J. Mol. Microbiol. Biotechnol. (2002) 4(4): 367–374.

JMMB Mini-review

Enhancer-Dependent Transcription in *Salmonella enterica* Typhimurium: New Members of the σ^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 σ^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.

*For correspondence. Email djs@soton.ac.uk; Tel. 023 8059 4324; Fax. 023 8059 4459.

Enhancer-Binding Proteins (EBPs) in Bacteria

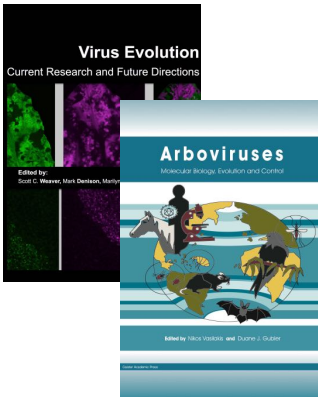
In prokaryotes, one of the best-studied examples of regulatory proteins that use DNA-looping is the NtrC family of enhancer-binding proteins (EBPs). Binding of the activated form of an EBP to its target sequence and DNA-looping place it in contact with a special form of RNA polymerase (RNAP) containing the σ^N subunit (Buck *et al.*, 2000; Rombel *et al.*, 1998). Sigma factors (σ) are subunits of RNAP that are required for promoter recognition and initiation of transcription. Among them σ^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 σ^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.

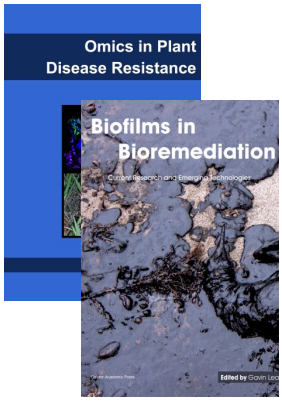
Recently there have been exciting developments in elucidating the molecular mechanism of enhancer-dependent transcription (e.g. Chaney *et al.*, 2001); but our understanding of the evolutionary origin and biological significance of this mode of regulation in bacteria still lags behind. The availability 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

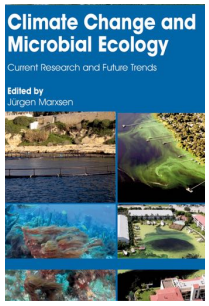
Further Reading

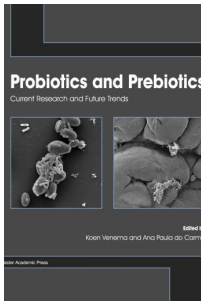
Caister Academic Press is a leading academic publisher of advanced texts in microbiology, molecular biology and medical research. Full details of all our publications at **caister.com**

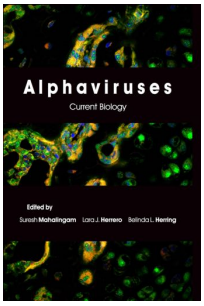
- **MALDI-TOF Mass Spectrometry in Microbiology**
  Edited by: M Kostrzewa, S Schubert (2016)
  www.caister.com/malditof
- ***Aspergillus* and *Penicillium* in the Post-genomic Era**
  Edited by: RP Vries, IB Gelber, MR Andersen (2016)
  www.caister.com/aspergillus2
- **The Bacteriocins: Current Knowledge and Future Prospects**
  Edited by: RL Dorit, 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 Tsaloglou (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 Romaní, H Guasch, MD Balaguer (2016)
  www.caister.com/aquaticbiofilms
- **Alphaviruses: Current Biology**
  Edited by: S Mahalingam, L Herrero, B Herring (2016)
  www.caister.com/alpha
- **Thermophilic Microorganisms**
  Edited by: F Li (2015)
  www.caister.com/thermophile

