Enhancer-Dependent Transcription in Salmonella enterica Typhimurium: New Members of the $\sigma^N$ Regulon Inferred from Protein Sequence Homology and Predicted Promoter Sites

David J. Studholme

Division of Biochemistry and Molecular Biology, School of Biological Sciences, University of Southampton, Bassett Crescent East, Southampton, SO16 7PX UK

Abstract

DNA-looping mediated by regulatory proteins is a ubiquitous mode of transcriptional control that allows interactions between genetic elements separated over long distances in DNA. In prokaryotes, one of the best-studied examples of regulatory proteins that use DNA-looping is the NtrC family of enhancer-binding proteins (EBPs), which activate transcription from $\sigma^N$-(sigma-N, sigma-54) dependent promoters. The completely sequenced genome of food-borne pathogen Salmonella enterica serovar Typhimurium LT2 contains seven novel EBPs of unknown function. Four of these EBPs have a similar domain organisation to NtrC whilst surprisingly the remaining three resemble LevR in Bacillus subtilis. Probable transcriptional targets are identified for each of the EBPs, including novel homologues of phosphotransferase system Enzyme II (EII) and several virulence-associated functions. Comparisons are made with the related enteric bacteria Salmonella Typhi, Escherichia coli and Yersinia pestis.

Introduction

DNA-looping, mediated by regulatory proteins, is a ubiquitous mode of transcriptional control that allows interactions between genetic elements separated over long distances in DNA (Rippe et al., 1995; Xu and Hoover, 2001). Generally, a protein bound to one genetic element some distance from a target gene interacts with another protein (or protein complex) associated with the promoter, thereby causing transcriptional activation or (occasionally) repression. One advantage of this type of regulation is that large macromolecular complexes can be pre-assembled at the promoter in readiness for activation by another protein that binds elsewhere, thereby increasing response efficiency.

Enhancer-Binding Proteins (EBPs) in Bacteria

In prokaryotes, one of the best-studied examples of regulatory proteins that use DNA-looping is the NtrC family of enhancer-binding proteins (EBPs). Binding of the activated form of an EBP to its target sequence and DNA-looping place it in contact with a special form of RNA polymerase (RNAP) containing the $\sigma^N$ subunit (Buck et al., 2000; Rombel et al., 1998). Sigma factors ($\sigma$) are subunits of RNAP that are required for promoter recognition and initiation of transcription. Among them $\sigma^N$ is unique (Merrick, 1993) in that it confers upon RNAP the requirement for activation by an EBP. Nucleotide-hydrolysis, catalysed by the EBP, drives the formation of an open transcription complex and hence activates expression of the target gene. Members of the NtrC family of EBPs have a modular domain organisation (Morett and Segovia, 1993). The middle domain is around 240 amino acids in length and contains seven highly conserved motifs including two ATP-binding motifs (Walker motifs A and B) and is responsible for nucleotide-hydrolysis and interaction with $\sigma^N$-RNAP. This central domain belongs to the AAA+ protein family (Neuwald et al., 1999) and may adopt a similar fold to that of other nucleotide-binding and hydrolysing proteins. The C-terminal domain contains a helix-turn-helix motif and is required for binding DNA. The N-terminus is the most variable domain and is absent in some EBPs. Its function is to receive sensory input, in the form of a phosphoryl-transfer or direct protein-protein interaction, and to modulate the transcriptional activation activity of the central domain.

