©2002 New Science Press Ltd new-science-press.com

Sigma Factors and Promoter Recognition in Bacteria

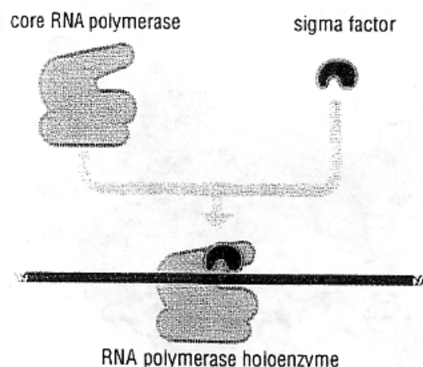

Figure 9-3.1 Promoter recognition by the bacterial RNA polymerase holoenzyme The core RNA polymerase associates with a sigma subunit which binds to a specific promoter sequence (red) in the DNA.

9-8 Sigma factors direct bacterial RNA polymerase to promoters

In order to recognize their promoters, bacterial RNA polymerase enzymes require a specialized subunit called the **sigma factor** (σ), which directly contacts the promoter sequence (Figure 9-3.1). The complex formed by the sigma subunit with the remaining polymerase core subunits constitutes the bacterial holoenzyme. Bacteria contain a variety of sigma factors that specifically recognize different promoter sequences. It is therefore the sigma factor that determines which genes are transcribed.

All cells have a **primary sigma factor**, which directs transcription from the promoters of essential housekeeping genes, and a variable number of **alternative sigma factors** whose levels or activities are increased in response to specific signals or stress conditions. Bacterial species differ significantly in the number of sigma factors they express, reflecting the different environmental conditions to which they must respond. *E. coli*, a symbiotic bacterium leading a relatively sheltered life in the gut of other organisms, has only seven sigma factors. In contrast, the genome of the soil organism *B. subtilis* encodes 17 sigma factors and *S. coelicolor*, which is also free-living, has 65. The large number of sigma factors makes possible a broader array of genetic programs, thus allowing these organisms to meet changing conditions and undergo major morphological changes upon starvation.

9-9 Sigma factors bind to specific –35 and –10 sequences that define bacterial promoters

Bacterial promoters are generally composed of two elements: one conserved sequence centered at **–35** nucleotides and a second conserved sequence centered at **–10** nucleotides from the start of the transcript, which is typically a purine nucleotide (A or G) (Figure 9-3.2). The terms –35 and –10 refer to the typical location of these sequences, although the number of base pairs separating them can vary among different promoters. As we shall see, variations in both promoter element spacing and sequence allow for differential regulation of genes by influencing the relative binding of different classes of sigma factors as well as the rates of promoter opening and transcription initiation. Many highly expressed genes have similar promoter sequences, with a consensus sequence of TTGACA at the –35 position and TATAAT at the –10 position (Figure 9-3.2). However, the –35 or the –10 sequence – or both – may diverge considerably from typical consensus sequences and few strictly match the consensus sequence. Some bacterial promoters contain sequences in addition to the –35 and –10 regions that can further modulate RNA polymerase binding and initiation. At some promoters lacking a conserved –35 sequence, initiation depends on recognition of an **extended –10 region** consisting of an additional two bases located upstream of the –10

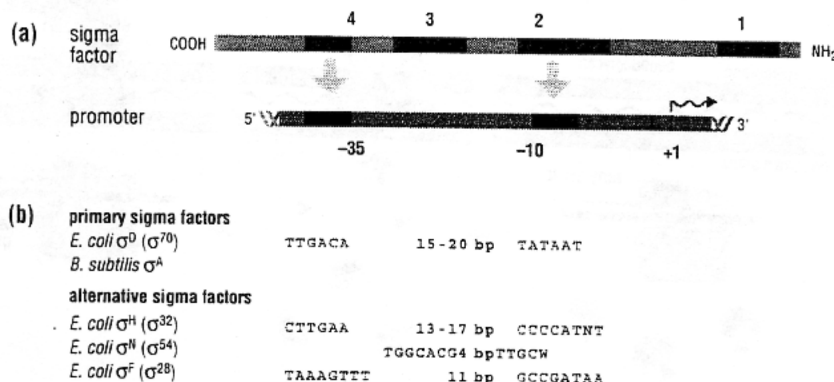

Figure 9-3.2 Structure of bacterial consensus promoters and regions of sigma subunits that contact DNA (a) The four major regions of sigma conservation are indicated on the diagram of a sigma factor. Regions of contact between the sigma factor and the promoter are indicated by arrows. **(b)** The consensus sequences for promoters recognized by the major *E. coli* and *B. subtilis* sigma factors and the *E. coli* σ^H, σ^N and σ^F alternative sigma factors. N = any nucleotide, W = A or T. Sigma factors are referred to either by a letter designation or by the molecular weight in kilodaltons.

