2. Introduction

2.1 Operation and regulation of the bacterial transcriptional machinery

Transcription involves the conveyance of genetic information from DNA to RNA. RNA polymerase (RNAP), the enzyme catalysing this process, uses one strand of the DNA duplex (template strand) as prototype for synthesizing a new RNA chain with identical sequence to the second DNA strand (coding strand). The new RNA chain can either function as further template for protein synthesis (mRNA), or serve its role as it is (rRNA, tRNA and sRNA; although some sort of procession/chemical modification usually follows RNA synthesis in all these three classes). Since transcription is the first stage in gene expression, it is energetically favourable for organisms to use it as the main way of regulating and reprogramming gene expression. In the next pages, a more analytical account regarding bacterial transcription and its regulation will follow; many times the focus will be on *Escherichia coli*, as most of the pioneering work on such issues has been done with this organism.

2.1.1 Composition and structure of RNA polymerase

A. Core RNAP

Bacterial RNAP is a multi-subunit enzyme with a ~400KDa catalytic core part ($\alpha_2\beta\beta'\omega$) that can only initiate promoter-specific transcription after associating with a σ factor (see also next chapter for sigma factors). The first high-resolution structural study on *Thermus aquaticus* RNAP core enzyme (Zhang et al., 1999) revealed a "crab-claw" shaped molecule, very similar to that of the yeast RNAP II (Cramer et al., 2001; Fu et al., 1999), with **β and β′ subunits** interacting extensively and forming the two pincers (Fig. 1A; names commonly used for several domains of the two subunits are noted in the figure). The active centre lies within a 27 Å wide internal channel/cleft formed by the two pincers (also called "jaws"), and is marked by a Mg^{2+} ion chelated at the base of the channel by three aspartate residues (NA**D**F**D**G**D** motif of β′ subunit). This main channel of RNAP is mostly positively charged, whereas negative residues predominantly cover the outside surface of the core enzyme. A secondary channel, 10-12 Å wide, also links the external milieu with the active site. Being too narrow for double stranded DNA to pass through, its role has been associated with that of an entrance gate for substrate nucleotides NTPs to the catalytical centre. Recent studies have highlighted how regulatory proteins alone (GreA/B), or together with small molecules (DksA and ppGpp), and peptide antibiotics (Microcin J25) can use the secondary channel for

reaching the RNAP active site and modifying its enzymatic properties (for review see Nickels and Hochschild, 2004).

Exposed in the surface of the RNAP molecule, opposite from the major cleft, are located the N-terminal domains (NTD) of the two identical **α subunits**, each of them contacting β and β′ subunit exclusively (Fig. 1A). The formation of the αNTD dimer marks the initial step of RNAP assembly (Ishihama et al., 1987). The C-terminal domain of α subunits, known to be binding the promoter's UP-element and interacting with adjacent transcriptional regulators, and the flexible linker that connects it to the NTD domain, are both disordered in all available RNAP structures (Murakami et al., 2002a; Murakami et al., 2002b; Vassylyev et al., 2002; Zhang et al., 1999). Nevertheless, the αCTD structure was resolved in other studies (Benoff et al., 2002; Jeon et al., 1995), whereas the linker connecting the two domains of α is thought to be unstructured and very flexible (Blatter et al., 1994; Husnain et al., 2004).

The **ω subunit** wraps around the carboxy-terminal tail of the β′ subunit (Fig. 1A), consistently with its proposed function to assist β′ subunit's folding and promote RNAP assembly (Minakhin et al., 2001). Recent studies have brought up into discussion a long-time discovered role of the ω subunit in sensitising the response of RNAP to ppGpp (Igarashi et al., 1989; Vrentas et al., 2005) and probably to negative supercoiling too (Travers and Muskhelishvili, 2005b).

The first modelling of nucleic acids into the core RNAP structure offered a more refined picture of the functional context of RNAP (Fig. 1B; Korzheva et al., 2000). Double-stranded downstream DNA enters the main channel formed by the pincers and separates in its two strands near the active site, only to reanneal again and form the upstream duplex at a 90° right angle to the downstream DNA. Nascent RNA remains as part of an RNA/DNA hybrid for 8-9 nucleotides before the β′ rudder (a "figure-8" shaped loop) separates it from the template DNA strand and allows it to exit the RNAP through the RNA-exit channel, located between the base of the β flap and the β′ lid; the length of nascent RNA by its exit is 14nt.

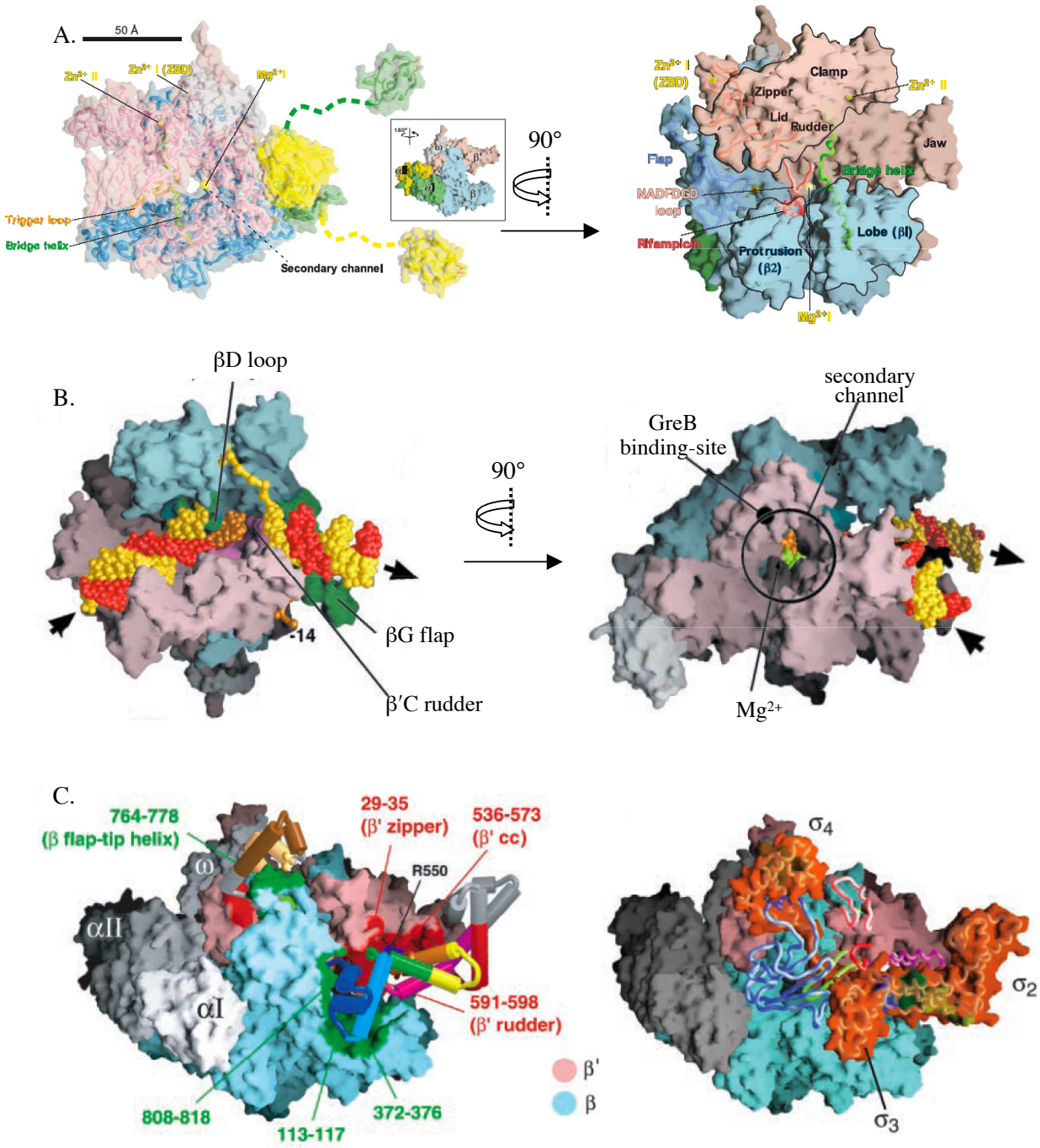


Fig 1: Structures of core RNAP (**A**) and RNAP holoenzyme (**C**), plus a TEC model (**B**), based on them (reproductions from Geszvain and Landick, 2004; Korzheva et al., 2000; Murakami et al., 2002a).

A. The downstream face of core RNAP is depicted on the left panel. The model is based on the coordinates of the *T. thermophilus* holoenzyme (PDB ID 1IW7; Vassylyev et al., 2002), with the σ subunit and a non-conserved region, not present in *E. coli* β′ (aa 164-448), removed and the RNAP conformation adjusted to that observed in the core *T. aquaticus* RNAP (PDB ID 1I6V; Zhang et al., 1999), by movement of RNAP mobile modules. A dashed line illustrates the path of the secondary channel. The α-carbon backbone is shown as a worm inside a semi-transparent surface. Subunits are colour-coded as follows: β′, pink, β, cyan, αI, green, αII, yellow, ω, grey. The β′ bridge helix and trigger loops are depicted as green and orange worms, respectively. Zn^{2+} and Mg^{2+} atoms are depicted as yellow balls. The αCTDs are shown in arbitrary positions 43 Å from core. They are shown as isolated domains, but may be present as a dimer in RNAP (Blatter et al., 1994). The boxed inset depicts the upstream face of core RNAP, illustrating the "crab-claw" shape of the enzyme. On the right panel, emphasis is drawn in the active-site channel. The initial RNAP model is shown rotated 90°

to the right. The subunits are shown as solid surfaces, except for the β′ Mg^{2+}-binding loop, ZBD, rudder, lid and zipper, which are shown as pink worms in a semi-transparent surface, and the β flap domain, which is shown as a dark blue worm in a semi-transparent surface. The β′ bridge helix is depicted as a green worm. The antibiotic rifampicin is depicted in red (β is rendered semi-transparent in front of Rif to reveal the antibiotic nestled in its binding pocket). The clamp, protrusion and lobe are outlined in black.
B. The TEC model consists of three components: (i) the *Taq* core RNAP crystal structure (Zhang et al., 1999), shown as colour-coded molecular surfaces (β, cyan; β′, pink; α and ω, white; catalytic Mg^{2+}, magenta sphere); (ii) the DNA template (template strand, red; nontemplate strand, yellow); and (iii) the RNA transcript (gold) plus incoming nucleotide substrate (green). The nucleic acid backbones are shown as worms. The directions of the entering downstream DNA and exiting upstream DNA are indicated (large arrows). On the left panel, a view of the TEC, perpendicular to the main active-site channel that runs roughly horizontal, is presented. βD loop I and βG flap are coloured in green, whereas β′C rudder is in rose. On the right panel, a view down the secondary channel is presented, showing the path for diffusion of the incoming nucleotide substrate (green) into the active site. The catalytic Mg^{2+} ion (magenta sphere) is just visible to the left of the substrate nucleotide.
C. The core component of the holoenzyme (left panel) is shown as a molecular surface, colour coded as follows: αI, αII, ω: grey; β: cyan; β′: pink. The σ subunit is shown as an α-carbon backbone worm on the left view of RNAP holoenzyme, with α helices shown as cylinders, colour coded as follows: region 1.2: red; 2.1: fuchsia; 2.2: orange; 2.3: yellow; 2.4: green; 3.0: light blue; 3.1: dark blue; 3.2: olive green; 4.1: tan; and 4.2: brown. Surfaces of RNAP within 4 Å of any σ atoms are colour-coded green (β) or red (β′) and labelled. Positions in σ region 2.2 (orange backbone) and 4.1 (tan backbone), where substitutions cause defects in core RNAP binding, are indicated by orange or tan α-carbon spheres, respectively. The exposed surface of β′Arg550 on the β′ coiled coil, important for β′interaction, is coloured blue and labelled. On the right panel, a similar view of the *Taq* RNAP holoenzyme structure, shown as a molecular surface, is provided, but this time important features of core RNAP are shown as α-carbon backbone worms without the corresponding surfaces (colour coding as the previous picture). In addition, the molecular surface of σ is transparent, allowing the orange α-carbon backbone worm to be seen as well. The Zn^{2+} ion bound in the β′ ZBD is shown as a light-green sphere. Surfaces of σ corresponding to residues important for promoter recognition and melting are colour coded as follows: melting/-10 element non-template strand binding, yellow; -10 element recognition, green; extended -10 element recognition, blue; -35 element recognition, brown.

B. RNAP Holoenzyme

Promoter-specific transcriptional initiation starts only upon formation of RNAP holoenzyme; this requires association of a single protein (sigma factor) in bacteria, whereas eukaryotes have to employ more than a dozen initiation factors (mounting a 750KDa assembly) for the same process (Hahn, 2004). The two bacterial holoenzyme crystal structures derive both from thermophilic organisms (Murakami et al., 2002b; Vassylyev et al., 2002; Fig. 1C comes from the former study) and reveal an extensive network of interactions between the housekeeping sigma factor and core RNAP (for review see Borukhov and Nudler, 2003), as previously predicted by various genetic or biochemical studies (Gruber et al., 2001; Owens et al., 1998; Sharp et al., 1999). Except for a strong contact between σ_2 (see next chapter for more details concerning conserved regions of sigma factors and their roles) and the coiled coil region of the β′ clamp, most of the other mapped contacts are relative weak, sometimes even destabilising, and dispersed over a large surface area. This observation fits to earlier findings

that had "depicted" $\sigma_{2.2}$-β′ coiled coil interaction as the primary interface between sigma and core RNAP (Arthur et al., 2000; Arthur and Burgess, 1998; Katayama et al., 2000). Several other contacts, including that of i) σ_3 with the β1 lobe, ii) σ_4 with the tip-helix of the β flap domain, iii) the linker between σ_2 and σ_3 with the β′ zipper and that of iv) the ~ 30 residue long flexible linker between σ_3 and σ_4 (33 amino acid long in *Thermus aquaticus* and 30 amino acid long in *Thermus thermophilus*) with several compartments of the RNA exit tunnel (β′ rudder, zipper and lid), support the notion that binding of sigma to core RNAP is a multi-step and cooperative process (Gruber et al., 2001).

In general, the conversion from core to holoenzyme is accompanied by various conformational changes in all subunits of RNAP. Most surprisingly of all, the pincers appear to close in the holoenyme form (by 10Å). Since σ_2 interacts with the larger pincer (β′ clamp) and σ_3 with the smaller (β1 lobe), it seems possible that σ can control closing and opening of the channel during the several steps of promoter recognition. Strong reorganisations are also apparent in the RNA exit channel, not only because the linker connecting σ_3 and σ_4 passes through it, but also because the β flap domain is shifted by 5-6 Å towards σ_4, narrowing the diameter of the channel from 19 to 11 Å. The N-terminal domains of β and αI seem to move in concert with the β flap, whereas the β flap tip helix seems to suffer an even more dramatic change in its orientation, as it is shifted 11 Å from its position in core. As a result, this whole region of the holoenzyme molecule appears to be strained and therefore its relaxation contributes to the first step of sigma release from the holoenzyme (Borukhov and Nudler, 2003; Nickels et al., 2005). Moreover, it is noteworthy that several conserved structural elements that appear disordered in the core RNAP structure (Zhang et al., 1999) seem to be stabilised and therefore, resolved in the holoenzyme structures (Murakami et al., 2002b; Vassylyev et al., 2002).

One of the most interesting regions missing in all structures available is region 1.1 of sigma. A FRET-based analysis of RNAP holoenzyme located $\sigma_{1.1}$ inside the main channel of RNAP, where its negative charge can interact with the basic surface of the channel walls (Mekler et al., 2002). Murakami *et al* (2002b) reached a similar conclusion based on a series of indirect structural data, supported by the fact that $\sigma_{1.1}$ is only exposed to hydroxyl radicals in the free σ^{70} form and the bound to DNA holoenzyme form ($E\sigma^{70}$); on the contrary, when $E\sigma^{70}$ is not bound to DNA, $\sigma_{1.1}$ is completely protected (Nagai and Shimamoto, 1997). The same authors proposed that $\sigma_{1.1}$ is presumably ejected from the main channel upon isomerisation of the holonzyme to an open complex formation (Mekler et al., 2002; Murakami et al., 2002b; see also Fig. 5C and the corresponding text in the transcriptional

initiation chapter).

2.1.2 Sigma factors

A. Classification and function of the different families/groups

Eubacteria contain a number of different sigma factors, ranging from a single one in *Mycoplasma genitalium* to more than 65 in *Streptomyces coelicolor*. A rough correlation seems to exist between the number of sigmas a bacterium encodes and the environmental complexity that is involved in its lifestyle. Most of our knowledge concerning sigma factors derives from model organisms like *Escherichia coli* and *Bacillus subtilis* that encode 7 and 19 sigmas respectively. Based on sequence comparisons bacterial sigma factors are classified in two large families. The broader σ^{70} family (reviewed in Gruber and Gross, 2003; Paget and Helmann, 2003) is named after the housekeeping σof *E. coli*, and contains all vegetative sigmas and most of the alternative ones. The structurally and functionally distinct σ^{54} family- the superscript reflects the molecular weight of the homonymous *E. coli* sigma factor- rarely has more than one representative in each bacterial species (reviewed in Studholme and Buck, 2000).