Recently there have been exciting developments in elucidating the molecular mechanism of enhancer-dependent transcription (e.g. Chaney et al., 2001); but our understanding of the evolutionary origin and biological significance of this mode of regulation in bacteria still lags behind. The availability of complete genome sequences allows us to analyse the complete complement of proteins synthesised within the cell. Inspection of the numerous microbial genome sequences now available indicates that diverse NtrC-like proteins (paralogues) are present in many organisms studied to date (Studholme & Buck, 2000). However, few of these have been characterised and, for the most part, their target genes are unknown. In those cases where the paralogous systems have been investigated, it appears that target genes are usually involved in adaptive responses allowing survival in hostile environments. For example, cellular functions dependent on...
• MALDI-TOF Mass Spectrometry in Microbiology
  Edited by: M Kostrzewa, S Schubert (2016)
  www.caister.com/malditof

• Aspergillus and Penicillium in the Post-genomic Era
  Edited by: RP Vries, IB Gelber, MR Andersen (2016)
  www.caister.com/aspergillus2

• The Bacteriocins: Current Knowledge and Future Prospects
  Edited by: RL Dorfl, SM Roy, MA Riley (2016)
  www.caister.com/bacteriocins

• Omics in Plant Disease Resistance
  Edited by: V Bhadauria (2016)
  www.caister.com/opdr

• Acidophiles: Life in Extremely Acidic Environments
  Edited by: R Quatrini, DB Johnson (2016)
  www.caister.com/acidophiles

• Climate Change and Microbial Ecology: Current Research and Future Trends
  Edited by: J Marxsen (2016)
  www.caister.com/climate

• Biofilms in Bioremediation: Current Research and Emerging Technologies
  Edited by: G Lear (2016)
  www.caister.com/biorem

• Microalgae: Current Research and Applications
  Edited by: MN Tsiaoglou (2016)
  www.caister.com/microalgae

• Gas Plasma Sterilization in Microbiology: Theory, Applications, Pitfalls and New Perspectives
  Edited by: H Shintani, A Sakudo (2016)
  www.caister.com/gasplasma

• Virus Evolution: Current Research and Future Directions
  Edited by: SC Weaver, M Denison, M Roossinck, et al. (2016)
  www.caister.com/virusevol

• Arboviruses: Molecular Biology, Evolution and Control
  Edited by: N Vasilakis, DJ Gubler (2016)
  www.caister.com/arbo

• Shigella: Molecular and Cellular Biology
  Edited by: WD Picking, WL Picking (2016)
  www.caister.com/shigella

• Aquatic Biofilms: Ecology, Water Quality and Wastewater Treatment
  Edited by: AM Romani, H Guasch, MD Balaguér (2016)
  www.caister.com/aquaticbiofilms

• Alphaviruses: Current Biology
  Edited by: S Mahalingam, L Herrero, B Herring (2016)
  www.caister.com/alphaviruses

• Thermophilic Microorganisms
  Edited by: F Li (2015)
  www.caister.com/thermophile

• Flow Cytometry in Microbiology: Technology and Applications
  Edited by: MG Wilkinson (2015)
  www.caister.com/flow

• Probiotics and Prebiotics: Current Research and Future Trends
  Edited by: K Venema, AP Carmo (2016)
  www.caister.com/probiotics

• Epigenetics: Current Research and Emerging Trends
  Edited by: BP Chadwick (2015)
  www.caister.com/epigenetics

• Corynebacterium glutamicum: From Systems Biology to Biotechnological Applications
  Edited by: A Burkovski (2015)
  www.caister.com/cory2

• Advanced Vaccine Research Methods for the Decade of Vaccines
  Edited by: F Bagnoli, R Rappuoli (2015)
  www.caister.com/vaccines

• Antifungals: From Genomics to Resistance and the Development of Novel Agents
  Edited by: AT Coste, P Vandeputte (2015)
  www.caister.com/antifungals

• Bacteria-Plant Interactions: Advanced Research and Future Trends
  www.caister.com/bacteria-plant

• Aeromonas
  Edited by: J Graf (2015)
  www.caister.com/aeromonas

• Antibiotics: Current Innovations and Future Trends
  Edited by: S Sánchez, AL Demain (2015)
  www.caister.com/antibiotics

• Leishmania: Current Biology and Control
  Edited by: S Adak, R Datta (2015)
  www.caister.com/leish2