Definitions

–10: promoter sequence located 10 nucleotides upstream of the start of transcription and comprising half of the sequence directing bacterial RNA polymerase to bind and initiate transcription, also called the Pribnow box.

–35: promoter sequence located 35 nucleotides upstream of the start of transcription and comprising half of the sequence directing bacterial RNA polymerase to bind and initiate transcription.

alternative sigma factor: sigma factor that recognizes promoters of specialized sets of genes and is expressed or activated in specific conditions when it associates with the core RNA polymerase enabling the specialized genes to be expressed.

anti-sigma factor: protein that binds a sigma factor and prevents it from binding to the core RNA polymerase enzyme.

extended –10 region: additional bases upstream of some –10 sequences that are bound by the σ_3 domain.

primary sigma factor: sigma factor that is present in all normal growth conditions and associates with the core RNA polymerase to bind the promoters of many highly expressed genes. Each bacterial species has only one primary sigma factor.

sigma factor: RNA polymerase subunit required for promoter recognition.

UP element: AT-rich sequence located just upstream of the –35 sequence and bound by the α subunit of RNA polymerase.

region. Some bacterial genes that are transcribed at a high rate contain an additional AT-rich sequence called the **UP element** upstream of the −35 region and which promotes tighter binding of RNA polymerase to DNA through contacts made with the α subunit.

The sigma subunit contains a poorly structured region at the N-terminus, followed by three distinct structured domains (domains 2-4), that correspond roughly to the four main regions of amino acid conservation found in sequence comparisons of diverse sigma factors (Figure 9-3.2). Within the holoenzyme, the three structured domains of sigma are positioned to recognize the promoter sequence (Figure 9-3.3). Domain 2 binds to the −10 region and is implicated in melting of the DNA to form the open complex, while the immediately adjacent domain 3 recognizes the two bases comprising the extended −10 region. Domain 4 recognizes the −35 region and is bound to a flexible flap of the β subunit that may allow movement of domain 4 to accommodate variably spaced −35 and −10 elements. The N-terminal unstructured region of sigma, termed region 1.1, inhibits DNA binding by the free sigma subunit and accelerates the rate of open complex formation at certain promoters.

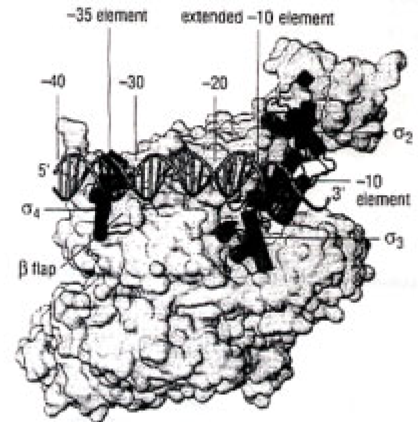

Figure 9-3.3 Structure of a bacterial RNA polymerase holoenzyme bound to a promoter The structure of the *T. aquaticus* holoenzyme shows how the three structural domains of the sigma subunit bind to the core enzyme in a position to recognize the promoter elements. The DNA is numbered relative to the transcription start site at +1. The σ_2 domain recognizes the −10 region (red), while the σ_3 domain binds to the flanking base pairs of the extended −10 region. The σ_4 domain, which binds to the −35 element (red), is anchored to a flexible flap of the β subunit that may allow movement of the σ_4 subunit to allow for different spacings between the −35 and −10 regions. (PDB 1IW7)

9-10 Sigma factor activity can be regulated in many ways

While the amounts of the core subunits of RNA polymerase remain relatively constant under different growth conditions, the levels and activities of many sigma factors can vary significantly in response to environmental or developmental signals, thereby enabling the cell to alter its pattern of gene expression to meet changing needs. This regulation of sigma factor activity can be accomplished in a variety of ways. Some sigma factors are synthesized as pro-sigma factors carrying inhibitory domains that must be cleaved off before the sigma factor can associate with the core RNA polymerase enzyme. The activities of other sigma factors are controlled by **anti-sigma factor** proteins that bind to specific sigma factors and prevent their interaction with the RNA polymerase core enzymes. The activities of anti-sigma factors, in turn, can be regulated by controlling their levels of expression in the cell or by sequestering them through interactions with other cellular proteins (called anti-anti sigma factors). As one example, the anti-sigma factor FlgM forms a complex with σ^F (also called σ^{28}), which directs transcription of genes required for completing the assembly of flagella in the bacterium *S. typhimurium*. During late stages of flagellar synthesis, FlgM is exported from the cell through the incomplete flagellar apparatus, releasing σ^F (σ^{28}) to activate transcription of the genes necessary for final assembly of the flagella (Figure 9-3.4).