The σ^{70} family has been divided into five different groups on the basis of phylogenetic relatedness (a representative phylogenetic tree can be seen in Fig. 2). **Group 1** includes σ^{70} and its orthologue housekeeping sigma factors of the different bacterial species. Housekeeping σs are indispensable for the cell, present during the entire growth cycle, and responsible for nearly all gene expression during vegetative growth (for review see Gross et al., 1998). **Group 2** members are present only in some species and are very closely related to housekeeping sigmas, however they are dispensable for growth. The master regulator for stress responses in *E. coli* and focus of this thesis, σ^{S} (RpoS), is the most well-known representative of the group (see chapter 1.2 and for review (Hengge-Aronis, 2000). **Group 3** σs are more divergent in sequence and function than the first two groups. Members of this group include sigmas involved in general stress response (for example σ^{B} in *B. subtilis*; Hecker and Volker, 2001; van Schaik and Abee, 2005), flagellar biosynthesis (e.g. σ^{28} in *E. coli* and *B. subtilis*; Aldridge and Hughes, 2002; Chilcott and Hughes, 2000), heat shock response (featuring σ^{32} in *E. coli* as its most well-studied member; Arsene et al., 2000; Nonaka et al., 2006; Yura et al., 2000), and sporulation (σ^{E}, σ^{F}, σ^{G}, σ^{H} and σ^{K} in *B. subtilis*; Eichenberger et al., 2004; Errington, 2003; Sonenshein, 2000). **Group 4** σs have only basic sequence similarity to σ^{70}. Since most of the group 4 sigmas are co-transcribed with a membrane bound anti-sigma factor and control genes involved in extracytoplasmic functions,

they are also known as ECF sigmas. *E. coli* has only two members belonging in this group, σ^{E} regulating envelope stress (Alba and Gross, 2004; Raivio, 2005; Ruiz and Silhavy, 2005) and σ^{FecI} securing iron transport (Braun et al., 2003; Visca et al., 2002). Other microorganisms with more complex lifestyles, such as *Bacillus subitlis*, *Caulobacter crescentus*, *Pseudomonas aeruginosa* and *Streptomyces coelicolor*, possess many more: 7, 13, 19 and 49 respectively. A number of recently discovered sigma factors in several *Clostridium* species, responsible for the production of toxins and bacteriocins (Dupuy et al., 2005; Dupuy et al., 2006; Raffestin et al., 2005), seem to be even further distantly related to σ^{70}, and therefore were proposed to form a novel class of their own, termed as **Group 5** (Helmann, 2002).

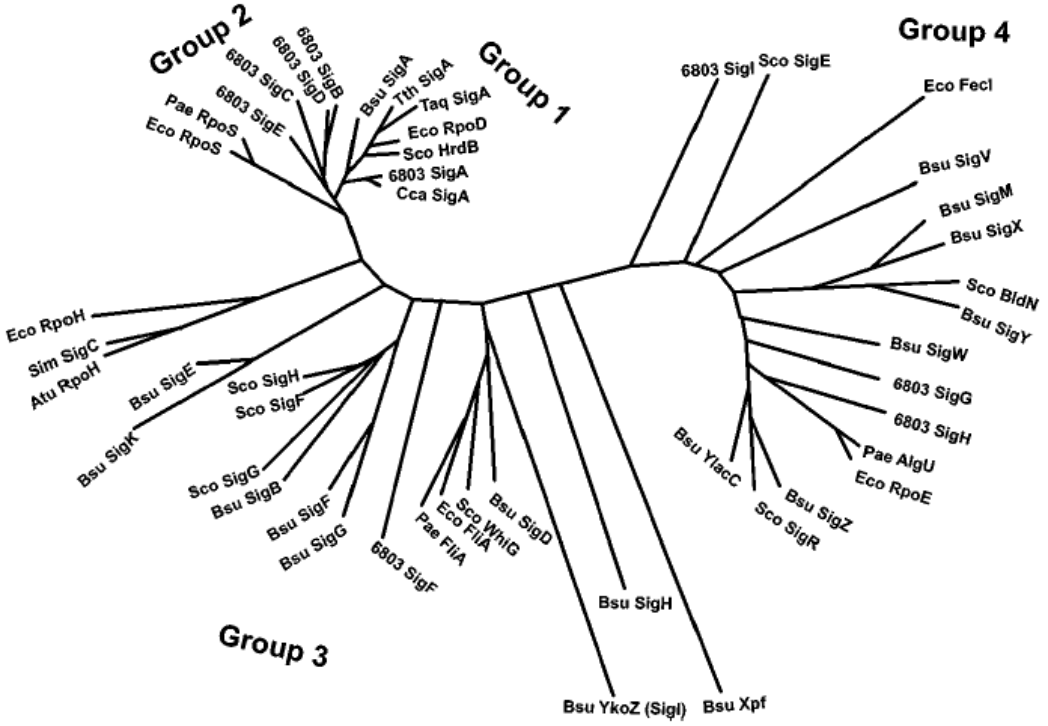


Fig. 2: The diversity of σ^{70}-family sigma factors. The Fitch-Margoliash phylogenetic tree (reproduction from Gruber and Gross, 2003) was generated for multiple sequence alignments spanning σ regions 1.2 to 2.4. The four different σ^{70} groups are indicated and sigma factors derive from various organisms (Eco, *E. coli*; Bsu, *B. subtilis*; 6803, *Synechocystis* sp. PCC 6803; Sco, *S. coelicolor*; Pae, *P. aeruginosa*; Tth, *T. thermophilus*; Taq, *T. aquaticus*; Cca, *Cyanidium caldarium* (chloroplast σ); Sim, *Sinorhizobium meliloti*; Atu, *Agrobacterium tumefaciens*).

The σ^{54} family shows no sequence similarity to any of the σ^{70} family members, and follows a distinct pathway for open complex formation. RNAP containing σ^{54} ($E\sigma^{54}$) requires the presence of enhancer proteins and ATP to initiate transcription (reviewed in Buck et al., 2000; Studholme and Dixon, 2003). Interestingly, some bacterial families, such as Gram-positive bacteria with high G/C content and cyanobacteria, do not seem to encode σ^{54}–homologues in their genomes (Studholme and Buck, 2000). Nevertheless when present,

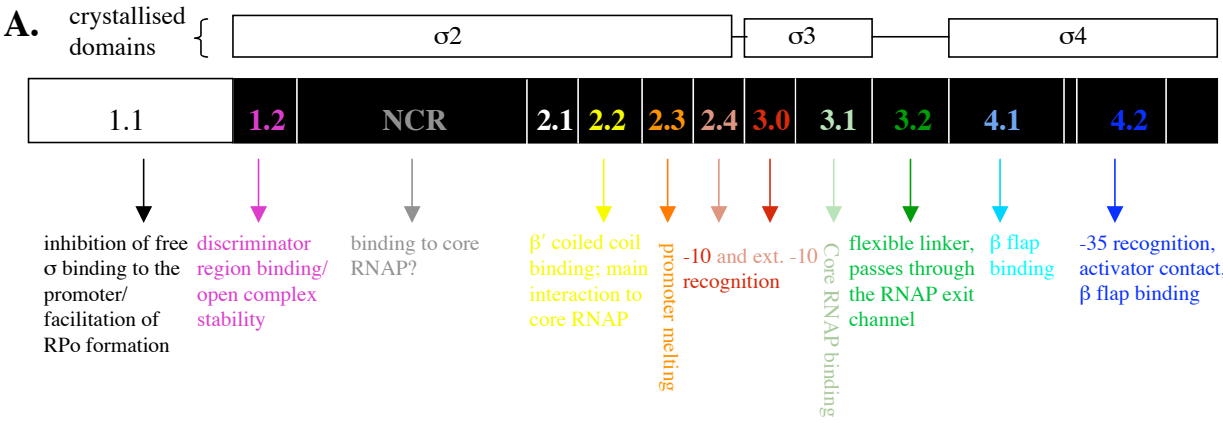
σ^{54} is responsible for expression of genes involved in various cellular processes. Nowadays, we are aware that the role of σ^{54} goes far beyond the mere control of nitrogen assimilation genes (Hunt and Magasanik, 1985; Kustu et al., 1989); genes associated with carbon metabolism (Cases et al., 2003), motility (Cases et al., 2003; McCarter, 2004; Millikan and Ruby, 2003), phage shock response (Lloyd et al., 2004; Weiner et al., 1995), developmental processes (Jelsbak et al., 2005) and virulence (Kazmierczak et al., 2005) are included in the σ^{54} (σ^{N}/RpoN) regulon in different bacteria. Moreover, σ^{54} controls the expression of σ^{s} in *Borrelia burgdorferi*, merging thus the two major alternative sigma factor-mediated responses into the same pathway (Hubner et al., 2001).

It is worth mentioning that sigma factors can be found also in plastids of photosynthetic eukaryotes, such as plants and algae (reviewed in Kanamaru and Tanaka, 2004; Shiina et al., 2005). Chloroplasts possess both a nuclear-encoded, phage-like, single subunit RNA polymerase (NEP) and a plastid-encoded, multi-subunit, eubacterial-like RNAP (PEP) (both reviewed in Hess and Borner, 1999). The multiple σs are nuclear-encoded, and therefore a targeting sequence in their N-terminal domain is required to allow them to be successfully transported to the chloroplasts (Isono et al., 1997). Their functions are diverse, usually associated with developmental (Kanamaru and Tanaka, 2004) or photosynthetic processes (Tsunoyama et al., 2004), and quite often tissue-specific (Isono et al., 1997; Lahiri et al., 1999; Tan and Troxler, 1999). Since plastids are believed to have a cyanobacterial ancestor (deriving from an endosymbiotic event; (Douglas and Raven, 2003; Douglas, 1998), it is not surprising that their sigma factors are closely associated to σ^{70} family members of cyanobacteria (Gruber and Gross, 2003). In contrast to that, mitochondria, which are also believed to be the end product of endosymbiosis between bacteria and eukaryotic hosts, are not known to utilise any bacterial-like RNAP. Nevertheless, the primitive protozoan, *Reclinomonas americana*, was found to encode a bacterial-like RNAP in its mitchondrial chromosome (Lang et al., 1997), suggesting that the mitochondria of this particular organism might be using a similar dual transcription mechanism like chloroplasts do.

B. Structural and functional domains of the σ^{70} family sigma factors

The multiple members of the σ^{70} family are modular proteins, consisting of up to four regions of sequence conservation, each divided into different sub-regions (Fig. 3A). Although a high-resolution structure of an intact sigma factor is missing, the current knowledge of crystallised portions from *T. aquaticus* σ^{A} (Campbell et al., 2002b) combined with the holoenzyme crystal structures (Murakami et al., 2002b; Vassylyev et al., 2002) has provided us with a picture of

those modular regions of sigma. Both σ^A of *T. aquaticus* (Fig. 3B) and *T. thermophilus* show three flexibly linked domains: σ_2 (regions 1.2-2.4), σ_3 (regions 3.0-3.1) and σ_4 (regions 4.1-4.2); as mentioned above region 1.1 is not visible in any of the structures. Note that the domain numbering used for σ here (also in Fig. 3) is the same as in Murakami *et al* (2002b); Vassylyev *et al* (2002) adopt a slightly different numbering, counting the linker between region 3 and 4 as a separate region of its own.



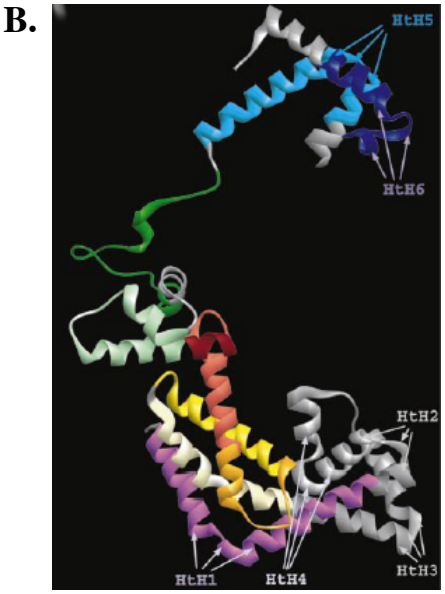


Fig. 3: Organisation and structure of a housekeeping σ. **A.** Schematic representation of the structural domains (above) and conserved regions (below) of a typical group 1 member of the σ^{70}-family of sigma factors. Each of the four conserved regions is divided into sub-regions, which are depicted in different colour and their functional assignments are based on σ^{70} of *E. coli*. **B.** The structure of the *T. aquaticus* σ^A presented as a ribbon diagram (adapted from (Vassylyev et al., 2002). The colour coding for the different sub-regions is the same as that used in the linear representation in A.

Domain 2 contains several of the most highly conserved regions in the σ^{70} family of σs, involved both in interactions to core RNAP and to the DNA promoter elements. Apart from the conserved regions 1.2, 2.1, 2.2, 2.3 and 2.4 (seen in Fig. 3B for *T. thermophilus*), domain 2 has also a non-conserved region between 1.2 and 2.1 (noted as NCR in Fig. 3A and shown as grey in Fig. 3B), which mostly appears in housekeeping σs, and significantly varies in its length and sequence. In both structures visible (Campbell et al., 2002b; Vassylyev et al., 2002), this region forms a bundle of different sized α-helixes, which might be modulating the binding of sigma to core RNAP (Vassylyev et al., 2002). Although region 2 has an extensive

interface with core RNAP, a polar exposed surface in the α-helix of region 2.2 is the one that provides the main anchor of sigma to core RNAP (its interaction partner is the β′ coiled coil; Arthur and Burgess, 1998), whereas the following long α-helix, comprising regions 2.3 and 2.4, carries the residues responsible for recognition and melting of the -10 promoter element. In detail, region 2.4 seems to mediate all sequence specific interactions to the -10 hexamer (Daniels et al., 1990; Kenney et al., 1989; Siegele et al., 1989; Waldburger et al., 1990; Zuber et al., 1989), whereas region 2.3 uses its conserved aromatic residues for promoter melting (deHaseth and Helmann, 1995; Jones et al., 1992; Juang and Helmann, 1994; Panaghie et al., 2000) and the basic ones for non-specific sequence interactions with the –10 promoter element (Tomsic et al., 2001). Furthermore, region $\sigma_{1.2}$ seems to directly contact the non-template strand at a position directly downstream of the –10 hexamer (in the discriminator region), affecting thus the open complex formation stability in rRNA promoters (Haugen et al., 2006).

Consistently with the sequence conservation of σ_2 among housekeeping sigmas, a crystallised fragment of region 2 of *E. coli* σ^{70} (Malhotra et al., 1996) is practically superimposable to the corresponding region of σ^A of *T. aquaticus* and *T. thermophilus* (Campbell et al., 2002b; Vassylyev et al., 2002). More intriguingly the region 2 structures of the far-related, group 4 belonging sigmas, σ^R of *S. coelicolor* (Li et al., 2002) and σ^E of *E. coli* (Campbell et al., 2003), seem to be almost identical to that of the group 1 σs (Campbell et al., 2002b; Vassylyev et al., 2002).

In contrast to domain 2, **domain 3** (regions 3.0 and 3.1) is less conserved among σ^{70} family members and is composed by a three α-helical bundle (Fig 3.B). The first of them is responsible for contacting the extended –10 promoter element, as identified by earlier genetic studies (Barne et al., 1997; Becker and Hengge-Aronis, 2001). The region was named 2.5 at that point, but since it was later discovered to be part of domain 3 in the RNAP holoenzyme structures, it was renamed to region 3.0 (Murakami et al., 2002b). Interestingly, all group 4 σs lack domain 3 (Gruber and Gross, 2003), and, instead of it, carry a short unstructured region similar to that connecting domains 3 and 4 in housekeeping σs (Campbell et al., 2003). This has as a straightforward consequence that group 4 σs are unable to recognise extended -10 promoter elements (Rhodius et al., 2006).

Domain 4 (regions 4.1 and 4.2) is composed by two Helix-turn-Helix (HtH) motifs, and binds both to DNA recognition determinants and to core RNAP. The crystal structure of σ_4-DNA complexes (Campbell et al., 2002b) confirmed earlier studies depicting the HtH motif of region 4.2 as responsible for recognition of the –35 promoter element (Gardella et al.,

1989; Kenney and Moran, 1991; Siebenlist et al., 1980; Siegele et al., 1989), and at the same time provided us with an extremely detailed picture of the individual σ_4-DNA contacts. In addition, domain 4, and especially its sub-region 4.1, was shown to form a hydrophobic pocket that the flap domain of the β subunit can latch onto, and form extensive contacts (Campbell et al., 2002b; Geszvain et al., 2004; Murakami et al., 2002b). This RNAP-σ_4 interaction site appears to be more pronounced in alternative sigma factors like σ^S of *E. coli* (Kuznedelov et al., 2002b). Interestingly, the same alternative σs exhibit weaker σ_2-β′ coiled coil interactions than σ^{70} (Raffaelle et al., 2005). Moreover, a short fifth helix at the CTD of region 4, appearing ordered only in some high-resolution structures of housekeeping σs (Lambert et al., 2004; Vassylyev et al., 2002), might be also involved in the σ_4-β flap interaction (Vassylyev et al., 2002). Finally, the last part of the second helix of the $\sigma_{4.2}$ HtH motif constitutes the upstream face of the RNAP holoenzyme, which is adequately positioned to interact with regulatory proteins (Campbell et al., 2002b) and/or the αCTD (Chen et al., 2003; Ross et al., 2003).

The 30-35 residue-long **flexible linker** (region 3.2) connects the globular domains 3 and 4. In the holoenzyme structures, it passes near the active site of RNAP and through the RNA exit channel, separating regions 2.4 and 4.2 in a way that RNAP can simultaneously reach the –10 and –35 promoter elements (Murakami et al., 2002a; Murakami et al., 2002b; Vassylyev et al., 2002). Its implications in promoter escape and transcriptional elongation are numerous (see later in the corresponding chapters for more details). Although group 4 sigmas, like σ^E of *E. coli*, do not show any sequence conservation in this region, the unstructured linker proceeding σ^E_2 could adequately replace the region 3.2-linker of housekeeping sigmas, when the σ^E structure was modelled over that of the RNAP holoenzyme (Campbell et al., 2003).

Finally, the non-conserved, and structurally unresolved, **region 1.1** of housekeeping sigmas has been proposed in the past to directly inhibit free sigma from recognising DNA promoter elements, by binding and occluding σ_4 (Bowers and Dombroski, 1999; Dombroski et al., 1993; Dombroski et al., 1992). However, more recent studies, have ruled out such a mechanistic scenario (Camarero et al., 2002; Geszvain et al., 2004), adding more mystery to the real function of $\sigma_{1.1}$ in the free sigma form. Its possible roles in holoenzyme formation and transcriptional initiation are discussed in the corresponding chapters.

C. *Bacteria globally change their pattern of gene expression by regulating the switch of σ in the RNAP assembly*

Upon drastic environmental changes bacteria have to adjust their gene expression and thus cope with the newly emerging conditions. A rapid and efficient way to succeed that is to switch the dissociable promoter-specific subunit of RNAP, i.e. the sigma factor, and thereby redirect the holoenzyme to a discrete new set of genes. The highly abundant housekeeping σ targets RNAP to most promoters of the cell, leading to expression of numerous genes that support vegetative growth. All other σs administrate more specific cellular responses, associated with external stress conditions, reception of intercellular signalling, nutrient/ion deprivation, and commitment to developmental programmes. Either uncontrolled expression of alternative σs during normal growth (for example see Bahl et al., 1987; Gehring et al., 2001), or insufficient/delayed expression of alternative σs under stress or during a developmental process (for example see King et al., 2004; Zupancic et al., 2001), can turn out to be deleterious for the cell. Therefore, the cell tightly regulates alternative σs expression, and restricts it to appropriate conditions.