• Acanthamoeba: Biology and Pathogenesis (2nd edition)
  Author: NA Khan (2015)
  www.caister.com/acanthamoeba2

• Microarrays: Current Technology, Innovations and Applications
  Edited by: Z He (2014)
  www.caister.com/microarrays

• Metagenomics of the Microbial Nitrogen Cycle: Theory, Methods and Applications
  Edited by: D Marco (2014)
  www.caister.com/n2
the $\sigma^N$-EBP system in bacteria include responses to nitrogen starvation, phage shock, utilisation of unusual carbon sources, flagellar motility, virulence and cellular differentiation (Merrick, 1993; Barrios et al., 1999; Studholme and Buck, 2000; Buck et al., 2000). The properties of the $\sigma^N$-RNAP/EBP system represent a highly sophisticated mode of transcriptional regulation (Buck et al., 2000) but at a cost; the looping mechanism of activator-RNAP contact requires large stretches of intergenic DNA. In bacteria, which have very compact genomes with little intergenic spacing, there is presumably a trade-off between the advantages of the DNA-looping mechanism and the evolutionary constraints on genome size. Therefore it is instructive to ask how many, and what type of functions are regulated in this way in different bacteria.

**Enhancer-Dependent Transcription in S. typhimurium LT2**

*Salmonella enterica* subspecies I, serovar Typhimurium (*S. typhimurium*) is a major cause of human gastro-enteritis and causes typhoid-like symptoms in mice. Inspection of the recently sequenced *S. typhimurium* LT2 genome (McClelland et al., 2001) reveals that it encodes 14 protein sequences that contain all of the characteristic motifs of a $\sigma^N$-dependent EBP (Table 1). Seven of these proteins have been well studied in *Salmonella* or *E. coli*: RtcR, NtcC, HydG, FhiA, PgtA, PspF and PrpR. The transcriptional targets of the remaining 7 EBPs are currently unknown. However, two well-established characteristics of EBPs enable us to make some predictions. Firstly, all of the bacterial EBPs studied so far transactivate their targets through direct interaction with $\sigma^N$-RNAP (Buck et al., 2000; Xu and Hoover, 2001). Therefore each transcriptional target of an EBP must be linked to a $\sigma^N$-RNAP-binding DNA sequence. Binding sites for $\sigma^N$-RNAP are fairly easy to detect and have the consensus sequence YTGG-CACGNNNTTGCW (Barrios et al., 1999). Secondly, the ORF encoding an EBP is almost invariably sited close (within a few kilobases) to one or more of its targets (e.g. Ledebur et al., 1990; Genschik et al., 1998; Model et al., 1997; Marques et al., 1998; Palacios and Escalante-Semerena, 2000; Leonhartsberger et al., 2001). Nevertheless, there may be additional targets at distant sites in the genome. Therefore I searched the complete genome sequence of *S. typhimurium* LT2 for potential $\sigma^N$-RNAP-binding sites (Table 2) using the PromScan computer program (http:\www.promscan.uklinux.net/index.htm). The algorithm assumes a probabilistic model and assigns a score to potential binding sites. This score, known as the Kullback-Liebler distance, reflects the theoretical binding energy of the DNA-protein interaction (Stormo, 2000). Interestingly, close to all of the EBP-encoding genes there are high-scoring sites, providing strong clues as to the EBPs’ targets (Figure 1). The probable targets of each of the novel EBPs are discussed in the following sections.

---

Table 1. ORFs encoding $\sigma^N$-dependent enhancer-binding proteins (EBPs) identified from complete genome sequences of enteric bacteria.