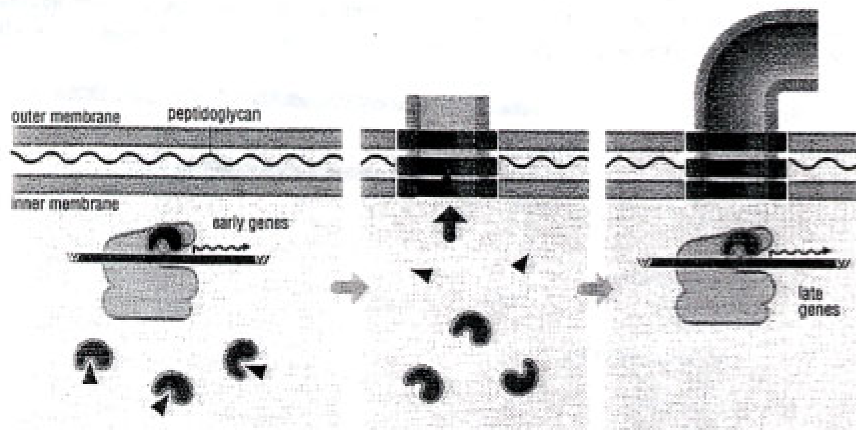

Figure 9-3.4 Regulation of *S. typhimurium* σ^F activity during flagellum biosynthesis σ^F (σ^{28}; green) is required for the expression of genes required late in the assembly of the flagellar motility motor. The genes needed for the synthesis and assembly of the initial hook-basal body complex, as well as σ^F and the anti-sigma factor FlgM (pink triangles), which inhibits σ^F, are transcribed by RNA polymerase containing the housekeeping sigma factor σ^D (σ^{70}; blue). FlgM binds σ^F and prevents the sigma factor from interacting with the RNA polymerase core enzyme. However, once the intermediate hook and basal body structure of the flagellum has been completed, FlgM, which carries late export determinants, is secreted out of the cell. The released σ^F then can interact with the RNA polymerase core enzyme and directs transcription of genes required for the completion of the flagellar structure.

References

Gross, C.A. *et al.*: **The functional and regulatory roles of sigma factors in transcription.** *Cold Spring Harb. Symp. Quant. Biol.* 1998, **63**:141-155.

Hughes, K.T. and Mathee, K.: **The anti-sigma factors.** *Annu. Rev. Microbiol.* 1998, **52**:231-286.

Moran, N.A.: **Microbial minimalism: genome reduction in bacterial pathogens.** *Cell* 2002, **108**:583-586.

Murakami, K.S. *et al.*: **Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex.** *Science* 2002, **296**:1285-1290.

Vassylyev, D.G. *et al.*: **Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution.** *Nature* 2002, **417**:712-719.

Two-component signal carriers employ a small conformational change that is driven by covalent attachment of a phosphate group

G proteins and motor ATPases are generally absent from prokaryotes, which use a different class of molecular switches. Their signaling pathways, referred to as **two-component systems**, are structured around two families of proteins. The first component is an ATP-dependent histidine protein kinase (HK). This is typically a transmembrane protein composed of a periplasmic sensor domain that detects stimuli and cytoplasmic histidine kinase domains that catalyze ATP-dependent autophosphorylation. The second component of the system, a cytoplasmic response regulator protein (RR), is activated by the histidine kinase. Signals transmitted through the membrane from the sensor domain modulate the activities of the cytoplasmic kinase domains, thus regulating the level of phosphorylation of the RR. Phosphorylation of the RR results in its activation and generation of the output response of the signaling pathway. Thus, although ATP hydrolysis is used to drive this switching process, it does so indirectly: a phosphoryl group originally derived from ATP covalently modifies the RR and serves as the trigger.