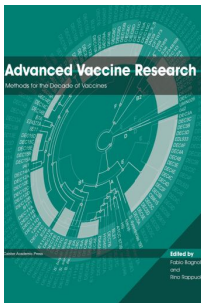


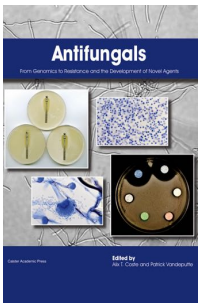












- **Flow Cytometry in Microbiology: Technology and Applications**
  Edited by: MG Wilkinson (2015)
  www.caister.com/flow
- **Probiotics and Prebiotics: Current Research and Future Trends**
  Edited by: K Venema, AP Carmo (2015)
  www.caister.com/probiotics
- **Epigenetics: Current Research and Emerging Trends**
  Edited by: BP Chadwick (2015)
  www.caister.com/epigenetics2015
- ***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**
  Edited by: J Murillo, BA Vinatzer, RW Jackson, et al. (2015)
  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/microarrays2
- **Metagenomics of the Microbial Nitrogen Cycle: Theory, Methods and Applications**
  Edited by: D Marco (2014)
  www.caister.com/n2

Order from caister.com/order

the σ^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 σ^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 σ^N-dependent EBP (Table 1). Seven of these proteins have been well studied in *Salmonella* or *E. coli*: RtcR, NtrC, HydG, FhlA, 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 σ^N-RNAP (Buck *et al.*, 2000; Xu and Hoover, 2001). Therefore each transcriptional target of an EBP must be linked to a σ^N-RNAP-binding DNA sequence. Binding sites for σ^N-RNAP are fairly easy to detect and have the consensus sequence YTGG-CACGrNNNTTGCW (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 σ^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 σ^N-dependent enhancer-binding proteins (EBPs) identified from complete genome sequences of enteric bacteria.

σ^N-dependent enhancer binding protein (EBP)				Proposed functions of genes regulated by EBP	References
S. typhimurium	*S. Typhi*	*E. coli*	*Y. pestis*		
NtrC	NtrC	NtrC	NtrC	Nitrogen assimilation	Reitzer and Magasanik, 1986.
PspF	PspF	PspF	PspF	Phage shock response	Model *et al.*, 1997.
YfhA	YfhA	YfhA	YfhA	Putative 2–component system	Unpublished.
YgaA	YgaA	YgaA	–	Putative flavoprotein and reductase	Gardner *et al.*, 2002.
HydG	HydG	HydG	–	Tolerance to Zn^{2+} and Pb^{2+}	Leonhartsberger *et al.*, 2001.
PrpR	PrpR	PrpR	–	Propionate catabolism	Palacios and Escalante–Semerena, 2000.
FhlA	FhlA	FhlA	–	Formate–hydrogen lyase, hydrogenase	Schlensog and Bock, 1990.
PgtA	PgtA	–	–	Phosphoglycerate transport	Jiang *et al.*, 1998.
RtcR	–	RtcR	–	RNA processing	Genschik *et al.*, 1998.
STM0652	STY0703	–	–	Hexuronate utilisation	Unpublished.
STM2361	STY2591	–	–	Amino acid transport	Unpublished.
STM3773	–	–	–	Putative PTS system EII transport protein	Unpublished.
STM0571	–	–	–	Putative PTS system EII transport protein	Unpublished.
STM4534	–	–	–	Putative PTS system EII transport protein, chemoreceptor	Unpublished.
–	–	AtoC	–	Short chain fatty acid utilisation	Jenkins and Nunn, 1987.
–	–	HyfR	–	Formate hydrogenlyase	Andrews *et al.*, 1997.
–	–	YgeV	–	Unknown, possibly quorum sensing	De Lisa *et al.*, 2001.
–	–	DhaR	–	Dihydroxyacetone utilisation	Paulsen *et al.*, 2000.
–	–	–	FleR	Flagellar motility	Unpublished.

Table 2. Close matches to the canonical σ^N-RNAP-binding sequence in the complete genome sequence of *Salmonella typhimurium*. **A**. 'Orphan' sites not linked to an nearby EBP gene. **B**. Promoters known or proposed to be activated by specific EBPs.