<table>
<thead>
<tr>
<th>$\sigma^N$-dependent enhancer binding protein (EBP)</th>
<th>Proposed functions of genes regulated by EBP</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>PspF</td>
<td>Phage shock response</td>
<td>Model et al., 1997.</td>
</tr>
<tr>
<td>YgaA</td>
<td>Putative flavoprotein and reductase</td>
<td>Unpublished.</td>
</tr>
<tr>
<td>HydG</td>
<td>Tolerance to Zn$^{2+}$ and Pb$^{2+}$</td>
<td>Leonhartsberger et al., 2001.</td>
</tr>
<tr>
<td>PgtA</td>
<td>Phosphoglycerate transport</td>
<td>Jiang et al., 1998.</td>
</tr>
<tr>
<td>RtcR</td>
<td>RNA processing</td>
<td>Genschik et al., 1998.</td>
</tr>
<tr>
<td>STM0652</td>
<td>Amino acid transport</td>
<td>Unpublished.</td>
</tr>
<tr>
<td>STM2361</td>
<td>Putative PTS system Ell transport protein</td>
<td>Unpublished.</td>
</tr>
<tr>
<td>STM3773</td>
<td>Putative PTS system Ell transport protein</td>
<td>Unpublished.</td>
</tr>
<tr>
<td>STM0571</td>
<td>Putative PTS system Ell transport protein</td>
<td>Unpublished.</td>
</tr>
<tr>
<td>STM4534</td>
<td>Putative PTS system Ell transport protein, chemoreceptor</td>
<td>Unpublished.</td>
</tr>
</tbody>
</table>

---
YgaA and YfhA

Two of the uncharacterised EBP-encoding genes in Salmonella, ygaA and yfhA, have orthologues in E. coli K12. YgaA is homologous to the nitric oxide reductase regulator NorR in Ralstonia eutropha and Gardner et al. (2002) recently showed that the orthologues of STM2840 and YgbD in E. coli are components of a nitric oxide detoxification system and that mutation of ygaA eliminates inducible anaerobic nitric oxide metabolism. Therefore, YgaG probably activates transcription of this nitric oxide detoxification system (Figure 1A).

The sequence of the N-terminal domain of the YfhA gene-product resembles a receiver domain of the classical two-component signal transduction system (Stock et al., 2000) and the nearby gene yfhK encodes a histidine kinase sensor protein. Thus it appears that yfhKGA specifies an auto-activating two component signal transduction system linked to yfhG, a gene of unknown function (Figure 1B).

STM0652 and STM2361

Two further EBPs in S. typhimurium LT2 have no orthologues in E. coli, but are present in S. serovar Typhi. STM0652 appears to activate transcription of a putative altronate hydrolase of unknown function encoded by STM0649 and STM0650. Immediately downstream is STM0651 encoding a 2-keto-3-deoxy-

### Table 2. Close matches to the canonical σ^rRNAP-binding sequence in the complete genome sequence of Salmonella typhimurium.