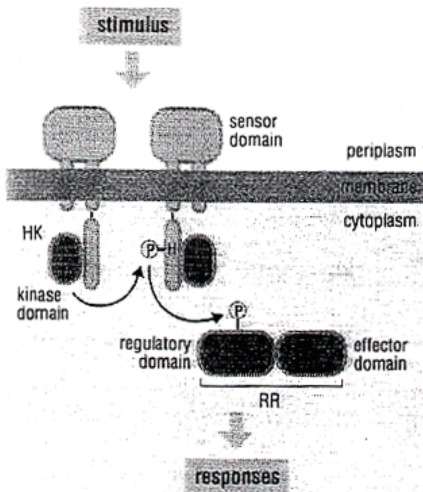

Figure 3-29 **Two-component signaling mechanisms** The basic two-component phosphotransfer system, found in bacteria, consists of a dimeric transmembrane receptor histidine kinase (HK) and a cytoplasmic response regulator (RR). Information flows between the two proteins in the form of a phosphoryl group (PO_3) that is transferred from the HK to the RR. HKs catalyze ATP-dependent autophosphorylation of a specific conserved His residue (H). The activities of HKs are modulated by environmental signals such as nutrients or osmotic stress. The phosphoryl group (P) is then transferred to a specific aspartic acid residue (D) located within the conserved regulatory domain of an RR. Phosphorylation of the RR typically activates an associated (or downstream) effector domain, which ultimately elicits a specific cellular response.

Two-component system proteins are abundant in most eubacterial genomes, in which they typically constitute ~1% of encoded proteins. For example, the *E. coli* genome encodes 62 two-component proteins, which are involved in regulating processes as diverse as chemotaxis, osmoregulation, metabolism and transport. In eukaryotes, two-component pathways constitute a very small number of all signaling systems. In fungi, they mediate environmental stress responses and hyphal development. In the slime mold *Dictyostelium* they are involved in osmoregulation and development, while in plants they are involved in responses to hormones and light, leading to changes in cell growth and differentiation. To date, two-component proteins have not been identified in animals and do not seem to be encoded by the human, fly or worm genomes. In most prokaryotic systems, the output response is generated directly by the RR; in many systems it functions as a transcription factor whose transcription-activating or repressing activity depends on its state of phosphorylation. Fungi and plants also contain RRs that function as transcription factors. In addition, eukaryotic two-component proteins are found at the beginning of signaling pathways where they interface with more conventional eukaryotic signaling strategies such as mitogen-activated protein (MAP) kinase and cyclic nucleotide cascades.

Similarly to most proteins in signaling pathways, two-component systems are modular in architecture. Different arrangements of conserved domains within proteins and different integration of proteins into pathways provide adaptations of the basic scheme to meet the specific regulatory needs of many different signaling systems. The prototypic prokaryotic two-component pathway (Figure 3-29) illustrates the fundamental phosphotransfer switching mechanism of both simple and more elaborate systems. Stimuli detected by the sensor domain of the histidine kinase regulate the kinase's activities. The kinase catalyzes ATP-dependent autophosphorylation of a specific histidine residue. The RR then catalyzes transfer of the phosphoryl group from this phosphorylated histidine to one of its own aspartate residues, located on the regulatory domain. Phosphorylation of the regulatory domain of the RR activates an effector domain on the same protein (or, on rare occasions, a separate effector protein) that produces the specific output response.

The regulatory domains of RRs have three activities. First, they interact with phosphorylated histidine kinase and catalyze transfer of a phosphoryl group to one of their own aspartate residues. Second, they are phosphatases that catalyze their own dephosphorylation—the counterpart to the GTPase activity of the G proteins. The phosphatase activity varies greatly among different RRs, with half-lives for the phosphorylated state ranging from seconds to hours, a span of four orders

Definitions

two-component systems: signal transduction systems found in bacteria and some eukaryotes involving a membrane-bound histidine kinase and a cytoplasmic response regulator protein that is activated by phosphorylation.

References

Parkinson, J.S. and Kofoid, E.C.: **Communication modules in bacterial signaling proteins.** *Annu. Rev. Genet.* 1992, **26**:71–112.

Stock A.M. *et al.*: **Two-component signal transduction.** *Annu. Rev. Biochem.* 2000, **69**:183–215.

Stock A.M. *et al.*: **Structure of the Mg^{2+}-bound form of CheY and mechanism of phosphoryl transfer in bacterial chemotaxis.** *Biochemistry* 1993, **32**:13375–13380.

West, A.H. and Stock, A.M.: **Histidine kinases and response regulator proteins in two-component signaling systems.** *Trends Biochem. Sci.* 2001, **26**:369–376.

Wurgler-Murphy, S.M. and Saito, H.: **Two-component signal transducers and MAPK cascades.** *Trends Biochem. Sci.* 1997, **22**:172–176.

©2004 New Science Press Ltd

of magnitude. And third, RRs regulate the activities of their associated effector domains (or effector proteins) in a phosphorylation-dependent manner. The different lifetimes of different regulators allow two-component signal systems to regulate a wide variety of cellular processes.

