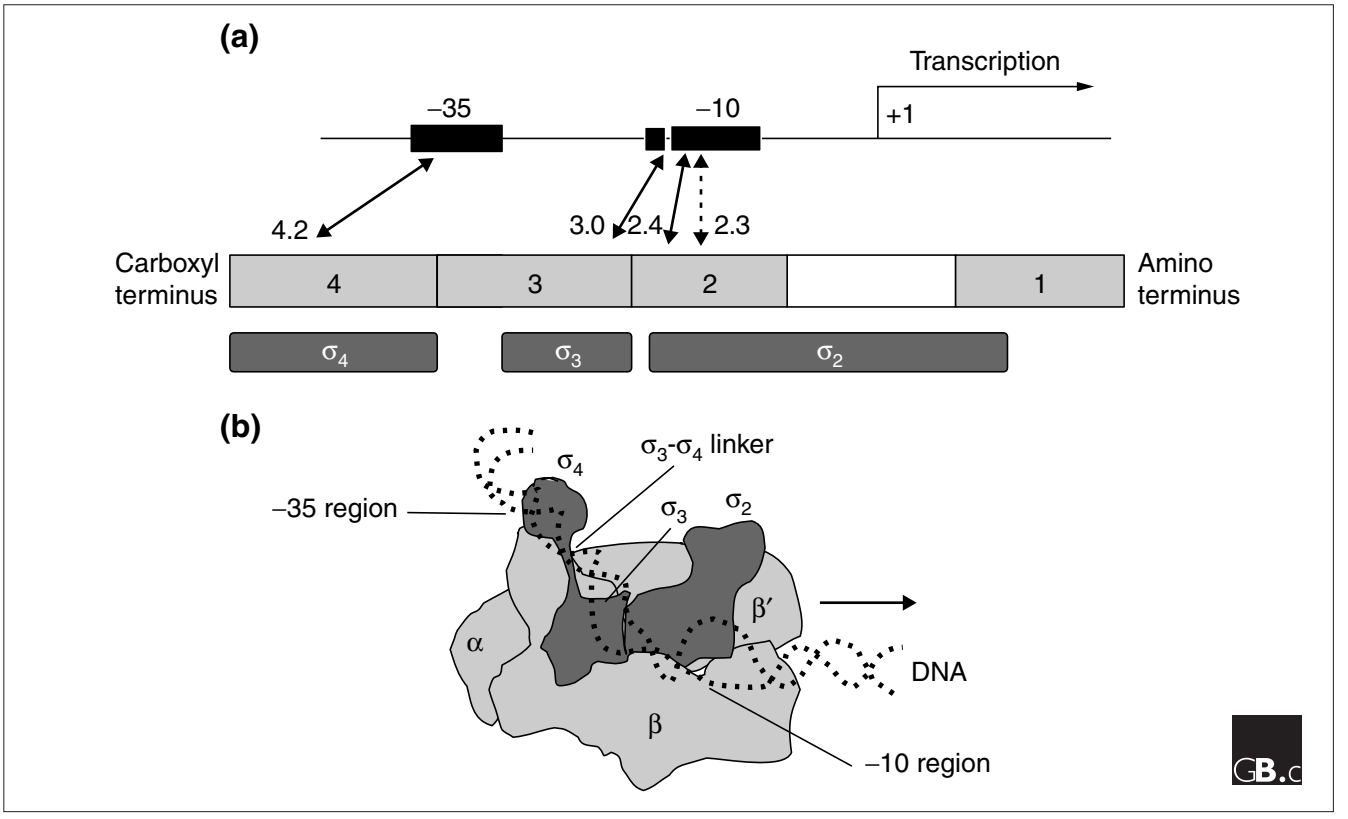

Figure 2
Structural characteristics of *E. coli* σ^{70}. **(a)** The protein sequence has been divided into four regions on the basis of sequence conservation with other members of the σ^{70} family. Residues in the carboxy-terminal part of region 4 (subregion 4.2) form a helix-turn-helix motif that contacts the -35 element of the promoter. Residues from conserved regions 2 and 3 cooperate to mediate recognition of the -10 region and melting of the DNA. A residue in the amino-terminal part of region 3 (3.0) contacts the conserved TG motif in the extended -10 element of certain promoters that do not require a -35 region. Residues from an α helix in region 2 that corresponds to the conserved subregions 2.3 and 2.4 interact intimately with the -10 element. Subregion 2.3 is thought to interact primarily with single-stranded DNA in the open complex (dashed arrow). The three domains of the σ factor observed by X-ray crystallography (σ_2, σ_3 and σ_4) are indicated underneath the linear structure. Note that the protein domains correspond closely (although not precisely) with the regions assigned by sequence comparisons. **(b)** A model for the interaction of RNA polymerase holoenzyme (containing β, β', two α, and one ω subunit in addition to the σ factor) with promoter DNA. The model is based on crystallographic analyses of σ domains, holoenzyme, and holoenzyme-model DNA complexes [10,14,15,25]. The major functional domains of the σ factor are shown in dark grey. The bold arrow indicates the direction of transcription. Although the template strand in the transcription bubble passes underneath the β unit and the σ_2 domain, the path of the DNA is shown throughout its length. Adapted from [21].

in *B. subtilis*) and cell-wall stress (controlled by, for example, σ^E in *S. coelicolor* and σ^W in *B. subtilis*). Developmental programs under the control of σ^{70} family members include flagella biosynthesis (involving σ^D in *B. subtilis* and σ^F in *Salmonella typhimurium*), endospore formation (mediated by, for example, σ^E, σ^K, σ^F and σ^G in *B. subtilis*), and exospore formation (σ^{WhiG} in *S. coelicolor*). Ancillary metabolic functions that are controlled by σ factors include iron uptake (σ^{FecI} in *E. coli* and σ^{PvdS} in pseudomonads). There are clearly many more functions to be discovered, especially amongst members of the ECF subfamily; for example, in *S. coelicolor* the function of only three of its 49 ECF σ factors is understood.

The activity of σ factors, and the consequent activity of the promoters they recognize, can be controlled at many different levels: *de novo* synthesis (at the transcriptional or translational level), by post-translational processing, by proteolytic degradation, and by post-translational inhibition. Indeed, some σ factors, such as σ^S factor of *E. coli*, are regulated at most of these levels [16]. Of widespread importance is post-translational inhibition by so-called anti-σ factors, proteins that reversibly bind to the σ factor thereby preventing its interaction with the core RNA polymerase [17,18]. In these cases, the signal that leads to the activation of the σ factor and the induction of the σ regulon somehow modify the σ-binding activity of the anti-σ factor.