<table>
<thead>
<tr>
<th>A</th>
<th>Promoter sequence</th>
<th>Gene name</th>
<th>Gene product/function</th>
</tr>
</thead>
<tbody>
<tr>
<td>2800</td>
<td>TGGCACCGTTTGGCT</td>
<td>rpoH</td>
<td>Heat shock sigma factor σ^r</td>
</tr>
<tr>
<td>2777</td>
<td>TGGCACGCTATTGCT</td>
<td>glutl</td>
<td>Glutamate/aspartate transporter</td>
</tr>
<tr>
<td>2688</td>
<td>TGGCACCGCTTTGCT</td>
<td>STM3521</td>
<td>Putative ribonucleotide related protein</td>
</tr>
<tr>
<td>2582</td>
<td>TGGCATAGTGTTGCT</td>
<td>yeaG</td>
<td>Protein kinase, similar to PrkA</td>
</tr>
<tr>
<td>2524</td>
<td>TGGCACAGCTGGCTA</td>
<td>astC</td>
<td>Succinylornithine transaminase, arginine metabolism</td>
</tr>
<tr>
<td>2507</td>
<td>TGGCGATATTGGCT</td>
<td>yleB</td>
<td>Putative negative regulator</td>
</tr>
<tr>
<td>2461</td>
<td>TGGGACAGCGGTGCT</td>
<td>STM2746</td>
<td>Putative excinuclease ATPase subunit, similar to RecN.</td>
</tr>
<tr>
<td>2439</td>
<td>TGGCAACGAATTTGCT</td>
<td>tpoA</td>
<td>DNA topoisomerase I</td>
</tr>
<tr>
<td>2426</td>
<td>TGGCATAGCCCTTGCT</td>
<td>STM0031</td>
<td>Similar to MarT and ToxR transcriptional regulators</td>
</tr>
<tr>
<td>2408</td>
<td>TGGCCGTTTATGGCC</td>
<td>cpsB</td>
<td>Cold shock protein</td>
</tr>
<tr>
<td>2400</td>
<td>TGGCTGTATCTGCTC</td>
<td>acrD</td>
<td>Multidrug transporter</td>
</tr>
<tr>
<td>2399</td>
<td>TGGTACGTATTTGCT</td>
<td>gilP</td>
<td>CilP protease</td>
</tr>
<tr>
<td>2389</td>
<td>TGGCACGGATATGCT</td>
<td>yrdA</td>
<td>Putative sideroophore-binding protein</td>
</tr>
<tr>
<td>2389</td>
<td>TGGCACGGATATGCT</td>
<td>yieD</td>
<td>Putative regulatory protein, possibly involved in gluconate utilisation</td>
</tr>
<tr>
<td>2389</td>
<td>TGGCACGGATATGCT</td>
<td>STM0649</td>
<td>Putative hydrolase STM0652</td>
</tr>
<tr>
<td>2383</td>
<td>TGGCACGGATATGCT</td>
<td>YfdG</td>
<td>Putative outer membrane efflux protein</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B</th>
<th>Promoter sequence</th>
<th>Gene name</th>
<th>Gene product/function</th>
<th>Associated EBP</th>
</tr>
</thead>
<tbody>