A

Score (Kullback-Liebler distance)	Promoter sequence	Gene name	Gene product/function
2800	TGGCACGGTTGTTGCT	*rpoH*	Heat shock sigma factor σ^H
2777	TGGCACGTCTATTGCT	*gltI*	Glutamate/aspartate transporter
2688	TGGCACGCTGGTTGCA	STM3521	Putative ribonucleotide related protein
2582	TGGCATGAGAGTTGCT	*yeaG*	Protein kinase, similar to PrkA
2524	TGGCACGAATGCTGCA	*astC*	Succinylornithine transaminase, arginine metabolism
2507	TGGCAGATATTTTGCT	*yjeB*	Putative negative regulator
2461	TGGGACAGCCGTTGCT	STM2746	Putative excinuclease ATPase subunit, similar to RecN.
2439	TGGCAACAGAATTGCT	*topA*	DNA topoiomerase I
2426	TGGCATAGGCCTTGCG	STM0031	Similar to MarT and ToxR transcriptional regulators
2408	TGGCGCGTTTATTGCC	*cspB*	Cold shock protein
2400	TGGCTCGTAGCTTGCC	*acrD*	Multidrug transporter
2400	TGTCACGTATTTTGCA	*clpP*	ClpP protease
2389	TGGCACGGAGTTAGCC	*yrdA*	Putative siderophore-binding protein
2389	TGGCACGGAGTTAGCC	*yieP*	Putative regulatory protein, possibly involved in gluconate utilisation
2389	TGGCACGGAGTTAGCC	*purH*	Purine biosynthesis
2370	TGGCATCCCCTTTGCG	*ytfJ*	Putative transcriptional regulator
2344	TGGAACGCCTCCTGCT	*dnaT*	Primosomal protein
2333	TGGCAAGGTCTATGCA	*ygjT*	Putative heavy metal transport/resistance protein
2329	TGGCACTGTAGTGGCA	*pagK*	PhoPQ-activated gene
2326	TGGCACCCTGGGTGCA	*yfcZ*	Putative protein
2323	TGGCACGCGCGTCGCC	STM2690	Putative outer membrane efflux protein

B

Score (Kullback-Liebler distance)	Promoter sequence	Gene name	Gene product/function	Associated EBP
2769	TGGCACGCCGCTTGCT	STM4535	Putative PTS EII component	STM4534
2683	TGGCACAACCTTTGCT	STM3772	Putative PTS EII component	STM3773
2653	TGGCACGCCGTTTGCC	STM0577	Putative PTS EII component	STM0571
2649	TGGCATGCCTTTTGCT	STM2360	Putative diaminopimelate decarboxylase	STM2361
2625	TGGCACATCCTTTGCA	*glnK*	Nitrogen regulatory protein	NtrC
2551	TGGCATAGCCTTTGCT	*prpB*	Propionate metabolism	PrpR
2550	TGGCATAAATATTGCT	*hypA*	Hydrogenase accessory protein	FhlA
2514	TGGCACGGAAGATGCA	*zraP*	Resistance to zinc and lead	HydG
2484	TGGCACGCCTTTTGAT	STM0649	Putative hydrolase	STM0652
2431	TGGCACAATTACTGCA	*yfhK*	Histidine sensor kinase	YfhA
2426	TGGCACGATAGTCGCA	*hisJ*	Histidine transport protein	STM2361
2407	TGGCACGCAAATTGTA	*pspA*	Phage shock response	PspF
2390	TGGCATGGAAAATGCT	*hycA*	Transcriptional repressor of formate hydrogenlyase	FhlA
2383	TGGCACACTAGCTGCA	STM2840	Putative flavoprotein	YgaA
2354	TGGCATGATCTCTGCT	*hydH*	Histidine sensor kinase, zinc/ lead resistance	HydG