The relative availability of an alternative sigma factor can be determined by numerous regulatory strategies, involving that of transcriptional regulation, control of protein synthesis and degradation rates, protein modifications that interconvert σ between an inactive and an active form, and sequestration of σ by a cognate anti-sigma factor. Especially the anti-sigma factor strategy is widespread in bacteria, even though the σ/anti-σ contacts (Campbell et al., 2002a; Campbell et al., 2003; Sorenson et al., 2004), and the release mechanism of sigma are highly diverse (reviewed in (Hughes and Mathee, 1998) and for more updated data concerning specific cases see also (Alba and Gross, 2004; Aldridge and Hughes, 2002; Yudkin and Clarkson, 2005).

However when the cellular pools of the different sigma factors (housekeeping and alternative) are taken into account, it becomes apparent that the cell takes additional steps to secure that increases or decreases in the relative availability of alternative σs end up in analogous changes of gene expression. Firstly, all current models regarding microbial gene regulation (mostly deriving from studies in *E. coli*) agree that amounts of RNAP are limiting for transcription (Ishihama, 2000). Moreover, the housekeeping sigma factor is present during the entire cell's lifecycle in relative high amounts; amounts that are close or even exceeding those of the core RNAP, and are higher than those of any other alternative σ (Grigorova et al., 2006; Jishage et al., 1996). In addition, σ^{70} exhibits the highest affinity to core RNAP among *E. coli* sigmas (Maeda et al., 2000). It is, thus, obvious that alternative σs have to compete under unfavourable conditions with the housekeeping σ, for their portion of limited amounts of core RNAP. In order to stand a fair chance in this competition and, thereby, mediate a

"holoenzyme switch", the cell has to orchestrate a reduction in the efficiency of the vegetative σ, apart from the increase in the cellular pool of the alternative σ in demand.

The alarmone ppGpp triggers recruitment of σ^N, σ^S, and σ^H to RNAP in *E. coli* by reducing the effectiveness of the dominating housekeeping σ^{70} (Jishage et al., 2002; Laurie et al., 2003; Magnusson et al., 2003). DksA was recently shown to act synergistically with ppGpp (Paul et al., 2004; Paul et al., 2005; Perederina et al., 2004). Since rRNA transcription employs 70% of $E\sigma^{70}$ upon exponential growth (Raffaelle et al., 2005), factors like DksA and ppGpp, which actively dissociate $E\sigma^{70}$ from rRNA loci upon entry into stationary phase, provide alternative σs with free core RNAP (Bernardo et al., 2006). Furthermore, overproduction of Rsd, a stationary phase induced protein (Jishage and Ishihama, 1998) with affinity for σ^{70} and core RNAP (Ilag et al., 2004), has similar effects to ppGpp with respect to "holoenzyme switching" (Jishage et al., 2002; Laurie et al., 2003). Finally, 6S RNA, a conserved small RNA (Barrick et al., 2005; Trotochaud and Wassarman, 2005), is active in stationary phase (Kim and Lee, 2004) and structurally mimics an open promoter complex that only fools $E\sigma^{70}$ to recognise it (Wassarman and Storz, 2000). Its presence decreases $E\sigma^{70}$ activity, thereby allowing alternative RNAPs to assume their role (Trotochaud and Wassarman, 2004; Wassarman and Storz, 2000). In summary, a variety of regulatory factors ensure a decrease in σ^{70} effectiveness upon conditions in which alternative sigma factors have to assume their role.

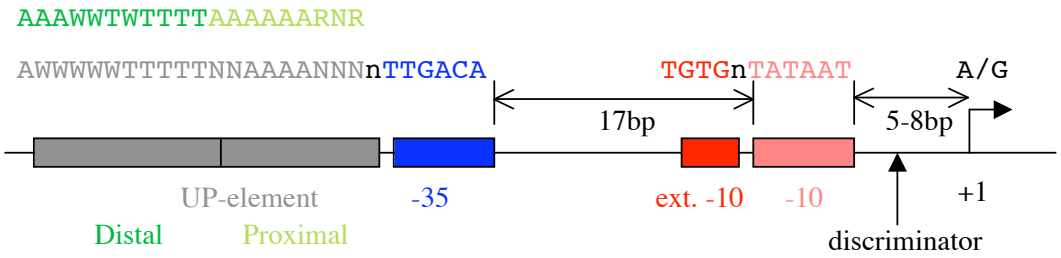
2.1.3 The bacterial transcriptional cycle

The transcriptional process can be divided into three distinct stages: initiation, elongation and termination. Although, transcription factors usually regulate the transcriptional initiation (see next section and for review (Browning and Busby, 2004), subsequent steps can also be influenced for regulatory purposes (for reviews see Borukhov et al., 2005; Mooney et al., 2005; Nickels and Hochschild, 2004; Nudler and Gottesman, 2002; Stulke, 2002; Tucker and Breaker, 2005).

A. Transcriptional initiation

Dissociable sigma factors bind core RNAP to form a holoenzyme (R), which is then able to recognise and bind its cognate promoters (P) in order to form the closed complex (RP_C); closed refers to the fact that the bound DNA duplex is still at this point unwound. This complex undergoes a series of structural intermediates/transitions, known as isomerisation steps (RP_I), before it ends up forming the transcription-competent open complex (RP_O). At

this point the DNA duplex is melted over a region spanning the transcriptional start, whereas nucleation is a prior intermediate stage in which strand separation is restricted to the –10 region. The open complex is capable of binding NTPs and initiating transcription, giving its turn to the initiation complex (ITC). Schematically the different stages of this reaction can be



depicted as: $R + P \leftrightarrow RP_C \leftrightarrow RP_I \leftrightarrow RP_O \rightarrow ITC$

Fig. 4: An illustrated cartoon depicting a promoter composed of all possible DNA determinants that are recognised by RNA polymerase containing a σ^{70}-family sigma factor. The consensus sequences for the DNA determinants recognised by σ^{70} (-35, -10 and extended -10 elements) are indicated above the promoter cartoon, and their colour is consistent with that of the sigma regions that recognise them in Fig. 3. The consensus sequence of the full UP-element site (coloured grey) and that of its two functionally distinct sub-sites (coloured as different shades of green) are also indicated above the promoter cartoon.

In the **closed complex** formation, RNAP holoenzyme containing a σ^{70} family member is recruited to its conserved DNA determinants, i.e. the promoter. A typical **promoter** (Fig. 4) for this sigma family is composed of up to four different sequence elements, with the most conserved being the –10 and –35 hexamers (Siebenlist et al., 1980). The spacing between them is optimally 17 bp, but it can be also 1-2bps longer or shorter (Harley CB, 1987; Mitchell et al., 2003; Mulligan et al., 1985; Stefano and Gralla, 1982). A subset of promoters carry an additional 4-5 bp-long recognition sequence directly upstream of the –10 region, called the extended –10 element; this element can either substitute the requirement for a –35 hexamer (Kumar et al., 1993) or can work in concert with it (Minakhin and Severinov, 2003; Mitchell et al., 2003). Finally the UP-element is the only promoter determinant recognised by another subunit of RNAP apart from sigma; it serves as a docking area for the αCTD(s) of RNAP in some of the stronger bacterial promoters (reviewed in Gourse et al., 2000). The UP-element is composed of a 20 bp-long A/T-rich motif that is located upstream of the –35 hexamer and can be divided into two functional half-sites (proximal and distal), each of them independently contacted by a single αCTD (Estrem et al., 1999).

Although no crystal structure of RP_C exists, a model has been developed based on the existing holoenzyme structures (Murakami et al., 2002a; Fig. 5A). Regions 2.4 and 4.2 of sigma lie across the same face of the holoenyme and are separated by a 75-76 Å distance

(Borukhov and Nudler, 2003; Murakami et al., 2002a). In order to allow simultaneous accommodation of the two σ regions (2.4 and 4.2) to their corresponding –10 and –35 promoter elements, which are separated by a 17 bp long spacer (or even 1-2 bp longer), the N-terminal Zn^{2+} binding domain (ZBD) of the β´ subunit binds the spacer region and induces an 8° bend at the middle of it (centred at –25; Murakami et al., 2002a). In addition, the plasticity acquired by $\sigma_{4.2}$ from its interaction with the β flap ($\sigma_{4.2}$ moves at least by 6 Å, relative to the DNA; (Murakami et al., 2002a) positions the region adequately towards the –35 element, even when the latter lies in sub-optimal positions relatively to the –10 hexamer (Kuznedelov et al., 2002b). Intriguingly, the T4 anti-sigma factor AsiA disrupts the $\sigma^{70}{}_4$-β flap interaction in order to inhibit transcription from the majority of housekeeping promoters (those that are dependent on both –35 and –10 elements for initiating transcription; Colland et al., 1998; Orsini et al., 2004; Severinova et al., 1998) and redirect RNAP to its own promoters (Gregory et al., 2004; Hinton, 2005).

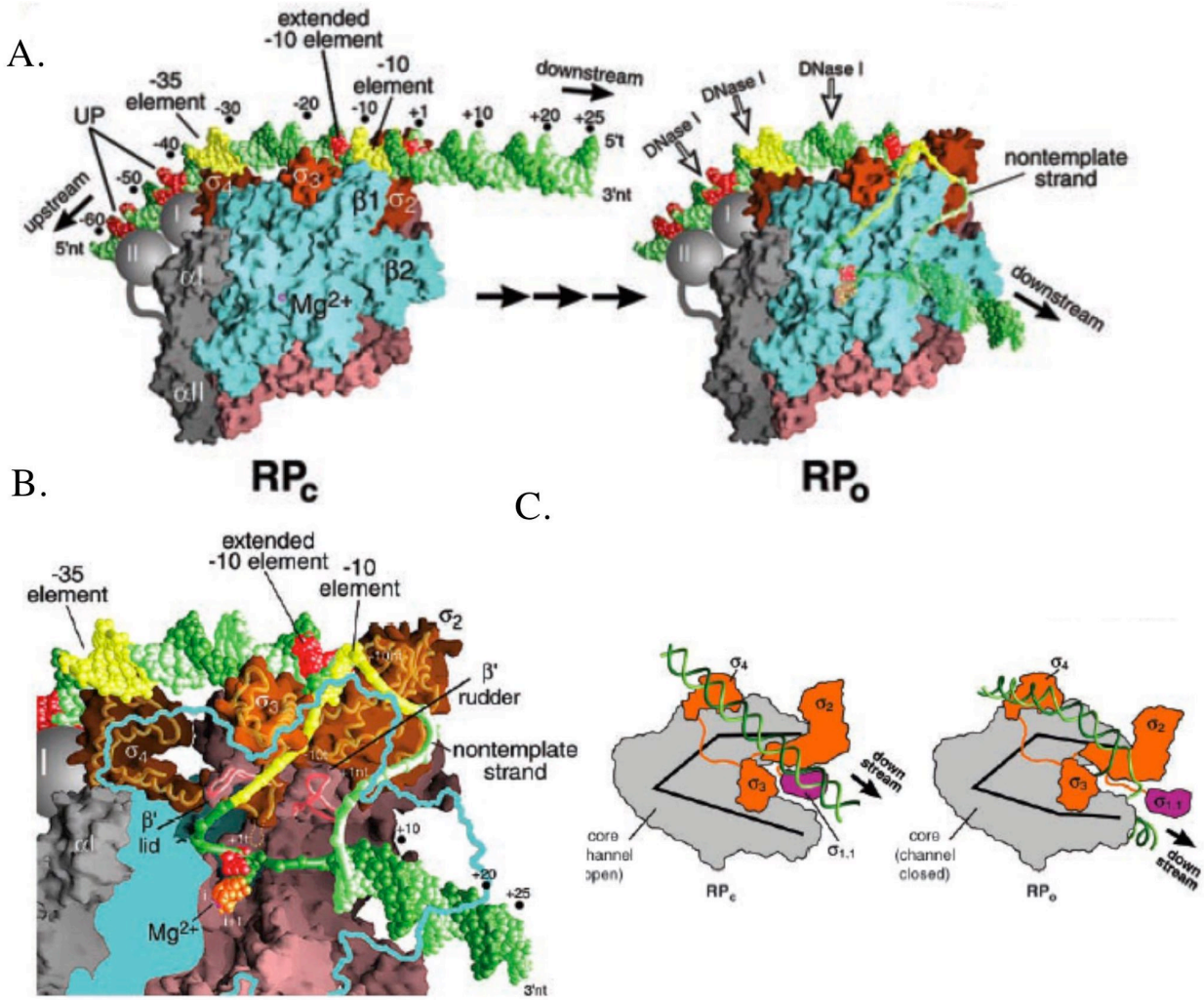


Fig. 5: RP_C and RP_O models (reproduced by Murakami et al., 2002a). In (**A**) and (**B**) are depicted views of holoenzyme-promoter DNA complexes along the pathway of open complex formation, which are based on available structural data from RNAP holoenzyme (Murakami et al., 2002b) and RNAP-DNA complex (Murakami et al., 2002a). Double-stranded DNA is shown as atoms, and single-

stranded DNA is shown as phosphate backbone worms with only the phosphate atoms visible. The template strand (t) is green, the non-template (nt) is light green, except for the -35 and –10 elements, which are yellow; and the UP elements, extended -10 element, and transcription start site on the template strand (+1), which are red. RNAP holoenzyme is shown as a molecular surface, colour coded as follows: aI, aII, ω, gray; β, cyan; β′, pink; and σ, orange. The possible disposition of the αCTDs (drawn as grey spheres, labelled "I" and "II") on the UP elements is illustrated. (**A**) Models of RP_C (left) and the final RP_O (right). The arrows in between denote that several intermediate steps exist along the pathway between these two states. The β subunit is rendered partially transparent to reveal the RNAP active site Mg^{2+} (magenta sphere) inside the main channel and the transcription bubble and downstream DNA enclosed inside the channel in RP_O. In RP_C, a numbering scale for the DNA position (-60 to +25) with respect to the transcription start site (+1) is shown above the DNA. In RP_O, RNA occupying the i and i+1 sites (catalytic centre) is shown as orange atoms. Sites of DNase I hypersensitivity in footprints of open complexes are denoted by open arrows in exposed minor grooves at approximately -45, -35, and -25. (**B**) Magnified view of RP_O, showing the details of the core promoter interactions, transcription bubble, and downstream DNA. Obscuring portions of the β subunit in front have been removed (the outline of β is shown as a cyan line) to reveal the structural elements inside the main RNAP channel. The molecular surfaces of the entire σ subunit, as well as of the β lid and β′ rudder are rendered transparent, revealing the α-carbon backbone worms (bright orange and pink, respectively) inside. The template strand DNA within the transcription bubble is directed through a protein tunnel framed by σ_2 and the σ_3-σ_4 linker underneath, an a-helix of σ_3 and the β′ lid on one side, σ_2 and the β′ rudder on the other side, and a domain of β (β1) in front, closest to the viewer, but seen only in outline. (**C**) Cartoon illustrating the proposed mechanism of the negatively charged σ region 1.1 in promoting open complex formation (reproduced by (Murakami et al., 2002b). Two states of the RNAP holoenzyme-promoter DNA complex are illustrated. The core RNAP is coloured grey, and σ is coloured orange, except region 1.1, which is coloured magenta. In the initial RP_C, σ region 1.1 is positioned inside the positively charged RNAP channel (protecting it from hydroxyl-radical cleavage; Nagai and Shimamoto, 1997), holding the channel open (indicated by thick black lines) to allow entry of double-stranded DNA. In the final RP_O, DNA has entered the RNAP main channel and the channel has closed, ejecting σregion 1.1, where it is exposed in solution to proteases (Murakami et al., 2002b) and hydroxyl-radical cleavage (Nagai and Shimamoto, 1997).

The interaction of the HtH motif of $\sigma_{4.2}$ with the DNA major groove causes an additional, sharper kink in DNA (36°) directly upstream of the –35 element (Campbell et al., 2002b). This observation correlates well with numerous DNase I footprint studies that had previously revealed a hypersensitive site at the same position, whereas the existence of additional hypersensitive sites more upstream (at -45 and often at -55) indicate that the binding of the aCTDs is further bending the DNA around RNAP (not pictured at Fig. 5A; Davis et al., 2005). It is worth mentioning that several studies point to an initial recruitment of RNAP from the upstream DNA determinants, i.e. the UP-element and the –35 hexamer, which is followed by the binding of RNAP to the –10 element (Buckle et al., 1999; Mecsas et al., 1991; Schickor et al., 1990; Sclavi et al., 2005). In addition both nuclease and hydroxyl radical footprinting of RP_C (stable only when it is "frozen" at low temperatures) had previously shown protection of a DNA region from –54 to –6 (Kovacic, 1987), supporting the notion that the downstream DNA does not enter the active channel at this stage. Murakami *et al* (Murakami et al., 2002b) based on structural and footprinting data suggested that during RP_C region 1.1 of sigma lies inside the main channel and keeps it wide open for duplex DNA

to enter (Fig. 5C). This would explain why $\sigma_{1.1}$ is shown to facilitate open complex formation at some promoters (Vuthoori et al., 2001). At some later intermediate step towards open complex formation, other signals might induce closure of the jaws that would eject $\sigma_{1.1}$ from the channel (Fig. 5C).

Before RNAP can initiate transcription, the downstream DNA must enter the main channel through the RNAP jaws, the DNA duplex has to be melted around the transcriptional start and the template strand should reach the active site to form the **open complex** (RP_O). Consistently, RP_O shows an extensive pattern of protection in footprinting studies, covering DNA regions up to +20 (Schmitz and Galas, 1979). A recent crystal structure of the RNAP holoenzyme bound to a fork junction DNA promoter fragment (encompassing double-stranded DNA from –41 to –12, followed by a single-stranded non-template strand from –11 to –7), RF, served as a base to model the RP_o structure (Fig. 5A and B; Murakami et al., 2002a), since RF is known to exquisitely mimic most of the properties of the RP_o.