<tr>
<td>2769</td>
<td>TGGGATCCGCTTTGCT</td>
<td>STM4535</td>
<td>Putative PTS EII component</td>
<td>STM4534</td>
</tr>
<tr>
<td>2683</td>
<td>TGGCAGCAGCTTTGCT</td>
<td>STM3772</td>
<td>Putative PTS EII component</td>
<td>STM3773</td>
</tr>
<tr>
<td>2653</td>
<td>TGGCAGCAGCTTTGCT</td>
<td>STM0577</td>
<td>Putative PTS EII component</td>
<td>STM0571</td>
</tr>
<tr>
<td>2649</td>
<td>TGGCAGCAGCTTTGCT</td>
<td>STM2360</td>
<td>Putative dianisopimelate decarboxylase</td>
<td>STM2361</td>
</tr>
<tr>
<td>2625</td>
<td>TGGCAGCAGCTTTGCT</td>
<td>gkhK</td>
<td>Nitrogen regulatory protein</td>
<td>NtrC</td>
</tr>
<tr>
<td>2551</td>
<td>TGGCAGCAGCTTTGCT</td>
<td>prpB</td>
<td>Propionate metabolism</td>
<td>PrpR</td>
</tr>
<tr>
<td>2550</td>
<td>TGGCAGCAGCTTTGCT</td>
<td>hyaA</td>
<td>Hydrogenase accessory protein</td>
<td>FhA</td>
</tr>
<tr>
<td>2514</td>
<td>TGGCAGCAGCTTTGCT</td>
<td>zraP</td>
<td>Resistance to zinc and lead</td>
<td>HydG</td>
</tr>
<tr>
<td>2484</td>
<td>TGGCAGCAGCTTTGCT</td>
<td>STM0649</td>
<td>Putative hydrolase</td>
<td>STM0652</td>
</tr>
<tr>
<td>2431</td>
<td>TGGCAGCAGCTTTGCT</td>
<td>yfhK</td>
<td>Histidine sensor kinase</td>
<td>YfhA</td>
</tr>
<tr>
<td>2426</td>
<td>TGGCAGCAGCTTTGCT</td>
<td>hisJ</td>
<td>Histidine transport protein</td>
<td>STM2361</td>
</tr>
<tr>
<td>2407</td>
<td>TGGCAGCAGCTTTGCT</td>
<td>psaA</td>
<td>Phage shock response</td>
<td>Psf</td>
</tr>
<tr>
<td>2390</td>
<td>TGGCAGCAGCTTTGCT</td>
<td>hyaA</td>
<td>Transcriptional repressor</td>
<td>FhA</td>
</tr>
<tr>
<td>2383</td>
<td>TGGCAGCAGCTTTGCT</td>
<td>STM2840</td>
<td>Putative flavoprotein</td>
<td>YgaA</td>
</tr>
<tr>
<td>2354</td>
<td>TGGCAGCAGCTTTGCT</td>
<td>yfhK</td>
<td>Histidine sensor kinase, zinc/lead resistance</td>
<td>HydG</td>
</tr>
</tbody>
</table>
Figure 1. Operon structures of 7 genes encoding uncharacterised EBPs in *Salmonella typhimurium*. Predicted αN-dependent promoters are indicated (○).
gluconate permease. Therefore it is likely that this gene cluster is involved in utilisation of sugar acids (hexuronates).