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-

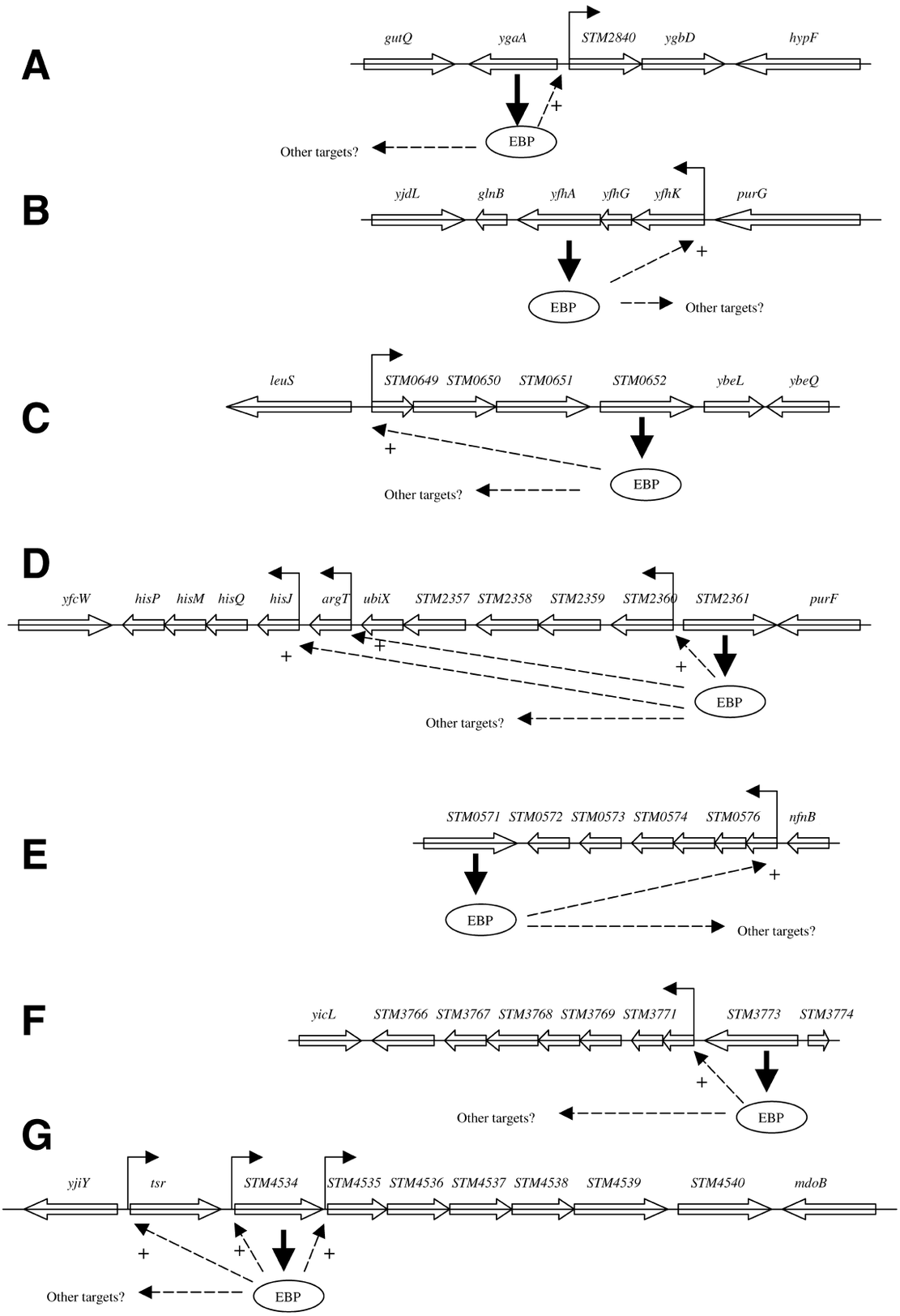


Figure 1. Operon structures of 7 genes encoding uncharacterised EBPs in *Salmonella typhimurium*. Predicted σ^N-dependent promoters are indicated (⌜).

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 σ^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 σ^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 Zhulin, 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 (EII) 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 σ^N-RNAP (e.g. Merrick and Coppard, 1989), a link possibly allowing coordination of carbon and nitrogen metabolism. However, the identification of σ^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 phosphorylatable 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* (Hechard *et al.*, 2001) five different LevR-like proteins each activate transcription of PTS EII components (Hechard *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 Enteritidis 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 bacterium. 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 σ^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' σ^N-Dependent Promoters

Several high-scoring potential σ^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