The RF structure shows DNA making a sharp bend at –16 (37°) towards the holoenzyme, whereas the strands are predicted to separate from base pair -11 and follow different pathways in the RP_o model. Amino acids of sigma regions 2.4 and 3.0 are surface exposed, in position to contact the DNA duplex at nucleotide –12 and at the extended –10 element. In addition, universally conserved basic residues of regions 2.2 and 2.3 of sigma interact with the negatively charged phosphate backbone of the non-template strand at the extended –10 element. Beyond position –11, where the strand separation starts, the template strand is inserted into the active site of the holoenzyme to base pair with the initiating nucleotides, after passing through a tunnel consisting of regions σ_2, σ_3, β1 lobe, β´ lid and rudder. Several of the conserved basic residues in region 2.4 and 3.0, located in the beginning of the tunnel can be guiding the template strand through the tunnel. On the other hand, several aromatic residues of region 2.3 are shown to directly interact with the single-stranded non-template DNA until position –7 in the RF structure, consistent with their role in promoter melting. Interestingly the upstream edge of the transcription bubble is formed by an interaction of one of those aromatic residues (Trp256/Trp433 in *T. aquaticus*/*E. coli*) and the base pairs at position –12 and –11; the formation of these contacts may be the defining step in DNA melting (Heyduk et al., 2006; Lim et al., 2001). In the RP_o model, the non-template strand between nucleotides –6 and –3 is unbound and exposed, consistent with footprinting analysis revealing no protection at this region against DNase I or hydroxyl radical attacks (Wang and Landick, 1997). Nucleotides –2 to +4 of the same strand are held in a groove between the two β lobes (not that β2 lobe is also called β protrusion, as in Fig. 1A). The two strands reanneal

at nt +5 and until roughly nt +12 are clamped between the two RNAP jaws. The interaction of this downstream duplex with RNAP is proposed to be closing further the RNAP jaws and bringing DNA further in the main cleft by facilitating the DNA unwinding (Saecker et al., 2002).

The **intiation complex** (ITC) is marked by the formation of the first phosphodiester bond between two nucleotide triphosphates (NTPs) at the catalytic centre of the holoenzyme (i and i+1 positions). It is important to highlight that RNAP has the ability to initiate transcription de novo from two NTPs and does not need a primer like DNA polymerases do. Upon synthesis of the first short nascent RNA chains, RNAP comes into a stage where it can either proceed further with transcript elongation, which requires RNAP to be disengaged from the promoter and translocate down the template DNA, or release the transcript and start over again. Usually before RNAP succeeds to translocate, it goes over several rounds of **abortive initiation**, in which a short RNA is synthesized and then released. Seeing that in terms of the holoezyme structure (Murakami et al., 2002b; Vassylyev et al., 2002), it is apparent that the linker between σ_3 and σ_4 (and especially a hairpin loop of it, in the 3.2 region) occupies the same space as the nascent RNA chain in the RNA exit channel, blocking, thus, the extension of the RNA chain past a few nucleotides (3-10nt). Consistent with this model, truncated sigmas lacking region 3.2 or sigmas with mutations in conserved amino acids of the σ_3-σ_4 linker exhibit minimal levels of abortive initiation (Murakami et al., 2002b; Sen et al., 1998). Once the transcript reaches a length of approximately 12nt then it has actively displaced the whole σ_3-σ_4 linker, and by a length of 14-15nt it exits the RNA channel, after disrupting the σ_4-β flap interaction (Murakami et al., 2002b; Nickels et al., 2005). The disruption of the σ_4-β flap results in destabilisation of the contacts of σ_4 to the –35 element, initiating thus the process of promoter escape and the transition to the elongation complex. Intrinsic promoter features determine probabilities and rates of abortive initiation and promoter escape (Vo et al., 2003).

B. Transcriptional elongation

According to the long-year established **σ-cycle** model, one of the dogmas in bacterial transcription, sigma dissociates from RNAP upon promoter escape (after having successfully orchestrated promoter recognition and melting), and core RNAP forms a stable elongation complex (TEC) that proceeds along with transcript synthesis until it encounters a terminator site; core RNAP is then released from DNA and can only reinitiate transcription after re-employing a sigma factor (Travers and Burgess, 1969). Consistently, biochemical and

structural data exhibited that sigma factor's affinity to core RNAP suffers a significant drop once the polymerase synthesizes a 9-14 nt RNA transcript and disengages from the promoter (Daube and von Hippel, 1999; Gill et al., 1991; Hansen and McClure, 1980; Krummel and Chamberlin, 1989; Metzger et al., 1993; Murakami et al., 2002b; Straney and Crothers, 1985; Straney and Crothers, 1987; Vassylyev et al., 2002). Moreover, several of these studies went further and proposed that **obligate release** of sigma was required for RNAP to form a stable TEC that can overcome abortive initiation and escape the promoter (Daube and von Hippel, 1999; Hansen and McClure, 1980; Krummel and Chamberlin, 1989; Metzger et al., 1993; Straney and Crothers, 1985; Straney and Crothers, 1987). This statement was adopted for many years- even by textbooks- as a general accepted truth. However, a series of independent recent studies challenged this belief, as sigma was often detected to be part of the TEC and to affect its behaviour (Bar-Nahum and Nudler, 2001; Brodolin et al., 2004; Kapanidis et al., 2005; Marr et al., 2001; Mooney and Landick, 2003; Mukhopadhyay et al., 2001; Nickels et al., 2004; Raffaelle et al., 2005; Ring et al., 1996; Wade and Struhl, 2004). Therefore, the obligate release version of the σ-cycle model had to be replaced by a new model that involves **stochastic release** of sigma upon elongation (already proposed by Shimamoto et al., 1986), combined with an ability of sigma to temporally rebind the TEC upon elongation (reviewed in (Mooney et al., 2005). According to this model, RNAP looses several of its contacts with sigma upon transition to a TEC (Murakami et al., 2002b; Nickels et al., 2005; Vassylyev et al., 2002) and this triggers a gradual release of sigma from the complex upon the elongation progress; sigma can weakly rebind TEC at any point of the elongation process, with the extent of this rebinding being probably dependent on σ's efficient concentration, the TEC processivity, and the existence of promoter-like DNA elements that can serve as pausing signals. Nevertheless, even according to the revised σ-cycle model, sigma is part of the TEC, mostly at the early stages of elongation (Raffaelle et al., 2005; Wade and Struhl, 2004).

The current structural model for the TEC (considered without sigma; see text in the Core RNAP chapter and Fig. 1B) provides clues about how the complex can remain extremely stable and highly processive. A series of interactions between RNAP, DNA and RNA are required for equipping TEC with such characteristics (Fig. 6): i) the RNA-DNA hybrid, both at its front and rear "zip-locks" (FZ, RZ; Komissarova et al., 2002; Komissarova and Kashlev, 1998; Kuznedelov et al., 2002a); ii) the single-stranded (ss) RNA in the RNA exit channel, probably interacting with the β rudder (Kuznedelov et al., 2002a; Uptain and Chamberlin, 1997; Wilson et al., 1999); iii) the downstream double-stranded (ds) DNA (between position +3 to +11)-RNAP interactions (Korzheva et al., 1998). The whole elongating complex seems

to be covering a 35bp long DNA area, as shown in numerous footprinting studies (Metzger et al., 1989; Wilson et al., 1999) and predicted by the structural model (Korzheva et al., 2000). However, recent atomic force microscopy measurements questioned this belief by showing ~60bp of DNA to be compacted within the TEC, which suggests that an extensive DNA wrapping against the enzyme takes place (Rivetti et al., 2003).

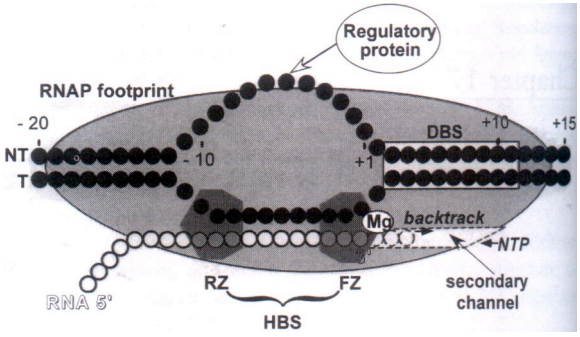


Fig. 6: Schematic model of the TEC (reproduced from (Artsimovitch, 2004). The DNA duplex (-20, +15; T and NT stand for template and non-template strand) is bound by RNAP and the duplex is melted for 12-13bp (transcription bubble). The non-template strand is exposed towards the surface of RNAP, where it becomes accessible to regulatory proteins. The nascent RNA (white circles) anneals with the template strand for 8-9 bp, and exits TEC at 14nt from the 3´end. During elongation the 3´end of the RNA chain is at the active centre, whereas in backtracked TECs it enters the secondary channel of RNAP (the NTP entry-gate). Three principal interactions stabilise the TEC: (i) the DNA binding site (DBS) where RNAP (with its jaws) contacts the DNA, (ii) the Front Zip lock (FZ; near the active site of RNAP) which marks the downstream end of the DNA-RNA hybrid and (iii) the Rear Zip lock (RZ; near the β´ rudder) which marks the upstream end of the DNA-RNA hybrid.

Despite the high resistance and processivity of TEC (Erijman and Clegg, 1998; Uptain et al., 1997), certain intrinsic DNA elements or nascent RNA structures might bring transcription to a halt (Artsimovitch, 2004). This halt, often followed by TEC backtracking along the DNA chain, can be transient (at **pausing** site), indefinite (at **arrest** sites), or even cause the release of the nascent RNA and the simultaneous breakdown of the complex (at **termination** sites). The DNA **backtracking** can be from 1-3bp in paused complexes and up to 18bp in arrested ones (Uptain et al., 1997). Several thermodynamical models have been employed to describe this ability of TEC to slide back and forth, or pause along the DNA chain (reviewed in Artsimovitch, 2004; Greive and von Hippel, 2005; and see Abbondanzieri et al., 2005; Bar-Nahum et al., 2005; Tadigotla et al., 2006 for more updated information). Moreover, several regulatory proteins have been directly implicated in regulating this "positional equilibrium" of the TEC (reviewed in Artsimovitch, 2004; Borukhov et al., 2005; Nudler and Gottesman, 2002).

More precisely, **GreA** and GreB suppress RNAP pausing in-vivo and in-vitro by stimulating the nucleolytic activity of RNAP (reviewed by Fish and Kane, 2002; recent structural data by Laptenko et al., 2003; Opalka et al., 2003; Sosunova et al., 2003). RNAP backtracks along the DNA after encountering a roadblock (pause site) and, as a result, the 3´-end of the nascent RNA chain enters the secondary channel; Gre factors induce the cleavage

of this 3´-end, providing, thus, RNAP with a new chance to resume transcription and surmount the roadblock. **Gfh1**, a Gre-like protein of *Thermus thermophilus*, uses also the secondary channel of RNAP, but this time to inhibit transcriptional elongation. Interestingly a pH-dependent conformational change is required before it can resume its role (Laptenko et al., 2006). In general, it appears that the use of the RNAP secondary channel is a widespread mechanism for regulating transcriptional elongation and/or initiation in bacteria (for review see (Nickels and Hochschild, 2004). Furthermore, **NusA** (its role reviewed in Gopal et al., 2001), a competitor of sigma in core RNAP binding (Gill et al., 1991; Greenblatt and Li, 1981), can both stimulate pausing and termination by itself, or trigger anti-termination with the aid of other auxiliary Nus factors (NusG, NusB or NuseE; Torres et al., 2004; Zellars and Squires, 1999). **Mfd**, the transcription-repair coupling factor, reactivates stalled and/or arrested TECs at DNA damaged sites (or releases them in conditions where NTPs are limiting), and recruits the DNA repair-excision machinery (reviewed in Roberts and Park, 2004). The reactivation of paused/arrested TECs is accomplished by promoting their forward translocation until the 3´-end of the nascent RNA chain comes out from the secondary channel (Park et al., 2002). Moreover, cold-shock induced RNA chaperones, like **CspA**, stabilize single stranded RNAs, preventing thus formation of RNA secondary structures (hairpins) that can arrest or terminate elongation (Phadtare et al., 2000). It is also conceivable that **topoisomerases** and the DNA **gyrase** change the ability and rate of RNAP's movement along the DNA, by affecting the extent of DNA negative supercoiling and the associated R-loop formation (Broccoli et al., 2004; Drolet, 2006). In contrast to all previous factors, **RfaH** (homologue to NusG) stimulates overall elongation rates and suppresses pausing of TEC at *specific* DNA sites (at a 12bp sequence, called ops site; Artsimovitch and Landick, 2002), facilitating this way expression of several virulent and biofilm associated genes in Gram negative bacteria (Beloin et al., 2006; Leeds and Welch, 1997; Nagy et al., 2002). Finally, various **bacteriophage anti-terminators** (λ Q and N, Xp10 p7, HK022 *put* RNAs) impose a lasting modification to TEC, either alone or with the aid of host-coding factors (NusA), so that the TEC becomes resistant to pausing or sometimes even to termination signals (Gusarov and Nudler, 2001; King et al., 1996; Nechaev et al., 2002; Nickels et al., 2002b; Ring et al., 1996; Roberts et al., 1998).

C. Transcriptional termination

Bacterial termination can occur in two fashions: factor-independent (intrinsic) or factor dependent. Both arts induce TEC dissociation from DNA, by releasing the nascent RNA

transcript. At **intrinsic** terminators, release is triggered at sites where nascent RNA folds into a stable, G/C-rich hairpin, followed by a U-rich stretch of nucleotides. Both the size of the hairpin and the sequence of the 3´end of the nascent RNA can vary. Nevertheless, predictions for such termination sites have been successfully made for several bacterial organisms (d'Aubenton Carafa et al., 1990; Ermolaeva et al., 2000; Lesnik et al., 2001,de Hoon, 2005 #625). The 3´end, U-rich RNA induces pausing of the TEC, 7-8 nt downstream from the base of the hairpin, and the RNA hairpin destabilises the TEC (by affecting both the DNA-RNA hybrid and the ssRNA-RNAP interactions), so that the transcript is disengaged from the ternary complex (Gusarov and Nudler, 1999; Komissarova et al., 2002; Uptain and Chamberlin, 1997; von Hippel, 1998). Several protein factors can influence the process (NusA, NusG), but are not entirely necessary (Uptain et al., 1997). On the contrary, **factor-dependent termination** is critically controlled by the presence of a regulatory protein that can be of bacterial (Rho, Mfd, TRAP, L4) or bacteriophagal (Nun, Alc) origin.

The most important factor-dependent termination mechanism involves protein **Rho** and is associated with up to 50% of all termination events in *E. coli* (Richardson, 2002). Initially Rho attaches to a ~ 60 nt-long nascent RNA fragment (*rut* site) that carries a high proportion of C and low proportion of G residues (Schneider et al., 1993). This attachment can be inhibited by ribosomes translating the nascent RNA, securing this way that Rho will terminate transcription efficiently only in non-coding regions, or when translation is inhibited. Once attached to the RNA, Rho moves along the chain while wrapping it up (using both mechanical forces and ATP hydrolysis), and catches up the advancing RNAP in the “termination zone” (many times RNAP has already paused at this site), located up to 100nt downstream of the *rut* site (reviewed in Banerjee et al., 2006). The moment Rho encounters the RNA:DNA hybrid its helicase activity helps it to unwind the duplex. Release of RNA is triggered either by Rho passively pulling RNA that has been transiently separated from DNA, or by Rho pushing RNAP forward, which would remove the 3´end of the nascent RNA from the active site (Richardson, 2002; Richardson, 2003). Many aspects regarding the fashion that the hexameric Rho assembly binds nascent RNA, translocates along it and unwinds it became more apparent through the recently obtained crystal structure (Skordalakes and Berger, 2003), and are analytically discussed in (Kaplan and O'Donnell, 2003; Richardson, 2003).

Apart from reactivating stalled TECs, **Mfd** can additionally mediate termination at pausing sites (independently from Rho or cis-signals), by inducing forward translocation of RNAP (Park et al., 2002). **RapA** has been also suggested to mediate release and recycling of trapped TECs on tightly supercoiled DNA or under high salt concentrations in *E. coli*

(Sukhodolets et al., 2001). The more specifically working **L4** (a protein involved in the 50S RNA assembly) has been shown to mediate transcriptional termination in the leader region of the *S10* operon, when it is present in its free form (Zengel and Lindahl, 1993). The mechanism requires initial pausing of TEC by NusA, followed by L4 stabilisation of the paused state and RNA release (Sha et al., 1995). Similarly, the **TRAP** protein in *B. subtilis* binds the leader mRNA of the *trp* operon after activation by tryptophan, and thereby stabilises a termination hairpin; uncharged tRNA(Trp) stimulates the synthesis of anti-TRAP protein (AT) that sequesters TRAP, permitting thus *trp* transcription (reviewed in Babitzke, 2004; Gollnick et al., 2005; Yanofsky, 2004). Finally, bacteriophage proteins, such as **Alc** (T4 phage), terminate host transcription after binding its non-modified DNA and inducing release of rapidly moving TECs (Kashlev et al., 1993). In contrast to that, **Nun** (HK022 phage) induces transcriptional termination of the phage's genes, after it binds the DNA boxB motif and anchors RNAP to the template strand, rendering it thus unable to translocate neither forwards nor backwards (Hung and Gottesman, 1995; Hung and Gottesman, 1997). The host's NusABEG proteins enhance its action (Watnick and Gottesman, 1998), whereas Mfd is required for the "frozen" TEC to be released from the DNA (Nudler and Gottesman, 2002).

A special case of factor-dependent termination and anti-termination mechanisms are the **riboswitches**. Since their first discovery four years ago (Mironov et al., 2002), reports have been accumulating and highlighting their importance and prevalence in bacteria (reviewed in (Nudler and Mironov, 2004; Tucker and Breaker, 2005; Winkler, 2005). Small metabolites (FMN, guanine, lysine, glycine, S-adenosylmethionine, glucosamine-6-phosphate and thiamine pyrophosphate) or even ions (Mg^{2+}; Cromie et al., 2006) can specifically bind and thereby stabilise secondary structures adopted by the 5´untranslated region of mRNAs (either termination or anti-termination hairpins). These mRNAs encode proteins responsible for the metabolism or uptake of the effectors. In the absence of the metabolite/ion, the leader mRNA resides to a more stable structural conformation that has the opposite impact in terms of transcriptional termination.