STM2361 lies adjacent to a cluster of genes implicated in amino acid transport (Figure 1D) including argT, which encodes the periplasmic lysine-arginine-ornithine binding protein (Schmitz et al., 1988; Kustu et al., 1979). Although argT is controlled by nitrogen status through a $\sigma^N$-dependent promoter, there are no nearby binding sites for the general nitrogen control protein NtrC (Schmitz et al., 1988; Wu et al., 1999). Therefore, presumably an alternative EBP is required to activate this $\sigma^N$-dependent promoter. The most likely candidate is STM2361, its gene being immediately adjacent to this cluster of amino acid transport genes. Whether there is a regulatory link between the classical Ntr nitrogen control system and STM2361 remains to be investigated. STM2361 has a PAS domain at its N-terminus. It is conceivable that in Salmonella STM2361 represents a second nitrogen-responsive transcriptional activator in addition to NtrC but it is very unlikely that STM2361 directly senses nitrogen-status; PAS domains are found in a variety of signal transduction proteins in both prokaryotes and eukaryotes where they have been implicated in sensing of redox, oxygen and light but there is no evidence that they can sense nitrogen (Taylor and Zhiuln, 1999).

**Novel EBPs and the Phosphoenolpyruvate-Dependent Phosphotransferase Systems (PTS)**

The remaining three EBPs in S. typhimurium LT2 (Figure 1E, F, G) are associated with genes resembling Enzyme II (EII) components of the phosphoenolpyruvate-dependent phosphotransferase system (PTS). The classical phosphoenolpyruvate-dependent phosphotransfer system (PTS) of E. coli carries out the fundamental process of substrate level phosphorylation, whereby the transport and activation of sugar substrates are thermodynamically coupled to dephosphorylation of the glycolytic intermediate phosphoenolpyruvate. The phosphoryl relay proceeds sequentially from PEP to Enzyme I, HPr, Enzyme II and finally to the incoming sugar that is transported across the membrane and concomitantly phosphorylated by EII. Additionally, PTS components participate in signal transduction, chemotaxis, and the regulation of some key physiological processes e.g. carbohydrate transport, catabolite repression, carbon storage and co-ordination of carbon and nitrogen metabolism (Postma et al., 1993; Reizer and Saier., 1997; Saier and Reizer, 1994; Siebold et al., 2001; Tchieu et al., 2001). Homologues of PTS-components have previously been implicated in modulating the transcriptional activity of $\sigma^N$-RNAP (e.g. Merrick and Coppar, 1989), a link possibly allowing coordination of carbon and nitrogen metabolism. However, the identification of $\sigma^N$-dependent promoters upstream of PTS EII homologues reveals a further novel and unanticipated link between these two crucial regulatory mechanisms in Gram-negative bacteria. Strikingly, examination of their sequences reveals that STM0571, STM3773 and STM4534 are similar to the LevR-type of EBP in Gram-positive bacteria. These proteins contain three domains, but are quite distinct from the NtrC-like EBPs. Domain A is similar to the central domain of NtrC-like EBPs whereas domains B and C are similar to the BglG family of antiterminators and each contains a phosphor- ylatable histidinyl residue (Debarouille et al., 1991). Phosphorylation of the histidinyl residue located in domain B of LevR (by the general PTS component Hpr) stimulates transactivation of the levanase operon. In contrast, a PTS-like component encoded in the levanase operon phosphorylates the domain C histidinyl residue, suppressing transactivation. This allows sophisticated fine-tuning of gene regulation and integration of two different signals (Martin-Verstraete et al., 1998). The Salmonella EBPs share about 30% amino acid sequence identity with LevR, and the critical histidinyl residues are conserved. Thus it is probable that they operate by a similar mechanism. The links between the EBPs and the PTS-like transport systems in Salmonella suggest a feedback loop whereby an activated EBP induces the expression of the PTS-like components that in turn may phosphorylate the cognate EBP to modulate its activity. Moreover, both systems are likely to regulate a number of additional targets. However, none of these targets has yet been identified.

LevR-like sequences have been described in the Gram-positive Clostridium acetobutylicum (Nolling et al., 2001), Listeria monocytogenes and L. innocua (Glaser et al., 2001). In Enterococcus faecalis (Hecdard et al., 2001) five different LevR-like proteins each activate transcription of PTS EII components (Hecdard et al., 2001). LevR-homologues have not previously been reported in Gram-negative bacteria and are absent in E. coli, Yersinia pestis, S. Typhi and all the Gram-negative organisms whose genomes have been fully sequenced. However, BLAST searches against unfinished genome sequences reveal LevR-like sequences in Salmonella serovars Enteritis and Dublin and the closely related Klebsiella pneumoniae. Thus although common in low-GC Gram-positive bacteria, these three PTS-EBP gene clusters, reminiscent of the B. subtilis levanase operon, are something of a novelty in a Gram-negative bacteria. Phylogenetic analysis revealed that the Salmonella LevR-like EBPs did not form a discrete clade (Figure 2) and therefore probably did not arise from duplications within Salmonella but rather by several lateral transfer events. The biological significance of this intertwining of the $\sigma^N$- and PTS-regulons in S. typhimurium is as yet unknown but may be related to its lifestyle as a host-generalist food-borne pathogen also able to live outside the host. Indeed the three LevR-like EBPs are not found in the closely related S.typhi, which is largely restricted to human hosts.

**‘Orphan’ $\sigma^N$-Dependent Promoters**

Several high-scoring potential $\sigma^N$-binding sites identified in S. typhimurium LT2 are not sited close to genes encoding EBPs and have not been shown to be associated with any particular EBP (Table 2A). If indeed
\( \sigma^H \)-RNAP binds to these sites then they may represent genuine promoters. If they are genuine promoters, then transcription must be activated by an EBP either bound to a nearby enhancer element or from solution. It has been demonstrated that some EBPs can activate transcription from \( \sigma^N \)-RNAP binding sites even in the absence of enhancers (e.g. Leonhartsberger et al., 2001), although the physiological significance of this is unclear. Alternatively, \( \sigma^N \)-RNAP binding sites may act as negative regulatory elements, as is the case for the negatively autoregulated \( \sigma^N \) gene in *Pseudomonas putida* (Köhler et al., 1994).