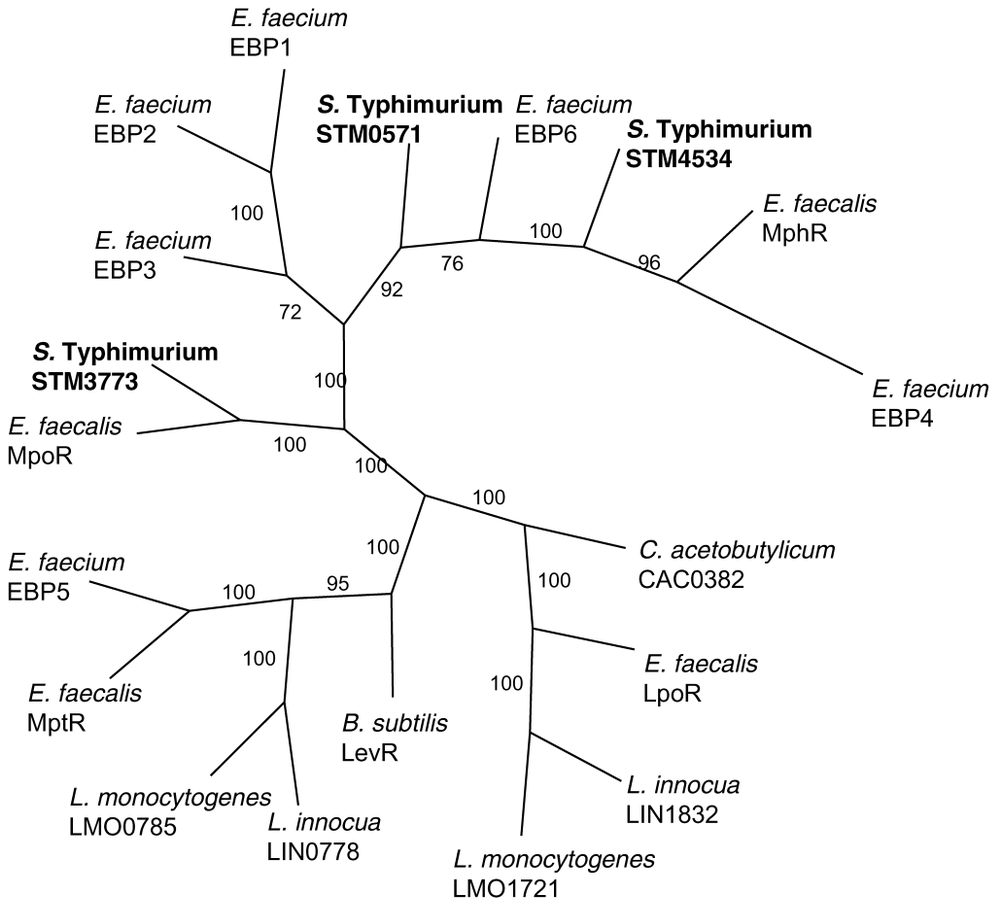


Figure 2. Phylogenetic tree of LevR-like protein sequences. Preliminary sequence data were obtained from The Institute for Genomic Research website at *http://www.tigr.org*. Neighbour joining analysis was carried out on 100 bootstrapped data sets using the PHYLIP computer package.

σ^N-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 σ^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, σ^N-RNAP binding sites may act as negative regulatory elements, as is the case for the negatively autoregulated *rpoN* gene in *Pseudomonas putida* (Köhler *et al*., 1994).

Pallen (1999) first noticed a potential σ^N-RNAP binding site upstream of *rpoH* (encoding the heat shock sigma factor σ^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 σ^N-RNAP strongly. In *E. coli* growing on nutrient-poor media, deletion of the *rpoN* gene (encoding σ^N) led to loss of transcription from the σ^H-dependent *ibpAB* promoter (Kucyńska-Wisńik *et al.,* 2001), further suggesting that σ^Hmay be dependent on σ^N under certain conditions. There is a strong potential σ^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 σ^H. Whether topA is additionally regulated by σ^N-RNAP, and under what conditions, remains to be tested. Nevertheless these observations, along with the established role of σ^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 σ^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

transcriptional start site upstream of *cspB*. However, it is still possible that the potential σ^N-RNAP binding site identified here is functional under some other conditions. In *Pseudomonas aeruginosa*, PA0961, encoding a protein 54% identical to CspB, is also preceded by a potential σ^N-RNAP (unpublished results) lending further support for biological significance.