Mechanism

Recent structural studies together with an extensive catalog of biochemical data are starting to shed light on the function of σ^{70} family members in transcription initiation. Once they

become part of the holoenzyme, the promoter-recognition determinants of subregions 2.4 and 4.2 of the σ factor are solvent-exposed and appropriately separated. This conformation allows subregions 2.4 and 4.2 to interact with the -10 and -35 elements, respectively, to form a so-called 'closed' complex in which the promoter DNA remains base-paired. At promoters that lack -35 regions or have -35 elements that deviate significantly from the consensus sequence, the σ_4 domain can stimulate formation of the closed complex by contacting activator proteins, such as λcI, which is bound upstream, or PhoB, which is bound downstream of the complex on the DNA [19].

In the following stage, the DNA in the region from position -11 to +4, which partially overlaps with the -10 element, melts in a process called isomerization, the mechanistic details of which are unresolved but probably include several kinetically distinct intermediate states. Once separated, the two DNA strands take different paths, with the template strand approaching the active site of the RNA polymerase and the non-template strand being held by conserved aromatic residues in region 2.3 of the σ factor that had previously been implicated in DNA melting [20]. The σ factor may also play a role in the next stage, that of *de novo* RNA synthesis, by donating a disordered loop from the σ_3-σ_4 linker into the active site of the RNA polymerase; this might perhaps stabilize the initiating nucleotide [15]. Alternatively, the disordered loop may stabilize the open complex by preventing reannealing close to the transcription start site [21]. Finally, after a nascent RNA of 8-10 nucleotides has been synthesized, the σ factor is released or moves out of the way to allow elongation to proceed further and RNA polymerase to escape the promoter. The discovery that the σ_3-σ_4 linker is located in the RNA exit channel of the RNA polymerase suggests a mechanism of promoter clearance that involves the nascent RNA displacing the linker, in turn weakening the interaction between the core RNA polymerase and the σ_4 domain and ultimately the rest of the σ factor [14,15]. Interestingly, σ^{70} of *E. coli* can in some instances interact with the exposed non-template strand early in the elongation process, and these interactions can lead to transient pausing of the elongation complex [22]. Furthermore, recent evidence suggests that σ^{70} may remain associated with the core RNA polymerase complex during the elongation process [23,24].