Apart from metabolites there are numerous other effectors, such as regulatory proteins (bacterial or phage-encoded), tRNAs and ribosomes that can facilitate **anti-termination**. Several of the bacterial proteins mentioned above as able to reactivate stalled/arrested TECs can also mediate anti-termination. Among them, the **NusABGE** apparatus aids RNAP to bypass ρ-dependent termination when RNA sequences containing *nut* or *nut*-like elements are present (e.g. ribosomal operons; (Nodwell and Greenblatt, 1993; Squires et al., 1993). Its effects are enhanced in the presence of the phage **λ N** protein (reviewed in Nudler and

Gottesman, 2002). Moreover, the ribosomal protein **S4,** a bacterial analogue of the λ N protein**,** binds RNAP directly and can facilitate anti-termination alone or together with the NusABGE apparatus at ρ-dependent terminators, even in the absence of *nut*-like elements (Torres et al., 2001). In contrast to regulatory proteins with general anti-terminating effects, the **BglG** protein acts very specifically (similarly to metabolites in riboswitches) by binding and stabilising an anti-termination hairpin at the leader mRNA of the *bgl* operon in *E. coli*. Thus, the formation of an overlapping and more stable termination hairpin is prevented (Houman et al., 1990). In a completely analogous manner, functions **GlcT** in *B. subtilis*, which regulates the glucose uptake (Schilling et al., 2004). *B. subtilis* also uses uncharged **tRNA**s for stabilising the leader mRNA of the cognate amino acid's synthetase genes in an anti-termination form (Grundy and Henkin, 1993; Grundy et al., 2002). This anti-termination form is known as the T-box, by the conserved 14 nt sequence that is part of it (Gutierrez-Preciado et al., 2005). *E. coli*, on the other hand, uses **stalled ribosomes** for allowing both anti-termination hairpins to be formed (and parallel preventing the formation of overlapping terminator hairpins) and for blocking ρ-dependent termination (Gong and Yanofsky, 2002; Yanofsky, 1981). Finally, phages use a vast repertoire of mechanisms in order to overcome transcriptional termination, including the transforming of RNAP to a processive form that bypasses termination signals (see in the chapter above), or the inactivation of Rho protein (Linderoth et al., 1997).

2.1.4 Transcription factors

Although bacteria can globally modulate gene expression by switching the sigma factor in the RNAP assembly and redirecting, thus, RNAP to a new set of promoter sequences (that are bound with a wide range of affinities, analogous to the resemblance of the promoter to the consensus sequence), this type of regulation is relative *static*. In order to make gene regulation more *adaptive*, the cell chromosome encodes a plethora of regulatory proteins, known as transcription factors, which can adjust the transcriptional output of a gene/operon in response to a variety of environmental signals. Transcription factors add a second layer of regulation in bacterial transcription that is wider and more signal-integrative than sigma factors. Consistently, transcription factors are estimated to be more than 300 in *Escherichia coli* (Madan Babu and Teichmann, 2003), whereas sigma factors are only seven. Most of transcription factors are responsible for controlling the expression of one gene or operon, but there are some that influence the expression of an extended number of genes (Martinez-Antonio and Collado-Vides, 2003). It is important to emphasize that in this section we

concentrate only on *orthodox transcription factors*, which are *DNA-binding proteins* that effect *transcriptional initiation* of a gene. This category of proteins embodies the major way bacteria control gene transcription. Proteins that regulate subsequent steps of the transcription cycle, i.e. elongation, termination and RNA stability (for the first two see earlier sections and for the third see for review Carpousis, 2002) are much rarer and belong only to a broader sense in transcription factors. Moreover, mechanisms like turning genes on or off by programmed site-specific recombinational events (some types of phase variation in *Salmonella* and *E. coli*, or σ^K expression in *Bacillus*; Blomfield, 2001; Kunkel, 1991) are just exceptions that verify nature's ability to adopt unlimited ways in order to achieve its developmental goals.

A transcription factor can bind the promoter region in a sequence-specific or non-specific manner (as monomer, dimer or even as higher multimer) and mediate repression or activation of transcriptional initiation. Some transcription factors function solely as activators or repressors, whereas others can play either role, depending on the promoter context. In some extreme cases the same regulator can both act as an activator and as a repressor at the same promoter, depending on the environmental conditions (see below for more information). Nevertheless, the cell has to ensure that the presence of a transcription factor is consistent to its current needs and therefore tightly regulates both the expression and activity of the transcription factor. Especially their activity is controlled by a variety of mechanisms including: (i) small ligands binding and allosterically changing the conformation of the transcription factor and its DNA-binding affinity; (ii) covalent modification of a transcription factor (phosphorylation, methylation, acetylation) that changes again its conformation and DNA-binding affinity; (iii) sequestration of the transcription factor to the membrane by another protein (this can lead both to inhibition (Plumbridge, 2002) and activation (Crawford et al., 2003; Krukonis and DiRita, 2003; Krukonis et al., 2000) of the transcription factor); and (iv) regulated proteolysis.

Transcriptional activators improve the promoter performance either by supporting initial recruitment of RNAP to the promoter or aiding later stages of the holoenzyme's isomerisation to a transcription-competent open complex. Regardless of the stage that an activator exerts its effects in transcriptional initiation, the most common way to accomplish that is by directly contacting one of the RNAP subunits, i.e. either σ or αCTD (for review see (Dove and Hochschild, 2004). On the other hand, the majority of repressors do not contact RNAP. Despite the diversity observed in the manner transcription factors act, some of the general mechanisms of activation/repression are conserved in bacteria.

Simple activation involves the function of a single activator and can be roughly divided in five categories. Analytically, *class I* activation involves a regulator binding upstream of the –35 element as a dimer, and recruiting RNAP to the promoter by directly contacting the αCTD and placing it adjacently downstream of it (Fig. 7A). The linker tethering αCTD to the rest of the holoenzyme is long and flexible (Meng et al., 2000), so that class I activators can bind at several positions upstream of the –35 element, but always at the same face of the DNA helix (usually centred at positions –61.5, -71.5, -82.5 and so on; for review see (Busby and Ebright, 1999). *Class II* activation describes the situation where the dimer-activator binds a site partially overlapping with the –35 element that $\sigma_{4.2}$ contacts (Fig. 7B). Activation is then mediated by a variety of contacts that the activator can build with αCTD, αNTD and/or $\sigma_{4.2}$. These contacts stabilise the holoenzyme's binding and/or improve its ability to form an open complex (Hochschild, 1994; Jain et al., 2004; Rhodius and Busby, 2000; Rhodius et al., 1997). Due to the fact that both RNAP and the activator should be accommodated in overlapping positions, in such a way that a steric clash is avoided and an optimal sigma-activator interface is provided, the location of the activator is strictly confined to defined positions (usually centred between –41 and -42; Lavigne et al., 1992). A third mechanism of simple activation is the one employed by the MerR family of proteins (Fig. 7C; reviewed in Brown et al., 2003): *MerR-type regulators* bind the spacer region of promoters with longer spacing between the –10 and –35 hexamers (19bp), and twist DNA so RNAP can optimally orientate its sigma regions 2.4 and 4.2 to recognise their cognate DNA elements (Heldwein and Brennan, 2001). The *MarA/SoxS/Rob family* of regulators in *E. coli* employs a completely different mechanism to activate transcription. Those activators interact with RNAP prior to promoter binding and thereby target it to promoters that carry an additional DNA determinant recognised by them (Fig. 7D; Martin et al., 2002). In all cases known, those regulators bind to αCTD, blocking its DNA-binding site (Dangi et al., 2004), and, thus, deprive RNAP of one of its important DNA-binding determinants (Gourse et al., 2000; Ross and Gourse, 2005). Therefore, parallel to activating their class of genes, this type of regulators indirectly repress genes that require the DNA-binding function of αCTD, or use an activator that binds to αCTD in order to recruit RNAP to the promoter. A similarly acting protein is Spx, found in *B. subtilis* (Nakano et al., 2003). Spx binds αCTD too (Newberry et al., 2005; Zhang et al., 2006), blocking thus transcription from many housekeeping genes (Nakano et al., 2003) and redirecting RNAP to its own regulon (Nakano et al., 2005). However, in the case of Spx, it is still unclear whether the regulator targets DNA on it own, or just alters the way αCTD binds DNA (Nakano et al., 2005). A fifth class of regulators, including NtrC and in some cases Fis,

facilitates transcriptional activation by binding to far *upstream promoter sites* (more than 100bp upstream of the promoter; in the case of NtrC even some thousand bp upstream; (Liu et al., 2001). This upstream binding triggers changes in the local DNA supercoiling, or buffers the local DNA supercoiling against rearrangements of the overall chromosomal DNA architecture (acting as a topological homeostat), so that RNAP can form an active open complex (Fig. 7E; Auner et al., 2003; Opel et al., 2004; Pemberton et al., 2002; Sheridan et al., 1998; Sheridan et al., 1999; Sheridan et al., 2001). In some cases the changes in local DNA superhelicity have been reported to be accompanied by wrapping of DNA against the RNAP, which is the exact reason for enhanced open complex formation (Auner et al., 2003; Pemberton et al., 2002). Curiously, recent reports have brought up examples where a transcription factor activates the promoter after contacting DNA-sites that are atypically situated downstream from the transcriptional start (the exact mechanism that activation is achieved is though unknown; Glinkowska et al., 2003; Munson and Scott, 2000).

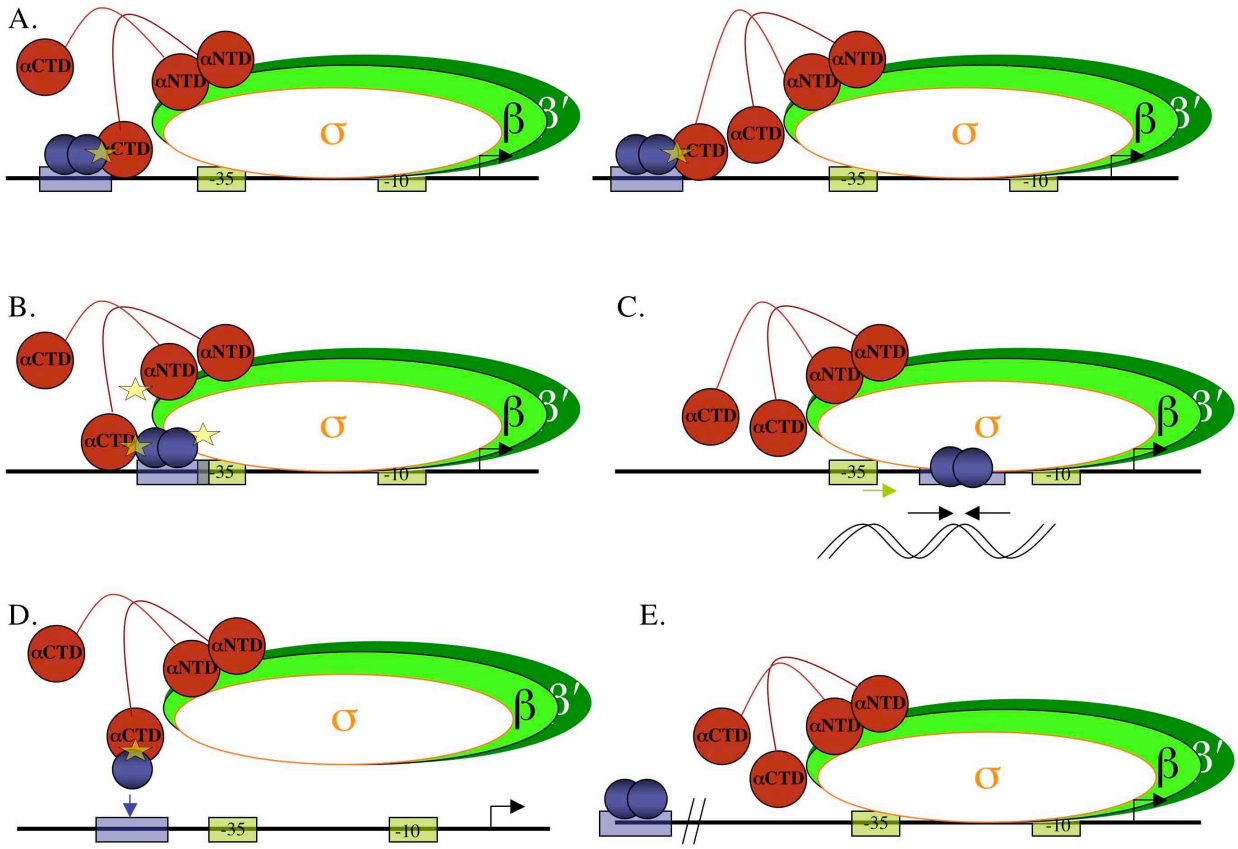


Fig. 7: Mechanisms of simple activation. This figure illustrates how RNAP and activator subunits (shown here in most cases as dimers) are organised during activation at simple promoters. A star indicates interactions between RNAP and activators. (**A**) Class I activation. The exact location of the activator's binding-site dictates whether one or both αCTD subunits can be accommodated downstream of the bound activator; in any case only one αCTD is contacted by the activator. (**B**) Class II activation; the activator (centred usually between –41 and –42) can contact αCTD, αNTD or σ_4 (the amount and art of contacts depend on the activator) in order to exert its effect. (**C**) MerR-type activation; the activator acts by contacting and twisting the DNA in the spacer region so that RNAP can recognise promoters with longer spacing between the –10 and –35 hexamers. (**D**) Activation by

targeting unbound RNAP; the activator interacts with RNAP prior to promoter binding and then exclusively directs it to promoters that carry an extra DNA determinant recognised by the activator. (**E**) "Distant" activation; the activator binds to far upstream sites from the promoter and ensures optimal local DNA supercoiling for RNAP to initiate transcription. Note that there are cases in which the activator recognises such far upstream sites, but nevertheless, acts as a class I type activator. In such cases a secondary transcriptional factor has to bind to DNA and loop it, so that the class I activator can reach and contact the αCTD (Weyand et al., 2001)

Analogously, **simple repression** defines the situation where a single protein can reduce or even shut down transcriptional initiation at a target promoter. There are four general mechanisms by which simple repression can be mainly achieved. The most common one is when the repressor recognises a site overlapping with the promoter, and this way occludes the binding of RNAP (Fig. 8A). In the majority of the cases, the repressor sterically hinders RNAP to bind to its core DNA determinants (-10 and –35 hexamers) but there are exceptions, in which the αCTD is prevented from making favourable interactions with the UP-element or other positive regulators (see Spx above). In other promoters, the repressor binds tandem or distant promoter-distal sites, and induces a DNA conformation (usually involving DNA looping) that blocks transcriptional initiation (Fig. 8B; Roy et al., 2004; Semsey et al., 2002). H-NS, a histone-like protein of *E. coli*, silences promoters usually this way (Dorman, 2004; Shin et al., 2005). A more unusual way of repression is when a transcription factor decreases the promoter activity, although it is bound at a classical activating site (class I position) and contacts RNAP the same way an activator does (in Fig. 8C through the αCTD). In such cases, the transcription factor irreversibly anchors RNAP to the promoter so that it cannot escape and initiate transcription (usually this requires the co-existence of core promoter elements that highly resemble the consensus sequence, so that RNAP binds strongly to them too; Monsalve et al., 1997; Monsalve et al., 1996). Finally, there are repressors that act via directly contacting the activator (after binding adjacently or further upstream of it) and preventing it from facilitating transcriptional initiation (Fig. 8D). Such type of repression exhibits CytR at several CRP-dependent promoters (Shin et al., 2001) or FNR when it binds at specific, tandem sites of a promoter (Barnard et al., 2003; Marshall et al., 2001).

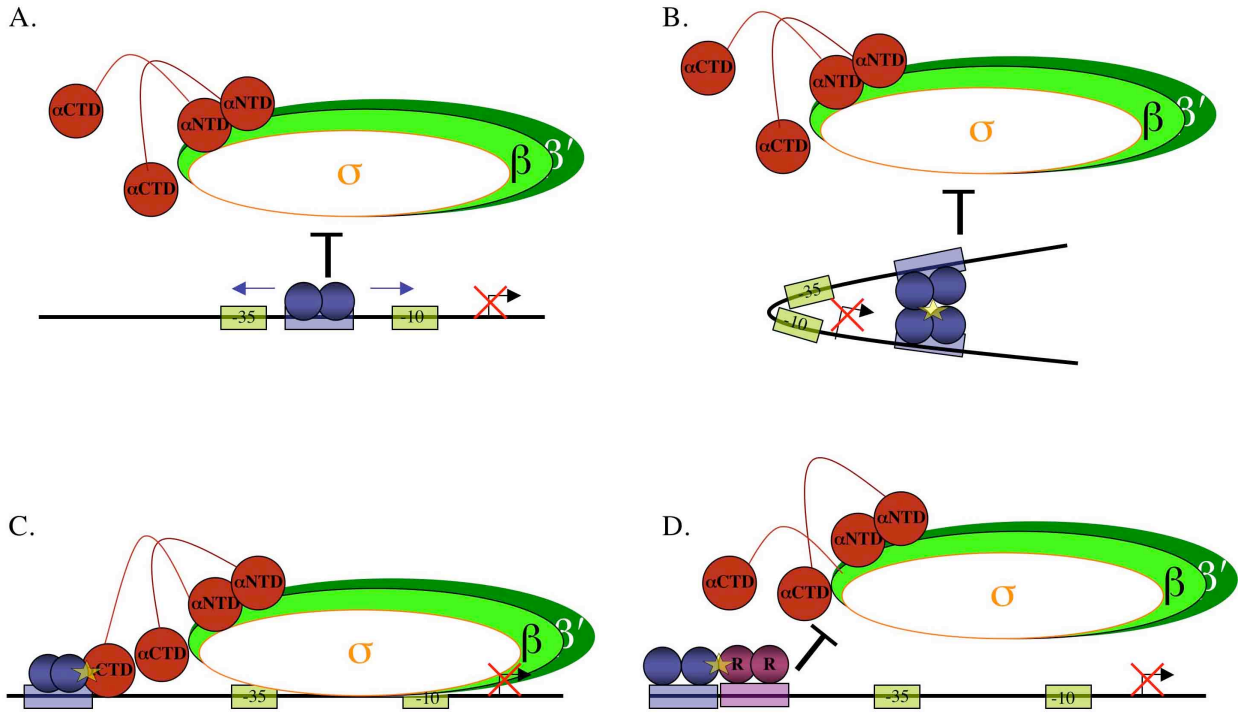


Fig. 8: Mechanisms of simple repression. This figure illustrates how RNAP and regulator subunits (shown here as dimers) are organised during repression at simple promoters. A star indicates interactions between RNAP and repressors. (**A**) Repression by steric hindrance; the repressor binds sites overlapping the core promoter determinants and blocks promoter recognition by RNAP. (**B**) Repression by DNA looping; the repressor binds tandem sites, downstream and upstream of the core promoter elements, inducing thus a bend in the DNA that occludes RNAP from the promoter. The DNA loop is sealed by interaction of the bound repressors. (**C**) Repression by RNAP "locking"; although the repressor is bound to a typical activator-site, its contact with RNAP is so strong that it prevents RNAP from escaping the promoter and completing successfully the transcriptional initiation process. (**D**) Repression by modulation of an activator protein. The repressor binds the activator and prevents it from stimulating transcriptional initiation.