Several of the potential σ^N-dependent promoters are associated with virulence factors in *Salmonella*. The σ^N-dependent *psp* operon, activated by PspF, is involved in virulence in *Y. enterocolitica*, *pspC* being required for normal growth when the Ysc type III secretion system was expressed in the laboratory (Darwin and Miller, 2001). The *psp* operon is conserved in *Salmonella.* It remains to be tested whether *pspC* is required for activity of the two type III secretion systems in *S. typhimurium*. Other genes implicated in *Salmonella* virulence that have potential σ^N-dependent promoters include *clpP* (Yamamoto *et al*., 2001), *yrdA* (encoding a siderophore-binding protein) and *pagK* (Gunn *et al*., 1998). Furthermore, *acrD,* encoding a multidrug resistance protein (Nishino *et al*., 2001) and *STM2690*, encoding an outer membrane efflux pump, have potential σ^N-dependent promoters. It would therefore be interesting to test whether deletion of the *Salmonella rpoN* gene had any effect on virulence as it does, for example, in *P. aeruginosa* (Hendrickson *et al*., 2001).

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 σ^N-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 regulon proposed here include a two component signal transduction system, hexuronate utilisation, an uncharacterised oxidoreductase system, ClpP protease, and three PTS EII components. The repertoire of functions regulated by σ^N-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

Andrews, S.C., Berks, B.C., McClay, J., Ambler, A., Quail, M.A., Golby, P., and Guest, J.R. 1997. A 12-cistron *Escherichia coli* operon (*hyf*) encoding a putative proton-translocating formate hydrogenlyase system. Microbiology 143: 3633–3647.

Barrios, H., Valderrama, B., and Morett, E. 1999. Compilation and analysis of σ^{54}-dependent promoter sequences. Nucleic Acids Res. 27: 4305–4313.

Buck, M., Gallegos, M.T., Studholme, D.J., Guo, Y., and Gralla, J.D. 2000. The bacterial enhancer-dependent σ^{54} (σ^N) transcription factor. J. Bacteriol. 182: 4129–4136.

Cases, I., and de Lorenzo, V. 2001. The limits to genomic predictions: role of σ^N in environmental stress survival of *Pseudomonas putida*. FEMS Microbiol. Ecol. 35: 217–221.

Chaney, M., Grande, R., Wigneshweraraj, S.R., Cannon, W., Casaz, P., Gallegos, M.T., Schumacher, J., Jones, S., Elderkin, S., Dago, A.E., Morett, E., and Buck, M. 2001. Binding of transcriptional activators to σ^{54} in the presence of the transition state analog ADP-aluminum fluoride: insights into activator mechanochemical action. Genes Dev. 15: 2282–2294.

Craig, J.E., Boyle, D., Francis, K.P., and Gallagher, M.P. 1998. Expression of the cold-shock gene *cspB* in *Salmonella typhimurium* occurs below a threshold temperature. Microbiology. 144: 697–704.

Darwin, A.J., and Miller, V.L. 2001. The *psp* locus of *Yersinia enterocolitica* is required for virulence and for growth in vitro when the Ysc type III secretion system is produced. Mol. Microbiol. 39: 429–444.

De Lisa, M.P., Wu, C.F., Wang, L., Valdes, J.J., and Bentley, W.E. 2001. DNA microarray-based identification of genes controlled by autoinducer 2-stimulated quorum sensing in *Escherichia coli*. J. Bacteriol. 183: 5239–5247.

Debarbouille, M., Martin-Verstraete, I., Klier, A., and Rapoport, G. 1991. The transcriptional regulator LevR of *Bacillus subtilis* has domains homologous to both σ^{54}- and phosphotransferase system-dependent regulators. Proc. Natl. Acad. Sci. USA. 88: 2212–2216.

Fischer, C., Geourjon, C., Bourson, C., and Deutscher, J. 1996. Cloning and characterization of the *Bacillus subtilis prkA* gene encoding a novel serine protein kinase. Gene. 168: 55–60.