Frontiers

The recent structural information on primary σ factors has had a major impact on our understanding of the mechanistic role of the σ^{70} family of σ factors in transcription initiation. Numerous puzzles remain, however. It is not clear how transcription-activator proteins can modulate the complex conformational changes that accompany promoter recognition and melting. The timing and extent of release of the σ factor during the transition to the transcript-elongation phase is a topic of continuing controversy. Finally, the extent to which σ factors might be retained in early, or perhaps later, elongation complexes and might mediate a sequence-responsive pause in transcription is not resolved. Although a reasonably detailed picture of the action of σ^{70} can now be envisaged, the sequence divergence noted within the σ^{70} family raises questions about how other family members mediate promoter recognition and melting and how they interact with their regulators. Finally, the many bacterial genome sequencing projects have revealed a huge gap in our understanding of the biological function of the many newly discovered σ factors. Even in some of the best characterized model systems, such as *B. subtilis*, there is a frustrating lack of knowledge regarding the regulation, roles, and possible redundancies among the various σ factors.

Acknowledgements

Work on σ factors in the laboratory of JDH is supported by the National Institutes of Health and by the BBSRC and the Wellcome Trust in the laboratory of MSBP.

References

1. Burgess RR, Travers AA, Dunn JJ, Bautz EK: **Factor stimulating transcription by RNA polymerase.** *Nature* 1969, **221:**43-46.
   A seminal paper describing the discovery of the σ subunit of RNA polymerase.
2. Losick R, Pero J: **Cascades of sigma factors.** *Cell* 1981, **25:**582-584.
   A classical minireview summarizing early studies on the role of σ factor cascades in the development of phage SPO1 and endospore formation in *B. subtilis*.
3. Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, *et al.*: **Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2).** *Nature* 2002, **417:**141-147.
   Complete sequence of the *S. coelicolor* genome, a genome that encodes 63 σ factors.
4. Isono K, Shimizu M, Yoshimoto K, Niwa Y, Satoh K, Yokota A, Kobayashi H: **Leaf-specifically expressed genes for polypeptides destined for chloroplasts with domains of σ^{70} factors of bacterial RNA polymerases in *Arabidopsis thaliana*.** *Proc Natl Acad Sci USA* 1997, **94:**14948-14953.
   The first description of genes that encode σ^{70}-like sigma factors in multicellular eukaryotes.
5. Lonetto M, Gribskov M, Gross CA: **The σ^{70} family: sequence conservation and evolutionary relationships.** *J Bacteriol* 1992, **174:**3843-3849.
   A classical review of the structure, function and evolutionary relationships between bacterial σ factors.
6. Helmann JD: **The extracytoplasmic function (ECF) sigma factors.** *Adv Microb Physiol* 2002, **46:**47-110.
   A comprehensive review of the ECF subfamily of σ factors.
7. Lonetto MA, Brown KL, Rudd KE, Buttner MJ: **Analysis of the *Streptomyces coelicolor sigE* gene reveals the existence of a subfamily of eubacterial RNA polymerase sigma factors involved in the regulation of extracytoplasmic functions.** *Proc Natl Acad Sci USA* 1994, **91:**7573-7577.
   The discovery of the ECF subfamily of σ factors.
8. Wösten MM: **Eubacterial sigma-factors.** *FEMS Microbiol Rev* 1998, **22:**127-150.
   A comprehensive review of bacterial σ factors.
9. Barne KA, Bown JA, Busby SJ, Minchin SD: **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* 1997, **16:**4034-4040.
   The isolation of a new *rpoD* mutant defines a region of σ^{70} involved in contacting promoters with 'extended -10' elements.