All RRs have the same general fold and share a set of conserved residues (Figure 3-30a). Phosphorylation of the active-site aspartate is associated with an altered conformation of the regulatory domain. A common mechanism appears to be involved in the structural changes that propagate from the active site. The phosphorylated aspartate is positioned by interaction of two of the phosphate oxygens with both a divalent cation (usually a magnesium ion) coordinated to three active-site carboxylate side chains and with the side chain of a conserved lysine residue. Two other highly conserved side chains have distinctly different orientations in the phosphorylated form of the regulatory domain compared to the unphosphorylated form: in the phosphorylated state a serine/threonine side chain is repositioned to form a hydrogen bond with the third phosphate oxygen and a phenylalanine/tyrosine side chain is reoriented towards the interior of the domain, filling the space that is normally occupied by the serine/threonine in the unphosphorylated protein.

Unlike the GTPases, which have large conformational changes in their switching mechanisms, the regulatory domains of two-component signaling pathways undergo less dramatic structural rearrangements on phosphorylation. Structural differences between the unphosphorylated and phosphorylated regulatory domains map to a relatively large surface involving several beta strands, alpha helices and adjacent loops, with backbone displacements ranging from 1 to 7 Å (Figure 3-30b).

In other words, like the G proteins, RR regulatory domains function as generic on–off switch modules, which can exist in two distinct structural states with phosphorylation modulating the equilibrium between them, providing a very simple and versatile mechanism for regulation. Surfaces of the regulatory domain that have altered structures in the two different conformations are exploited for protein–protein interactions that regulate effector domain function.

Not all effector domains are DNA-binding transcriptional regulators. Some function, for example, as regulators of bacterial flagellar rotation. There are many different strategies for regulation of effector domains by RR regulatory domains, including inhibition of effector domains by unphosphorylated regulatory domains, allosteric activation of effector domains by phosphorylated regulatory domains, dimerization of effector domains mediated by domain dimerization, and interaction of RRs with heterologous target effector proteins. Different RRs use different subsets of the regulatory domain surface for phosphorylation-dependent regulatory interactions. Thus, although RRs are fundamentally similar in the design of their phosphorylation-activated switch domains, there is significant versatility in the way these domains are used to regulate effector activity.

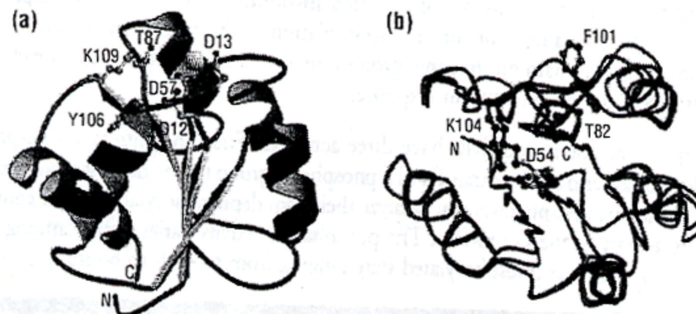

Figure 3-30 Conserved features of RR regulatory domains **(a)** A ribbon diagram of the bacterial RR CheY, whose protein fold is just a bare regulatory domain, is shown with ball-and-stick representations of the side chains of highly conserved residues. Residues that are highly conserved in all RR regulatory domains are clustered in two regions: an active site formed by loops that extend from the carboxy-terminal ends of strands 1, 3 and 5, and a pair of residues that form a diagonal path extending across the molecule from the active site. Three aspartic acid residues (D12, D13 and D57) position a Mg^{2+} ion (yellow) that is required for catalysis of phosphoryl transfer from an HK to aspartate 57. Three additional residues (K109, T87 and Y106) are important in propagation of the conformational change that occurs on phosphorylation. The regions of the backbone that have been observed to differ in unphosphorylated and phosphorylated regulatory domains are shown in magenta. **(b)** The conserved mechanism involved in the phosphorylation-induced conformational change is illustrated by the structures of the unphosphorylated (blue) and phosphorylated (magenta) regulatory domains of the RR protein FixJ. When D54 is phosphorylated, K104 forms an ion pair with the phosphate (yellow). T82 also forms a hydrogen bond with the phosphate and F101 has an inward orientation, positioned in the space occupied by T82 in the unphosphorylated structure. These changes trigger other rearrangements of secondary structure elements, creating a different protein-binding surface in the two states. Graphic kindly provided by Ann Stock. From West, A.H. and Stock, A.M.: *Trends Biochem. Sci.* 2001, **26**:369–376, with permission.

©2004 New Science Press Ltd