Gardner, A.M., Helmick, R.A., and Gardner, P.R. 2002. Flavorubredoxin: an inducible catalyst for nitric oxide reduction and detoxification in *Escherichia coli*. J. Biol. Chem. 277: 8172–8177.

Genschik, P., Drabikowski, K., and Filipowicz, W. 1998. Characterization of the *Escherichia coli* RNA 3'-terminal phosphate cyclase and its σ^{54}-regulated operon. J. Biol. Chem. 273: 25516–25526.

Glaser, P., Frangeul, L., Buchrieser, C., Rusniok, C., Amend, A., Baquero, F., Berche, P., Bloecker, H., Brandt, P., Chakraborty, T., Charbit, A., Chetouani, F., Couve, E., de Daruvar, A., Dehoux, P., Domann, E., Dominguez-Bernal, G., Duchaud, E., Durant, L., Dussurget, O., Entian, K.D., Fsihi, H., Portillo, F.G., Garrido, P., Gautier, L., Goebel, W., Gomez-Lopez, N., Hain, T., Hauf, J., Jackson, D., Jones, L.M., Kaerst, U., Kreft, J., Kuhn, M., Kunst, F., Kurapkat, G., Madueno, E., Maitournam, A., Vicente, J.M., Ng, E., Nedjari, H., Nordsiek, G., Novella, S., de Pablos, B., Perez-Diaz, J.C., Purcell, R., Remmel, B., Rose, M., Schlueter, T., Simoes, N., Tierrez, A., Vazquez-Boland, J.A., Voss, H., Wehland, J., and Cossart, P. 2001. Comparative genomics of *Listeria* species. Science. 294: 849–852.

Gunn, J.S., Belden, W.J., and Miller, S.I. 1998. Identification of PhoP-PhoQ activated genes within a duplicated region of the *Salmonella typhimurium* chromosome. Microb. Pathog. 25: 77–90.

Hechard, Y., Pelletier, C., Cenatiempo, Y., and Frere, J. 2001. Analysis of σ^{54}-dependent genes in *Enterococcus faecalis*: a mannose PTS permease (EII^{Man}) is involved in sensitivity to a bacteriocin, mesentericin Y105. Microbiology. 147: 1575–1580.

Hendrickson, E.L., Plotnikova, J., Mahajan-Miklos, S., Rahme, L.G., and Ausubel, F.M. 2001. Differential Roles of the *Pseudomonas aeruginosa* PA14 *rpoN* Gene in Pathogenicity in Plants, Nematodes, Insects, and Mice. J. Bacteriol. 183: 7126–7134.

Hübner, A., Yang, X., Nolen, D.M., Popova, T.G., Cabello, F.C., and Norgard, M.V. 2001. Expression of *Borrelia burgdorferi* OspC and DbpA is controlled by a RpoN-RpoS regulatory pathway. Proc. Natl. Acad. Sci. USA. 98: 12724–12729.

Jenkins, L.S., and Nunn, W.D. 1987. Regulation of the *ato* operon by the *atoC* gene in *Escherichia coli*. J. Bacteriol. 169: 2096–2102.

Jiang, S.Q., Yu, G.Q., Li, Z.G., and Hong, J.S. 1988. Genetic evidence for modulation of the activator by two regulatory proteins involved in the exogenous induction of phosphoglycerate transport in *Salmonella typhimurium*. J. Bacteriol. 170: 4304–4308.

Köhler, T., Alvarez, J.F., and Harayama, S. 1994. Regulation of the *rpoN*, ORF102 and ORF154 genes in *Pseudomonas putida*. FEMS Microbiol. Lett. 115: 177–184.

Kuczyńska-Wisńik, D., Laskowska, E., and Taylor, A. 2001. Transcription of the *ibpB* heat-shock gene is under control of σ^{32}- and σ^{54}-promoters, a third regulon of heat-shock response. Biochem. Biophys. Res. Commun. 284: 57–64.

Kustu, S.G., McFarland, N.C., Hui, S.P., Esmon, B., and Ames, G.F. 1979. Nitrogen control of *Salmonella typhimurium*: co-regulation of synthesis of glutamine synthetase and amino acid transport systems. J. Bacteriol. 138: 218–234.