However, since bacteria often require their gene expression to be responsive to diverse environmental cues, many of their promoters are controlled by two or more transcription factors. This art of **complex regulation** enables the promoter to integrate both global and specific metabolic signals in its activity (for reviews see Barnard et al., 2004; Browning and Busby, 2004). There are many mechanisms to accomplish such an integration of multiple signals in transcription initiation; most of them involve co-regulation of the promoter by two or more activators or a combination of repressor and regulator proteins. A frequently reoccurring pattern is when two activators bind *independently* the promoter and make separate contacts to RNAP (often referred to as class III mechanism of activation; Fig. 9A); those activators can function either both by a class I mechanism (Beatty et al., 2003; Tebbutt et al., 2002) or one by class I and the other by class II mechanism (McLeod et al., 2002; McLeod et al., 2000). A more complicated version of this transcriptional activation way is when two or more activators bind the promoter *cooperatively* (Fig. 9B); hence, one activator

is unable to bind in the absence of the other (Wade et al., 2001). Another fascinating mechanism is that of *repositioning* of an activator to a functional location (Fig. 9D). In this case, a secondary activator is required to trigger the relocation of the primary activator from a position where it is unable to activate transcription to one where it can support RNAP to assume its role. This repositioning can either occur by the secondary activator directly shifting the primary activator to a new DNA binding-site (Richet, 2000; Richet et al., 1991), or by the secondary activator triggering a DNA conformational change (e.g. looping) in order to optimise the relative position of the primary activator towards RNAP (IHF is a protein that often triggers such DNA looping; Schroder et al., 1993; Weiner et al., 1995). Finally, a last category of complex transcriptional activation involves an *anti-repression* strategy (Fig. 9C; Browning et al., 2002; Browning et al., 2000; Wu et al., 1998). This mechanism requires a secondary activator binding DNA and alleviating the inhibitory effect of a repressor protein on the primary activator (usually the secondary activator displaces the repressor).

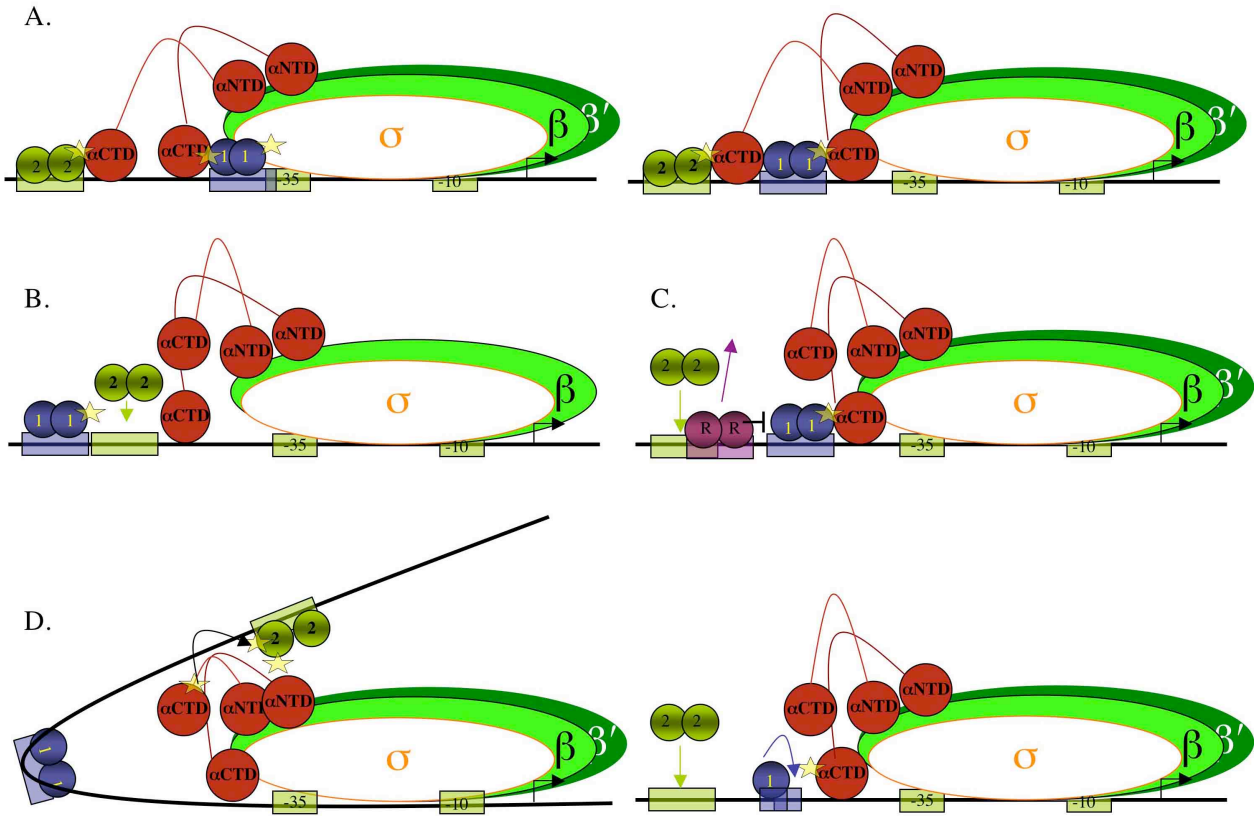


Fig. 9: Mechanisms of complex activation (co-dependence on two activators). This figure is a modified version from a figure appearing in (Browning and Busby, 2004). The primary and secondary activator (shown as dimers in most cases) are noted with 1 and 2 respectively; the repressor is noted with an R. A star indicates interactions between the activators or RNAP and an activator. (**A**) Class 3 mechanism; independent contacts by both activators are required for maximal promoter activity. In the left panel one activator functions by a class II mechanism and the second one by a class I mechanism, whereas in the right panel both act as class I activators. (**B**) "Cooperative binding" mechanism; the first activator recruits the second to its binding site and enables it thus to function. (**C**) "Anti-repression" mechanism; the binding of the secondary activator alleviates the inhibitory effect of the

repressor to the primary activator, and thus allows the latter to exert its positive effect in transcriptional activation of the promoter. (**D**) "Re-positioning" mechanism; the secondary activator binds DNA and alters either the actual location of the primary activator (right panel) or the relative location of the primary activator towards RNAP (left panel). In both cases the primary activator is then in position to interact with RNAP (αCTD and/or αNTD in the left panel, and αCTD in the right) and facilitate transcriptional initiation.

One has to keep in mind that there are several more intricate cases of simple or complex transcriptional regulation that do not belong to any of the general categories above, or are just slight deviations of them. For example a single protein can redistribute to new DNA binding-sites at a promoter, after an effector molecule binds to it, and stimulate or repress transcriptional initiation (AraC at the *araBAD* promoter, or Fur at some promoters in *H.pylori*; Delany et al., 2002; Delany et al., 2003; Schleif, 2003). This "dancing" of a transcription factor between activating and repressing sites at the DNA can become more sophisticated, and involve additional transcriptional regulators, plus modification of the binding sites by *dam* DNA methylation (like in the case of the *pap* promoter, which controls the expression of pyelonephritis-associated pili in the uropathogenic *E. coli*; Hernday et al., 2002; Hernday et al., 2004; Hernday et al., 2003; Weyand et al., 2001).

2.2 The stress-response and stationary phase sigma factor, σ^S/RpoS, in *Escherichia coli*

The σ^S subunit of RNA polymerase, a group 2 member of the σ^{70} family of sigma factors, has been repeatedly characterised as the master regulator of stress response and stationary phase in *Escherichia coli*. From its early discovery and the accompanying indications of its broad-ranged role about 15 years ago (Lange and Hengge-Aronis, 1991b; Mulvey and Loewen, 1989; Tanaka et al., 1993; Touati et al., 1991), up to the recent evaluation of its regulon extent, which can reach up to 10% of the organism's genes depending on the environmental conditions (Weber et al., 2005), our view of the multifaceted influence that σ^S has in cell physiology and morphology has enormously expanded (reviewed in Hengge-Aronis, 2000). σ^S (RpoS), a dispensable sigma factor upon vegetative growth (even conferring cells a selective disadvantage upon growth in nutrient limited chemostat cultures; reviewed in Ferenci, 2003), becomes of major importance to cells that reach stationary phase or are confronted with a series of stresses, including oxidative stress, hyperosmolarity, acidic shock, potentially lethal heat shock, low temperature, nutrient deprivation, ethanol, UV irradiation (reviewed in Hengge-Aronis, 2000). In contrast to other alternative sigma factors, which just help the cell to cope with a specific stress, σ^S triggers a complex cellular response that allows

the organism to *adapt* and *survive* the actual stressful condition, but also to *prepare* and be more tolerant against new upcoming stresses (cross-protection). Consistently, most of the σ^S–dependent genes encode proteins that primarily *prevent* stress or increase the cell's tolerance against it, rather than repair the cellular damage caused by it (Hengge-Aronis, 2000). Members of the σ^S regulon are genes conferring multiple resistance (to oxidative stress, acid, osmotic shock and ethanol), genes redirecting metabolism, transcription factors, chaperones, genes affecting the cell envelope and the overall morphology, genes triggering programmed cell death, members of the quorum sensing response, genes that affect biofilm formation, virulence genes etc (Hengge-Aronis, 2000; Weber et al., 2005). Most interestingly, the whole genetic reprogramming that σ^S orchestrates has an easily and rapidly *reversible* character, in contrast to developmental processes such as sporulation in *B. subtilis*, which show a committing, hierarchical and strictly forward-moving organisation (reviewed in (Errington, 2003; Hilbert and Piggot, 2004; Sonenshein, 2000).

In order for σ^S to effectively confer increased stress tolerance to the cell under diverse environmental stimuli, a variety of signals have to be integrated in the control of its expression and activity. Astonishingly, the cell controls every possible level of *rpoS* expression, so that it can integrate the different stimuli (Fig. 10), and employs a variety of mechanisms to succeed that (reviewed in Hengge-Aronis, 2002a).

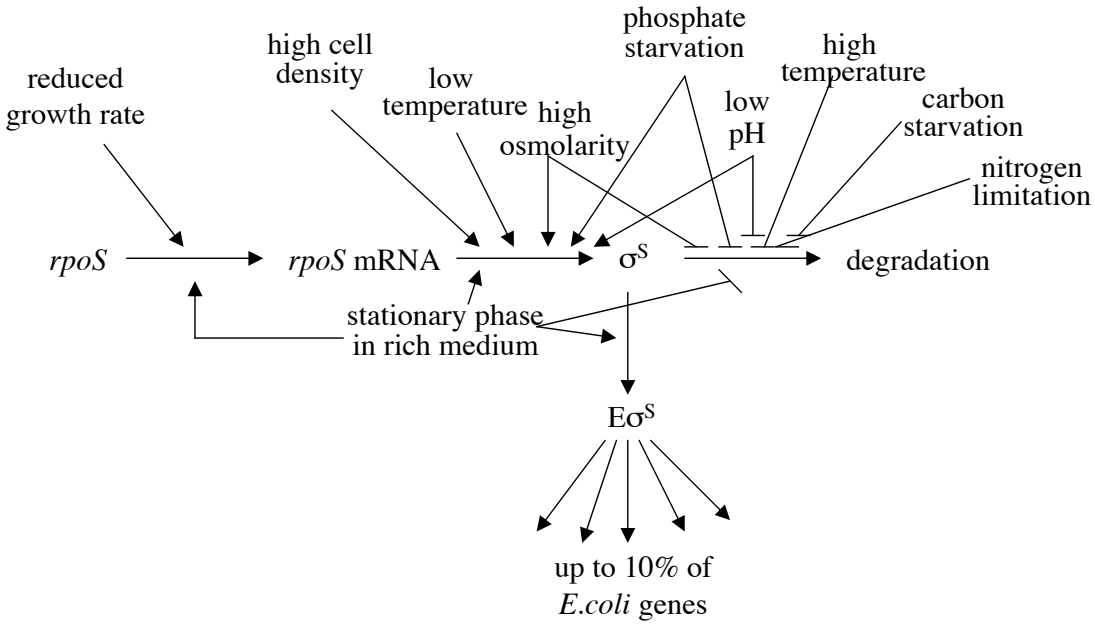


Fig. 10: Signal integration in the various levels of σ^S regulation. An increase in σ^S cellular levels can be achieved either by stimulating *rpoS* transcription/translation or by inhibiting σ^S proteolysis, which is extremely rapid during vegetative growth. A more efficient cell response is obtained by exerting combined effects on more than one level. Efficient σ^S-dependent gene expression is though not achieved by a mere increase in σ^S cellular levels, but additionally requires enhanced σ^S activity. Although until now only stationary phase is known to facilitate $E\sigma^S$ formation (indirectly; see text), it

is expected that other stressful conditions for the cell will generate a similar enhancement in σ^S activity.

2.2.1 Multifaceted control of σ^S cellular levels and activity

A. Regulation of rpoS transcription

Although transcriptional control has a major impact on σ^S cellular levels in other γ-proteobacteria such as *Pseudomonas* (Venturi, 2003), the case seems to be different in *E. coli*. Levels of *rpoS* mRNA tend to be relatively high even at time points where barely any σ^S protein is detected, and only slightly increase (in a degree ranging from 2- to 10-fold, depending on the growth media and conditions) upon most conditions that result in strongly elevated σ^S cellular levels (reviewed in (Hengge-Aronis, 2002a). Earlier studies identified that *rpoS* is transcribed by three different σ^{70}–dependent promoters: two of them are located in front of its upstream-positioned gene, *nlpD*, and produce basal levels of a polycistronic *nlpD-rpoS* mRNA during all phases of growth; a third promoter, *rpoS*p accounts for nearly all gene expression and the entire stationary phase induction, and is located within the *nlpD* coding region, producing thus a monocistronic *rpoS* mRNA with an unusually long 5´untranslated region (5´-UTR) (Lange et al., 1995; Lange and Hengge-Aronis, 1994b). The *rpoS*p activity seems to be regulated by a variety of proteins and/or non-proteinaceous effectors, most of the times though by unknown, complex or indirect means (see Fig. 11 for a depicted summary).

In detail, cellular **growth rate** seems to be of central importance in the regulation of *rpoS* transcription. Both continuous reduction of growth rate in complex media and drop in the quality of carbon substrate in defined media (that also produces slower growth rates) were shown to directly stimulate *rpoS* transcription (Lange and Hengge-Aronis, 1994a; Liu et al., 2005). Although this inverse correlation between growth rate and *rpoS* expression has also been observed repeatedly in the past in chemostat cultures, both in terms of σ^S protein levels (Ihssen and Egli, 2004; Teich et al., 1999; Zgurskaya et al., 1997) and in terms of σ^S–dependent gene expression (Notley and Ferenci, 1996), our knowledge still remains elusive about the exact mechanism that reduction of growth rate is sensed by *rpoS* transcription, and whether other levels of *rpoS* regulation use growth rate as a regulatory signal (Cunning and Elliott, 1999). A plausible candidate for the growth rate effects on *rpoS* transcription would be ppGpp (see also below), which accumulates in the cell in a growth rate dependent manner too. However the response of ppGpp to changes in nutrient availability is often distinct to that of σ^S (Teich et al., 1999).

Similarly, the repressive effect of **cAMP-CRP** in *rpoS* transcription during exponential phase of growth (Lange and Hengge-Aronis, 1991a; Lange and Hengge-Aronis, 1994a) is

mediated by an unidentified mechanism. Recent evidence suggests that this negative effect of cAMP-CRP in *rpoS* transcription is indirect (Mika F. and Hengge R., unpublished data), and might even be a consequence of reduced growth rates that the *cya* or *crp* mutants exhibit. On the other hand, cAMP-CRP directly activates *rpoS* expression in stationary phase, after recognising a typical class I binding-site (centred at –62.5; Mika F. and Hengge R., unpublished data). In addition, the repressive effect in *rpoS* transcription during exponential growth reported for **EIIA(Glc)** -the glucose-specific EII component of the PTS system- is due to the positive effect of EIIA(Glc) on the activity of the adenylate cyclase (Ueguchi et al., 2001). Additional positive roles in *rpoS* transcription have been reported for **ppGpp** and **polyphosphate**, but the way they are exerted (after transcriptional initiation, in elongation phase?) remains hazy and complicated (Gentry et al., 1993; Hirsch and Elliott, 2002; Lange et al., 1995; Shiba et al., 1997).

Apart from the direct activating effect of CRP in stationary phase expression of *rpoS* (Mika F. and Hengge R., unpublished data), recent reports have identified additional transcription factors that directly control the main promoter's (*rpoS*p) activity. Phosphorylated **ArcA** specifically binds the promoter region at two positions, one upstream of the –35 element (centred at –63) and the other one downstream of the transcriptional start (centred at +23), and induces, thus, a DNA looping (seen as an extreme hypersensitive site at -15, in DNaseI footprints) that represses *rpoS*p especially during exponential phase (Mika and Hengge, 2005). Since the upstream site overlaps with the activating CRP-binding site, it seems plausible that upon entering stationary phase, cAMP-CRP accumulates and can substantially compete with ArcA-P (though it cannot completely replace it, as ArcA shows a residual inhibitory effect also in stationary phase) for DNA binding. This replacement of ArcA by CRP in *rpoS*p in stationary phase could be supported by the fact that less phosphorylated ArcA is predicted to be present during this time (the available energy/oxygen ratio is low keeping the quinones oxidised and ArcB non-phosphorylated; Mika and Hengge, 2005). Furthermore **Fis**, an abundant histone-like protein during exponential growth, binds to the promoter in an atypical site (centred at –50) and inhibits its activity during vegetative growth (Hirsch and Elliott, 2005a). The mechanism of action of Fis or whether interplay between Fis and ArcA-P takes place (by Fis stabilising the proposed inhibitory DNA looping mediated by ArcA-P) are still unaddressed issues. Conflicting reports concerning the sensor kinase **BarA** have attributed both positive and negative effects to its function in *rpoS* transcription during exponential growth (Mukhopadhyay et al., 2000; Sugiura et al., 2003). Intriguingly, those effects are hardly dependent on the cognate response regulator of BarA,

UvrY (Hengge-Aronis, 2002a; Sugiura et al., 2003), and thus might happen through cross talk of BarA with other response regulators.