10. Campbell EA, Muzzin O, Chlenov M, Sun JL, Olson CA, Weinman O, Trester-Zedlitz ML, Darst SA: **Structure of the bacterial RNA polymerase promoter specificity sigma subunit.** *Mol Cell* 2002, **9:**527-539.
Crystal structures of the stable σ_2, σ_3 and σ_4 domains of *Thermus aquaticus* and of the σ_4 domain complexed with promoter DNA.
11. Malhotra A, Severinova E, Darst SA: **Crystal structure of a σ^{70} subunit fragment from *E. coli* RNA polymerase.** *Cell* 1996, **87:**127-136.
Crystal structure of conserved subregions 1.2-2.4 of *E. coli* σ^{70}.
12. Wei L, Stevenson CEM, Burton N, Jakimowicz P, Paget MSB, Buttner MJ, Lawson DM, Kleanthous C: **Identification and structure of the anti-sigma factor-binding domain of the disulphide-stress regulated sigma factor factor σ^{R} from *Streptomyces coelicolor.*** *J Mol Biol* 2002, **323:**225-236.
Crystal structure of conserved region 2 of an ECF σ factor.
13. Young BA, Anthony LC, Gruber TM, Arthur TM, Heyduk E, Lu CZ, Sharp MM, Heyduk T, Burgess RR, Gross CA: **A coiled-coil from the RNA polymerase β' subunit allosterically induces selective nontemplate strand binding by σ^{70}.** *Cell* 2001, **105:**935-944.
A 48-residue coiled-coil domain from the β' subunit of RNA polymerase is sufficient to stimulate recognition of the -10 region non-template strand by σ^{70}.
14. Vassylyev DG, Sekine S, Laptenko O, Lee J, Vassylyeva MN, Borukhov S, Yokoyama S: **Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution.** *Nature* 2002, **417:**712-719.
Crystal structure of the RNA polymerase holoenzyme of *Thermus thermophilus*.
15. Murakami KS, Masuda S, Darst SA: **Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution.** *Science* 2002, **296:**1280-1284.
Crystal structure of the RNA polymerase holoenzyme of *Thermus aquaticus*.
16. Hengge-Aronis R: **Signal transduction and regulatory mechanisms involved in control of the σ^{S} (RpoS) subunit of RNA polymerase.** *Microbiol Mol Biol Rev* 2002, **66:**373-395.
A review of the *E. coli* stationary phase sigma factor σ^{S}.
17. Hughes KT, Mathee K: **The anti-sigma factors.** *Annu Rev Microbiol* 1998, **52:**231-286.
This article and [18] are comprehensive reviews of anti-σ factors.
18. Helmann JD: **Anti-sigma factors** *Curr Opin Microbiol* 1999, **2:**135-141.
See [17].
19. Gross CA, Chan C, Dombroski A, Gruber T, Sharp M, Tupy J, Young B: **The functional and regulatory roles of sigma factors in transcription.** *Cold Spring Harb Symp Quant Biol* 1998, **63:**141-155.
A detailed review of the role of the σ factor in transcription initiation.
20. Juang YL, Helmann JD: **A promoter melting region in the primary sigma factor of *Bacillus subtilis*. Identification of functionally important aromatic amino acids.** *J Mol Biol* 1994, **235:**1470-1488.
The characterization of mutants in subregion 2.3 of *Bacillus subtilis* σ^{A} that are specifically impaired in open-complex formation.
21. Young BA, Gruber TM, Gross CA: **Views of transcription initiation.** *Cell* 2002, **109:** 417-420.
A review that places the recent RNA polymerase structural studies into the wider context.
22. Ring BZ, Yarnell WS, Roberts JW: **Function of *E. coli* RNA polymerase sigma factor σ^{70} in promoter-proximal pausing.** *Cell* 1996, **86:**485-493.
This article provides evidence that σ^{70} remains bound to core RNA polymerase when contacts between the σ factor and the -10 promoter element are broken, and then moves to a pause-inducing sequence further downstream.
23. Bar-Nahum G, Nudler E: **Isolation and characterization of σ^{70} retaining transcription elongation complexes from *Escherichia coli*.** *Cell* 2001, **106:**443-451.
Identification of σ^{70}-retaining elongation complexes, which are especially prevalent during stationary phase and have a higher ability to support multiple rounds of initiation compared to elongation complexes lacking σ^{70}.
24. Mukhopadhyay J, Kapanidis AN, Mekler V, Kortkhonjia E, Ebright YW, Ebright RH: **Translocation of σ^{70} with RNA polymerase during transcription: fluorescence resonance energy transfer assay for movement relative to DNA.** *Cell* 2001, **106:**453-463.
In the majority of transcription complexes, σ^{70} is not released from RNA polymerase upon transition from initiation to elongation.
25. Murakami KS, Masuda S, Campbell EA, Muzzin O, Darst SA: **Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex.** *Science* 2002, **296:**1285-1290.
The first structural study of RNA polymerase holoenzyme complexed with promoter DNA.