Pallen (1999) first noticed a potential \( \sigma^N \)-RNAP binding site upstream of \( \sigma^H \) (encoding the heat shock sigma factor \( \sigma^H \)) in *E. coli*, and it is conserved in *Salmonella*. This sequence is very similar to the consensus of Barrios et al. (1999) and would be expected to bind \( \sigma^N \)-RNAP strongly. In *E. coli* growing on nutrient-poor media, deletion of \( \sigma^N \) gene (encoding \( \sigma^H \)) led to loss of transcription from the \( \sigma^N \)-dependent \( ibpAB \) promoter (Kucynska-Wisnik et al., 2001), further suggesting that \( \sigma^H \) may be dependent on \( \sigma^N \) under certain conditions. There is a strong potential \( \sigma^N \)-RNAP-binding site upstream of \( topA \), encoding topoisomerase I, which is required for temperature-dependent transcriptional regulation and for thermotolerance in several bacteria (Tse-Dinh et al., 1997). Four \( topA \) promoters have been previously identified, one of which is dependent on \( \sigma^H \). Whether \( topA \) is additionally regulated by \( \sigma^N \)-RNAP, and under what conditions, remains to be tested. Nevertheless these observations, along with the established role of \( \sigma^N \)-RNAP in regulating the *pspABCDE* and *ipbAB* heat shock operons (Model et al., 1997; Kucynska-Wisnik et al., 2001), may indicate intimate links between the heat shock and enhancer-dependent regulons.

The YeaG gene-product resembles *Bacillus subtilis* protein kinase PrkA, which phosphorylates an unidentified 60kDa protein in *B. subtilis*. PrkA was discovered during a screen of a *B. subtilis* genomic library for genes that would interact with the PTS. However, interaction with PTS was dependent on an adjacent ORF immediately upstream rather than on prkA itself (Fischer et al., 1995). Nevertheless this suggests yet another possible link between the \( \sigma^N \) regulon and the PTS in *Salmonella*.

In *Salmonella*, cspB encodes a small DNA-binding cold-shock protein, CspB. In *S. typhimurium* cells grown at 10°C, Craig et al. (1998) found a single
Concluding Remarks

Complete genome sequences are the starting point for an integrative approach towards understanding how genes and molecules are networked to form a biological system. However, while sequence data are accumulating rapidly, our current knowledge of biology constitutes only a small fraction of what remains to be discovered. Therefore, we are increasingly forced to rely on induction and inference to build models from existing data. In this paper I have shown how such deductive methods can take us some way towards understanding the ‘wiring diagram’ of a genetic regulatory network in a food-borne pathogen, S. typhimurium, that involves DNA-looping. The σ54-dependent regulon in S. typhimurium includes known-genes for assimilation of ammonia and amino acids, RNA-processing, tolerance to lead and zinc, formate-hydrogenlyase, phosphoglycerate utilisation, the heat/phage shock response, and propionate metabolism. Novel members of the region proposed here include a two component signal transduction system, hexuronate utilisation, an uncharacterised oxidoreductase system, CipP protease, and three PTS EII components. The repertoire of functions regulated by σ54-RNAP varies significantly even between quite closely related bacteria (Table 1) suggesting that this regulon is actively subject to natural selection and/or horizontal transfer events. Between more distantly related bacteria the differences are even greater (Studholme and Buck, 2000).

Of course predictions derived from in silico surveys such as this are prone to errors (see Cases and de Lorenzo, 2001). However, this approach has already generated hypotheses that were subsequently confirmed by laboratory-based experiments (Matthews and Timms, 2000; Hübner et al., 2001) and provides a framework for future investigations.

Acknowledgements

Preliminary sequence data were obtained from The Institute for Genomic Research website at http://www.tigr.org.

I am grateful to Prof David O’Connor for discussions and support.

References


Reizer, L.J., and Magasanik, B. 1986. Transcription of glnA in E. coli is stimulated by activator bound to sites far from the promoter. Cell. 45: 785–792.