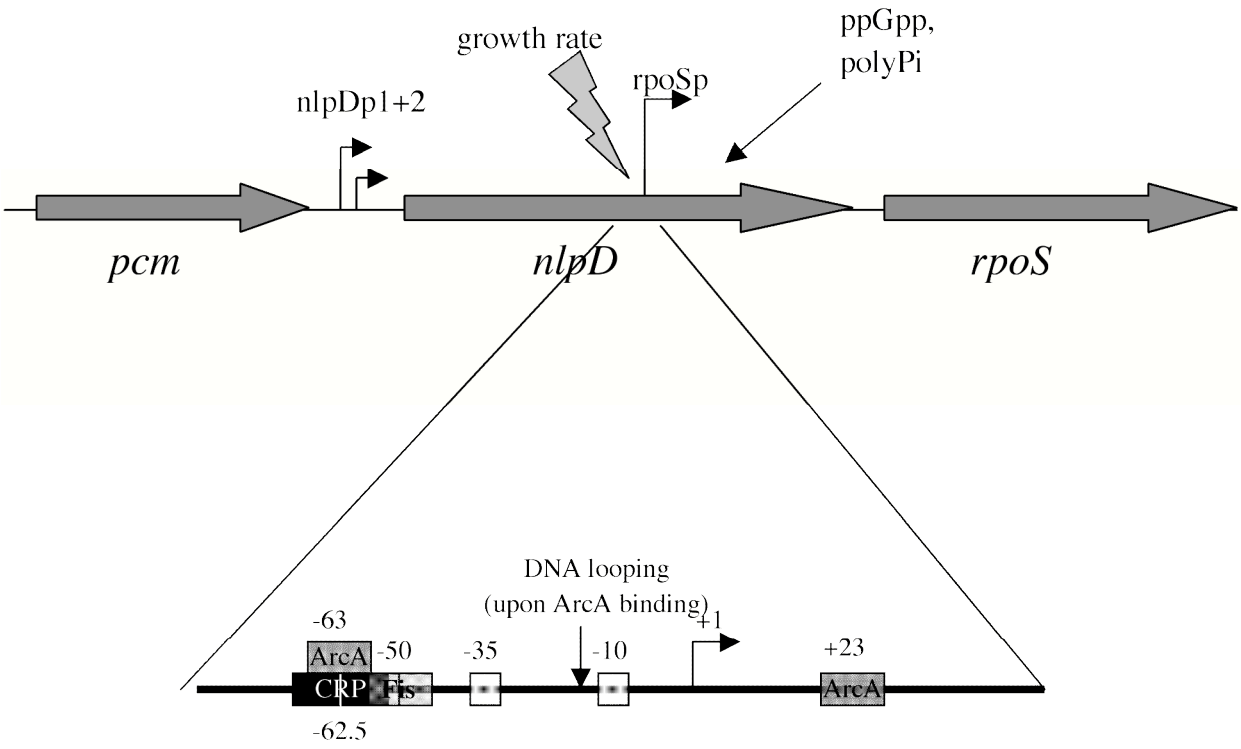


Fig. 11: Regulation of *rpoS* transcription. The operon promoters (*nlpD*p1 and *nlpD*p2) contribute to basal and constant expression of *rpoS*, whereas the *rpoS* promoter (*rpoS*p) is responsible for most of *rpoS* transcription, and is subject to growth-rate and phase regulation (see text for details). The architecture of the *rpoS*p is shown in detail and the DNA binding-sites of trans-acting factors known to be directly regulating promoter's expression, and the –10/–35 elements, are drawn as rectangles. The site where the DNA makes a loop upon ArcA-P binding is noted with an arrow (see text and Mika and Hengge, 2005).

B. Regulation of rpoS translation

The 5´-untranslated region of the major *rpoS* transcript (originating from the *rpoS*p promoter) seems to adopt two structural conformations, a repressive one that masks the translational initiation region (TIR) and makes it inaccessible to ribosomes during exponential phase, and an activating one, that allows unobstructed translation upon entering stationary phase and upon several stress signals (reviewed in Hengge-Aronis, 2002a). Several theoretical predictions, combined with genetic evidence, have proposed a series of structural versions for the repressive form of *rpoS* mRNA (Brown and Elliott, 1997; Cunning and Elliott, 1999; Hirsch and Elliott, 2005b; Lease et al., 1998; Majdalani et al., 1998). Which version of them is correct remains elusive (Hengge-Aronis, 2002a). However, it seems that a conserved

primary sequence around the ribosome binding site (RBS) is present in all enterobacteria and is absolutely necessary for the existence of both the repressive and activating forms of the *rpoS* mRNA (Hirsch and Elliott, 2005b). In addition many regulatory proteins and/or sRNAs can actively shift this equilibrium from one structural form to the other (Fig. 12; reviewed in Hengge-Aronis, 2002a; Repoila et al., 2003).

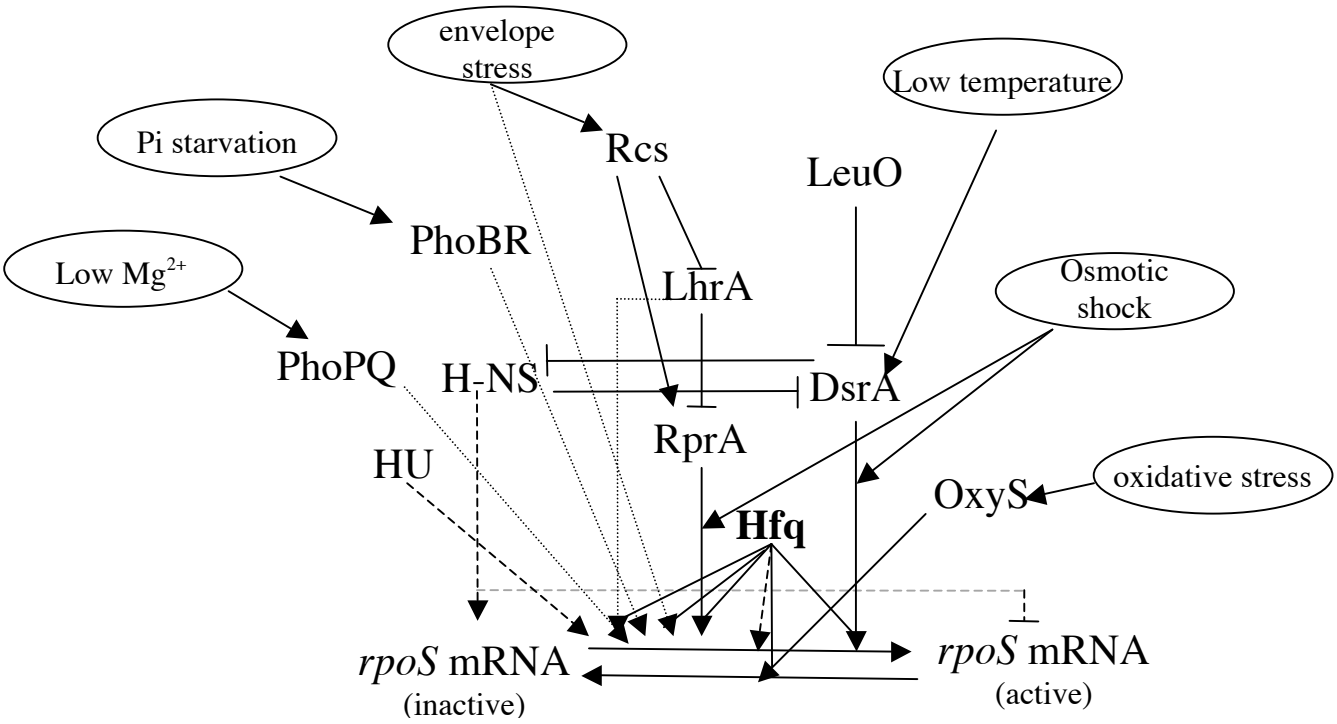


Fig. 12: Regulation of *rpoS* translation. *rpoS* mRNA can adopt at least two different conformations, one that has the translation initiation region (TIR) base-paired and thus blocks ribosomes from reaching it (inactive form), and one that renders TIR free for ribosomes to contact, and thereby permits unobstructed translation of *rpoS* (active form). Hfq has a central role in *rpoS* translation since it binds *rpoS* mRNA and recruits several known (DsrA, RprA and OxyS) or unknown small RNAs (PhoBR- or LhrA- dependent) that can directly stabilise the active or inactive *rpoS* mRNA form. Since most of them favour the formation of the active *rpoS* mRNA form, the presence of Hfq is shown to significantly stimulate *rpoS* translation in-vivo (whether though the binding of Hfq at *rpoS* mRNA per se favours the formation of the active or inactive form of *rpoS* mRNA is not clear). HU and H-NS are also very likely to exert their opposing effects in *rpoS* translation by directly binding *rpoS* mRNA (see text for more details).

The RNA-binding protein, **Hfq**, seems to play a central role in the translational efficiency of *rpoS* mRNA (Fig. 12; (Brown and Elliott, 1996; Muffler et al., 1996b). In its absence the repressive form of *rpoS* mRNA prevails all along the growth cycle (Brown and Elliott, 1997) and deprives the cell of physiological amounts and induction of σ^S (Muffler et al., 1996b). It is now evident that the hexameric ring formed by Hfq (homologous to eukaryotic Sm and Lsm proteins; (Arluison et al., 2006; Sauter et al., 2003) binds its target mRNAs (its role reviewed in (Valentin-Hansen et al., 2004), and stabilises directly one of the mRNA adapted conformations and/or serves as recruiter/chaperone to other small RNAs (sRNAs) that also

target and (de)stabilise a certain mRNA conformation (Geissmann and Touati, 2004; Lease and Woodson, 2004; Moller et al., 2002; Schumacher et al., 2002; Zhang et al., 2002). In the case of *rpoS*, different sRNAs bind Hfq and support the prevalence of one of the two structures that the mRNA can undergo (reviewed in Repoila et al., 2003). Recent evidence suggests that Hfq might even dissociate from the ternary complex after facilitating the binding of the sRNA (Lease and Woodson, 2004). Among those sRNAs, **DsrA** significantly stimulates *rpoS* translation at *low temperatures* (Sledjeski et al., 1996). This "thermometer" function of DsrA is based both on the enhanced transcription and stabilisation of its intact form -87 nt- at low temperatures (Repoila and Gottesman, 2001; Repoila and Gottesman, 2003). DsrA acts positively on *rpoS* translation by an "anti-antisense" mechanism, after base-pairing to an upstream cis antisense element of the *rpoS* mRNA, which is assumed to block the TIR region (Lease and Belfort, 2000; Lease et al., 1998; Lease and Woodson, 2004; Majdalani et al., 1998). Hfq seems to firstly accelerate/stabilise DsrA binding at the *rpoS* mRNA (Lease and Woodson, 2004; Mikulecky et al., 2004), without though changing the secondary structure of DsrA (Brescia et al., 2003), and successively to be released from the ternary complex, leaving DsrA alone to exert its function (Lease and Woodson, 2004). A further sRNA, **RprA**, has a secondary positive role on *rpoS* translation (only in the absence of DsrA) via a yet unidentified mechanism (Majdalani et al., 2001). Since its regulation is associated with the Rcs phosphorelay system (Majdalani et al., 2002), which senses *cell surface stress* (reviewed in Majdalani and Gottesman, 2005), it is plausible that RprA might stimulate σ^S levels upon such conditions. In addition both DsrA and RprA are partially responsible for the increase in *rpoS* translation observed under osmotic shock (Majdalani et al., 2001). Finally, a small RNA called **OxyS** has been shown to mediate a repressive effect in *rpoS* translation (Altuvia et al., 1997). This repressive effect goes through Hfq (Zhang et al., 1998). Since Hfq facilitates OxyS recruitment at its mRNA targets (Zhang et al., 2002), it seems plausible that a ternary complex, formed by the two of them and *rpoS* mRNA, stabilises the repressive form of the *rpoS* mRNA. Alternatively, OxyS might facilitate mRNA destabilisation, by functioning as part of ribonucleoprotein complexes containing RNases, as it was shown to be the case for other "repressor" sRNAs in *E. coli* (Masse et al., 2003; Morita et al., 2005). OxyS and σ^S are both induced under *oxidative stress*, but OxyS represses σ^S expression, resulting, at first glance, to a paradoxical situation. A plausible explanation is that the cell aims to maintain a balance between the two separate (and competing) response mechanisms induced under these conditions: the σ^S-dependent one and another one mediated by σ^{70}/OxyR.

HU and **H-NS** are abundant nucleoid-associated proteins in *E. coli*, known to frequently alter its gene regulation by changing the general and/or topological DNA architecture (for review see Dorman and Deighan, 2003; Travers and Muskhelishvili, 2005a). In the case of *rpoS*, H-NS and HU mediate opposing effects on its translation, possibly by *directly* contacting the *rpoS* mRNA (Balandina et al., 2001; Barth et al., 1995; Yamashino et al., 1995). HU was shown to bind in-vitro the leader *rpoS* mRNA in a region that encompasses the TIR (Balandina et al., 2001; Traulsen D. and Hengge R., unpublished data) and thereby it was proposed to directly impose its *positive* effect on *rpoS* translation during late logarithmic phase (Balandina et al., 2001). On the contrary, H-NS has a *negative* effect on *rpoS* translation (Barth et al., 1995; Yamashino et al., 1995), which might be direct since H-NS binds specifically both *rpoS* mRNA and DsrA (Brescia et al., 2004). Surprisingly, H-NS appears to enhance the in vitro cleavage of DsrA and *rpoS* mRNA by ribonucleases, and decrease, thus, the stability of the complex (Brescia et al., 2004). This implies that H-NS might act in-vivo similarly to some sRNAs that inhibit gene translation in *E. coli* by facilitating the formation of ribonucleoprotein complexes containing RNases (Morita et al., 2005).

Several other reported effectors of *rpoS* translation have been shown/predicted to use more indirect pathways (see also Fig. 12). **LeuO**, a LysR-like transcriptional regulator, represses the expression of *dsrA* and this way affects *rpoS* expression at low temperatures (Klauck et al., 1997; Repoila and Gottesman, 2001). **LrhA**, another LysR-like transcriptional regulator indirectly represses *rpoS* translation, by affecting the expression of RprA and possibly another unidentified sRNA (Peterson et al., 2006). The expression of *lrhA* is repressed by the **Rcs** phosphorelay system, which also directly activates *rprA* transcription (Majdalani et al., 2002), providing, thereby, an additional indirect acting factor on *rpoS* translation (Peterson et al., 2006). Interestingly, an Rcs-independent pathway seems to stimulate *rpoS* translation upon envelope stress, caused by the loss of the inner core LPS (actually the loss of *hldD*, encoding for an ADP-L-glycerol-D-mannoheptose-6-epimerase that produces a precursor for the synthesis of inner-core LPS; (Joloba et al., 2004). The **PhoBR** two-component system has also been proposed to act positively on *rpoS* translation through influencing the expression of an unidentified sRNA (Ruiz and Silhavy, 2003). Recently, **PhoPQ** was proposed to positively influence *rpoS* translation in *Salmonella* through an indirect mechanism (Tu et al., 2006). **UDP-glucose** has also been reported to have a negative effect on *rpoS* translation (Böhringer et al., 1995; Hengge-Aronis, 2002a). In addition, **EIIA(Glc)** significantly contributes to repression of *rpoS* translation, via an

unidentified and probably indirect mechanism (Ueguchi et al., 2001). Finally, **DksA/ppGpp** (Paul et al., 2005) exert some indirect positive control on this level of σ^S expression, probably by directly influencing the transcription of some of the above-mentioned effectors (Brown et al., 2002; Webb et al., 1999).

C. Regulation of σ^S proteolysis

Although *rpoS* transcription and translation allow some basal σ^S synthesis during vegetative growth, σ^S cellular levels are barely detectable in exponential phase, due to rapid degradation (Lange and Hengge-Aronis, 1994a). The half-life of σ^S is a couple of minutes for logarithmically growing cells in minimal medium, and even lower for cells growing in rich medium. Rapid stabilisation occurs upon carbon starvation (Lange and Hengge-Aronis, 1994a), osmotic up-shift (Muffler et al., 1996c) and shift to acidic pH (Bearson et al., 1996). The stabilisation process is more gradual upon entering stationary phase in rich medium and during heat shock (Muffler et al., 1997).

ClpXP is the responsible protease for σ^S degradation (Schweder et al., 1996). Although most known ClpXP substrates are recognised by the protease alone, σ^S needs a specific recognition factor, RssB, which targets it to the protease (Muffler et al., 1996a; Pratt and Silhavy, 1996). **RssB** is an unorthodox response regulator with a typical phosphorylatable N-terminal receiver domain (Klauck et al., 2001), but with an unusual C-terminal output domain that carries no homology to DNA binding proteins. The phosphorylated form of RssB binds to σ^S with high affinity (Becker et al., 1999; Zhou et al., 2001) at a location in region 3.0 of the sigma factor, known as the *turnover element* (Becker et al., 1999). Residue K173 of σ^S seems to be the major contact site, whereas amino acids E174 and V177 seem to play also some additional role in the interaction (Becker et al., 1999). Moreover, K173 is a crucial amino acid for extended –10 element recognition in σ^S–dependent promoters (Becker and Hengge-Aronis, 2001). The binding of RssB to σ^S and the formation of a 1:1 complex (Klauck et al., 2001) enable ClpXP to recognise an N-terminal motif in σ^S (Stüdemann et al., 2003) and form a quaternary complex (Zhou et al., 2001). RssB possibly contacts ClpX in the quaternary complex (Moreno et al., 2000), and somehow facilitates the subsequent unfolding and degradation of σ^S by the protease (Stüdemann et al., 2003). It is still unknown, though, whether RssB has to be phosphorylated for exerting this additional role in σ^S degradation (apart from its role to deliver σ^S to the protease). Finally, RssB does not get co-degraded and at some point it is released from the complex, playing thus a catalytic role in the process (Klauck et al., 2001; Zhou et al., 2001; see also Fig. 13).

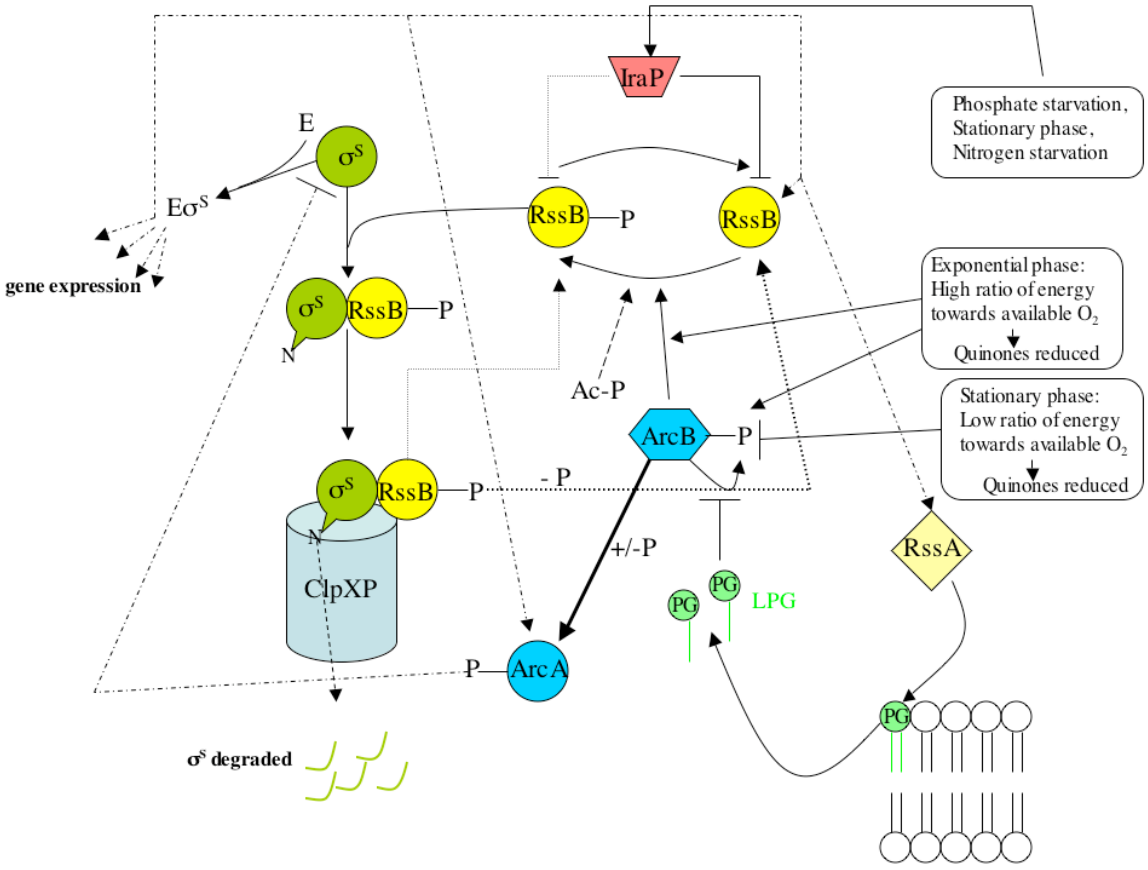


Fig. 13: The σ^S degradation pathway. The response regulator RssB acts as specific recognition factor for σ^S, with phosphorylation dependent affinity to it. An N-terminal motif of σ^S is exposed upon formation of the σ^S-RssB-P complex (depicted as a "tail" in the sphere representing RssB), and serves as the recognition signal for the ClpXP protease. After formation of the quaternary complex, σ^S-RssB-ClpXP, RssB facilitates initial unfolding of σ^S and subsequently is released from the complex, whereas σ^S gets completely degraded. Although it is not known whether RssB release is accompanied by obligatory dephosphorylation (actually dephosphorylation might even happen earlier, upon formation of the quaternary complex), it is tempting to speculate that this is true, since extensive efforts have failed to identify a phosphatase for RssB. On the contrary, phosphorylation of RssB is mediated by "cross-talk" with a sensor kinase, ArcB, and to a lesser extent by acetyl phosphate (Ac-P). ArcB is autophosphorylated in response to the presence of reduced quinones (see also text), and then predominantly functions as a phospho-donor for ArcA. This means that ArcA has a dual role on regulating σ^S levels; on one hand it directly represses *rpoS* transcription along the growth curve (especially during log phase), and on the other hand it indirectly stabilises σ^S in stationary phase by competing with RssB for phosphorylation by ArcB. IraP acts as an anti-RssB protein (it is not clarified yet whether it can also bind phosphorylated RssB) and RssA uses its phospholipase activity to hydrolyse certain membrane phospholipids (PG), producing this way a lipid (LPG) that inhibits ArcB autophosphorylation. A series of positive and negative feedback loops fine-tune the pathway, since $E\sigma^S$ is responsible for *rssAB* expression and partially for that of *arcA*.

Although RssB has a catalytic role in σ^S degradation, its low concentration seems to be the rate-limiting factor for the whole process in-vivo (Pruteanu and Hengge-Aronis, 2002). Accordingly, sudden and strong increases in σ^S synthesis result in stabilisation of σ^S due to RssB titration (Pruteanu and Hengge-Aronis, 2002), and therefore part of the σ^S stabilisation observed during stress shock conditions, such as osmotic upshift, pH downshift or phosphate

starvation, might be partially a consequence of an increased *rpoS* mRNA translation. Interestingly the cell can adjust its RssB levels and counterbalance small increases in σ^S synthesis, since *rssB* transcription is solely dependent on a weak σ^S–controlled promoter (Pruteanu and Hengge-Aronis, 2002).

Nevertheless the main control-point of RssB function is its phosphorylation-dependent activity. Although the σ^S-RssB complex can also be formed in the presence of unphosphorylated RssB and σ^S can be subsequently degraded both in-vivo and in-vitro (Bouché et al., 1998; Peterson et al., 2004; Zhou et al., 2001), phosphorylated RssB shows increased affinity to σ^S (Becker et al., 1999; Zhou et al., 2001) and thereby makes the degradation process much faster and efficient in-vivo (Becker et al., 2000; Peterson et al., 2004). Keeping the amounts of RssB low and, thus, critical for degradation, enables the cell to use the phosphorylation as the central mechanism for regulating RssB activity. This, of course, does not exclude that in conditions where RssB is not a limiting factor for σ^S degradation, its phosphorylation acquires a rather minor importance (Peterson et al., 2004).

RssB does not have any cognate sensor kinase and thereby its phosphorylation has to be mediated by other means. Acetyl phosphate seems to play some role in this process (Bouché et al., 1998), but most phosphorylation derives from another sensor kinase, **ArcB**, which acts as a phospho-donor to RssB both in-vivo (during vegetative growth) and in-vitro (Mika and Hengge, 2005). The trans-phosphorylation occurs in an irreversible manner (ArcB does not act as a phosphatase for RssB), though significantly slower than the ArcB-mediated phosphorylation of its cognate response regulator, ArcA (Mika and Hengge, 2005). ArcB responds in vivo to the redox state of quinones (Georgellis et al., 2001), being able to autophosphorylate when quinones are reduced (Malpica et al., 2004). Therefore autophosphorylated ArcB would be expected to be present both during anaerobic growth, and during aerobic vegetative cell growth where oxygen is relatively low compared to the energy supply (Mika and Hengge, 2005; Salmon et al., 2005). In addition, it seems possible that other sensor kinases show a similar cross talk with RssB, since the effects of ArcB to σ^S degradation are only conditional and are not matching those of RssB (Mika and Hengge, 2005).

There are many factors that seem to directly influence the degradation process of σ^S. **RssA**, the product of the gene preceding *rssB* (and being co-transcribed with it), has been recently shown to possess phospholipase-A activity that enables it to target membrane phospholipids (Marquardt M., Hengge R., manuscript in preparation). One of the fatty acid products of the RssA-mediated phospholipid hydrolysis, LPG, was shown to directly inhibit

autophosphorylation of ArcB and thus decrease the trans-phosphorylation of RssB (Marquardt M., Hengge R., manuscript in preparation). Furthermore, a protein induced under phosphate starvation, **IraP**, stabilises σ^S after directly interacting with RssB and preventing it to direct σ^S to the protease (Bougdour et al., 2006). Although the exact mechanism that this anti-adaptor protein acts remains unidentified, it seems that IraP also plays some role in other conditions that stabilise σ^S, i.e. in stationary phase in rich medium and under nitrogen starvation conditions (Bougdour et al., 2006). Recently it was shown that *iraP* is expressed under the control of the PhoPQ two-component system in *Salmonella typhimurium* (Tu et al., 2006). An indirect positive effect of **H-NS** in σ^S proteolysis (Zhou and Gottesman, 2006) might be mediated by a direct repressive effect of H-NS to such anti-adaptor proteins (it is known that H-NS does not affect *rssB* expression, and that H-NS directly influences transcriptional regulation only by acting as a repressor). However, this anti-adaptor protein seems not to be IraP, or at least not only IraP (Zhou and Gottesman, 2006). Finally the association of σ^S with **core RNAP** protects it from degradation (Zhou et al., 2001) and therefore, conditions that directly/indirectly support $E\sigma^S$ formation, also stabilise σ^S.

D. Regulation of σ^S activity

In contrast to other alternative sigma factors, there are no known proteins that modulate σ^S activity. RssB has the potential to act as a typical anti-sigma factor, since when it binds to σ^S, it blocks $E\sigma^S$ formation (Zhou et al., 2001). Consistently, in *clp*$^-$ background and with its concentration considerably increased (higher levels than the physiological), RssB assumes the role of an anti-σ^S factor (Becker et al., 2000). However physiological conditions that something like this would happen are unlikely to exist in *E. coli*, since ClpXP is constantly present in the cell and RssB levels are clearly lower than those of σ^S under all conditions tested (Becker et al., 2000). Consistently, in the majority of cases, formation of the σ^S-RssB complex in-vivo leads to σ^S degradation (during carbon starvation, the formation of σ^S-RssB complex does not lead to σ^S degradation but to σ^S inactivation; Becker, 2003). A more plausible scenario is that of RssB being earlier an anti-sigma factor that at some point in evolution got recruited by the proteolysis machinery as a targeting factor (Becker et al., 2000). An interesting twist of this story is that *rpoS* was found to possess an internal secondary translation initiation region (TIR) that leads to expression of a functional σ^S protein, depleted though of its first 53 amino acids (Subbarayan and Sarkar, 2004b); such a protein is active (Subbarayan and Sarkar, 2004b) but cannot not be degraded (Stüdemann et al., 2003). The secondary TIR is active only in *E. coli* strains carrying an amber mutation that

prematurely ends σ^S synthesis at the 33rd codon (Subbarayan and Sarkar, 2004c). The produced σ^S levels from the internal TIR are lower than normal (Subbarayan and Sarkar, 2004b). Intriguingly, many wild-type *E. coli* strains are found to carry such an amber mutation (Subbarayan and Sarkar, 2004a), producing thus the truncated, non-degradable form of σ^S. This could mean that we are in the middle of an uncompleted event of evolution that changes σ^S regulation from a typical case of sigma/anti-sigma factor pair to a case of targeted proteolysis.

σ^S like any other alternative sigma has to compete with the housekeeping sigma for limiting amounts of core RNAP. As described in detail above, cells use a series of mechanisms and effectors to decrease σ^{70} effectiveness upon entering stationary phase or upon stress encounter. All alternative sigmas (σ^S included) benefit from this decrease of σ^{70} effectiveness, and their activity is tightly bound to the presence of factors that mediate it. In other words, factors like Rsd, 6S RNA, DksA and ppGpp enhance σ^S activity, though in an indirect manner (Bernardo et al., 2006; Jishage et al., 2002; Kvint et al., 2000; Trotochaud and Wassarman, 2004). However stationary phase is mainly the territory of the master regulator for stress responses, σ^S. Eσ^S is actively engaged in transcription of more genes than any other alternative sigma factor, with the majority of them being also activated in stationary phase (Weber et al., 2005). Nevertheless σ^S exhibits the lowest affinity for core RNAP of all *E. coli* sigma factors in-vitro (Colland et al., 2002; Maeda et al., 2000), and despite the strong increase of its protein levels upon entering stationary phase (Hengge-Aronis, 2002a), σ^S only reaches about one third of the σ^{70} levels under the same conditions (Jishage et al., 1996). Therefore it seems plausible that *E. coli* uses additional factors to specifically help σ^S recruit core RNAP at its promoters upon stationary phase and/or during several other stress conditions. Such factors could be Crl (known to stimulate expression of a series of σ^S–dependent genes; Pratt and Silhavy, 1998) and ppGpp (postulated to change the affinity of core RNAP to σ^{70} and possibly also to other alternative sigmas; Jishage et al., 2002), although direct proof for that was missing by the beginning of my PhD thesis.

2.2.2 Promoter recognition by Eσ^S; the σ^S/σ^{70} paradox

As mentioned before one of the primary means bacteria use to globally reprogram their gene expression is to switch sigma factor in the RNAP assembly and redirect RNAP to a new set of genes, which carry an appropriate promoter sequence, specifically recognised by the corresponding sigma factor. However, the case seems to be different for the housekeeping sigma factor in *Escherichia coli*, σ^{70} and σ^s. These two sigma factors show a high degree of

sequence similarity and were found to bind optimally to identical -35 and -10 elements in vitro (Gaal et al., 2001). In vivo, however, σ^s efficiently activates the expression of a distinct regulon, which, depending on the conditions comprises up to 10 % of the *E. coli* genes (Weber et al., 2005). This paradoxical behaviour of σ^s promoter selectivity has puzzled scientists for many years, and by the beginning of my thesis considerable progress had been done on this issue (reviewed in Hengge-Aronis, 2002b).

Natural selection has resorted to a strategy involving either an increased tolerance of $E\sigma^S$ to minor deviations from the optimal consensus promoter (that both holoenzymes favourably recognise in-vitro), or a developed preference of $E\sigma^S$ to use promoter sequence features that are counter-selective for $E\sigma^{70}$. Therefore it seems that a series of "minor" sequential features, usually *surrounding* the core promoter elements (-10 and -35 recognition sites), are able to establish promoter selectivity (see also Fig. 14): i) C and T at positions -13 and -14, respectively, in the extended -10 region favour $E\sigma^s$ transcription, and C(-13) is at the same time counter-selective for $E\sigma^{70}$ (Becker and Hengge-Aronis, 2001); ii) an A+T-rich region downstream of the -10 box stimulates $E\sigma^s$-mediated transcription (Ojangu et al., 2000; Pruteanu and Hengge-Aronis, 2002), iii) sequence degeneration in the -35 or the -10 region can be tolerated more efficiently by $E\sigma^s$ (Gaal et al., 2001; Lacour et al., 2003); iv) -35 elements appear to be more dispensable for σ^s –dependent promoters (Espinosa-Urgel et al., 1996; Lee and Gralla, 2001); and v) a distal UP-element sub-site was shown in the case of *csiD*p to increase σ^s-selectivity (Germer et al., 2001).

-35		**ext-10**	**-10**		**+1**	
TTGACA -	12bp	- TRTG N	**TATAAT**	-5-8bp-	R	$E\sigma^{70}$
(TTGACA)-	12bp	- TRT**K** **C**	**TA*TAC*T**	-**AT-rich**-	R	$E\sigma^S$

Fig. 14: Consensus sequences of promoters preferentially recognised and transcribed by $E\sigma^{70}$ (upper lane) and $E\sigma^S$ (lower lane). Elements shown in bold are conserved in the majority of the $E\sigma^{70}$- or $E\sigma^S$-dependent promoters. The –35 element is often degenerate in $E\sigma^S$-dependent promoters and therefore is in parenthesis and not bold-faced, whereas nucleotides that often deviate from the consensus –10 sequence in $E\sigma^S$-dependent promoters are shown in italics. Underlined elements (together with –10 and –35 elements that can deviate from the optimal) contribute to $E\sigma^S$ selectivity of a promoter. Nucleotide symbols are used in accordance with the genetic nomenclature, and therefore K stands for T/G and M designates A/C.

Furthermore, $E\sigma^s$ selectivity can be generated by global transcription factors, e.g. CRP, Fis, IHF, Lrp or H-NS (Colland et al., 2000; Germer et al., 2001; Shin et al., 2005; Xu and Johnson, 1995). This involves transcriptional activators that differentially activate $E\sigma^s$ and $E\sigma^{70}$, depending on the location of their DNA binding-site (Germer et al., 2001), or

transcriptional activators that possess an optimal interface towards one of the holoenzymes (Nickels et al., 2002a). Alternatively, transcription factors (H-NS, IHF, Lrp) can selectively repress (or even silence) $E\sigma^{70}$-derived transcription more effectively than $E\sigma^{S}$-dependent one (Bouvier et al., 1998; Colland et al., 2000). Interestingly, in the case of H-NS, the selective repression is mediated by a unique ability of $E\sigma^{70}$ to act as a co-factor in the extensive H-NS promoter binding, which results in DNA wrapping around RNAP and making a sealed loop, which traps RNAP to the promoter (Shin et al., 2005).

Finally, the influence of global effectors that support increased $E\sigma^{S}$ formation/activity (or alternatively decrease the ability of $E\sigma^{70}$; for a detailed description see the previous chapter) can be decisive whether a promoter is favoured by $E\sigma^{70}$ or $E\sigma^{S}$ in-vivo. Apart from them, salt (Ding et al., 1995), glutamate concentration (Ding et al., 1995; Gralla and Vargas, 2006; Lee and Gralla, 2004) and global DNA supercoiling (Bordes et al., 2003; Kusano et al., 1996) have been proposed to increase $E\sigma^{s}$ effectiveness in-vivo.