Ledebur, H., Gu, B., Sojda, J. 3rd, and Nixon, B.T. 1990. *Rhizobium meliloti* and *Rhizobium leguminosarum dctD* gene products bind to tandem sites in an activation sequence located upstream of σ^{54}-dependent *dctA* promoters. J. Bacteriol. 172: 3888–3897.

Leonhartsberger, S., Huber, A., Lottspeich, F., and Bock, A. 2001. The *hydH/G* genes from *Escherichia coli* code for a zinc and lead responsive two-component regulatory system. J. Mol. Biol. 307: 93–105.

Marques, S., Gallegos, M.T., Manzanera, M., Holtel, A., Timmis, K.N., and Ramos, J.L. 1998. Activation and repression of transcription at the double tandem divergent promoters for the *xylR* and *xylS* genes of the TOL plasmid of *Pseudomonas putida*. J. Bacteriol. 180: 2889–2894.

Martin-Verstraete, I., Charrier, V., Stulke, J., Galinier, A., Erni, B., Rapoport, G., and Deutscher, J. 1998. Antagonistic effects of dual PTS-catalysed phosphorylation on the *Bacillus subtilis* transcriptional activator LevR. Mol. Microbiol. 28: 293–303.

Mathews, S.A., and Timms, P. 2000. Identification and mapping of sigma-54 promoters in *Chlamydia trachomatis*. J. Bacteriol. 182: 6239–6242.

McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W., Latreille, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F., Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A., Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Florea, L., Miller, W., Stoneking, T., Nhan, M., Waterston, R., and Wilson, R.K. 2001. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. Nature. 413: 852–856.

Merrick, M.J. 1993. In a class of its own–the RNA polymerase sigma factor σ^{54} (σ^{N}). Mol. Microbiol. 10: 903–909.

Merrick, M.J., and Coppard, J.R. 1989. Mutations in genes downstream of the *rpoN* gene (encoding σ^{54}) of *Klebsiella pneumoniae* affect expression from σ^{54}-dependent promoters. Mol. Microbiol. 3: 1765–1775.

Model, P., Jovanovic, G., and Dworkin, J. 1997. The *Escherichia coli* phage-shock-protein (*psp*) operon. Mol. Microbiol. 24: 255–261.

Morett, E., and Segovia, L. 1993. The σ^{54} bacterial enhancer-binding protein family: mechanism of action and phylogenetic relationship of their functional domains. J. Bacteriol. 175: 6067–6074.

Neuwald, A.F., Aravind, L., Spouge, J.L., and Koonin, E.V. 1999. AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. Genome Res. 9: 27–43.

Nishino, K., and Yamaguchi, A. 2001. Analysis of a complete library of putative drug transporter genes in *Escherichia coli*. J. Bacteriol. 183: 5803–5812.

Niu, S., Jiang, S.Q., and Hong, J. 1995. *Salmonella typhimurium pgtB* mutants conferring constitutive expression of phosphoglycerate transporter *pgtP* independent of *pgtC*. J. Bacteriol. 177: 4297–4302.

Nolling, J., Breton, G., Omelchenko, M.V., Makarova, K.S., Zeng, Q., Gibson, R., Lee, H.M., Dubois, J., Qiu, D., Hitti, J., Wolf, Y.I., Tatusov, R.L., Sabathe, F., Doucette-Stamm, L., Soucaille, P., Daly, M.J., Bennett, G.N., Koonin, E.V., and Smith, D.R. 2001. Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. J. Bacteriol. 183: 4823–4838.

Palacios, S., and Escalante-Semerena, J.C. 2000. *prpR*, *ntrA*, and *ihf* functions are required for expression of the *prpBCDE* operon, encoding enzymes that catabolize propionate in *Salmonella enterica* serovar Typhimurium LT2. J. Bacteriol. 182: 905–910.

Pallen, M. 1999. RpoN-dependent transcription of rpoH? Mol. Microbiol. 31: 393.

Paulsen, I.T., Reizer, J., Jin, R.Z., Lin, E.C., and Saier, M.H., Jr. 2000. Functional genomic studies of dihydroxyacetone utilization in *Escherichia coli*. Microbiology 146: 2343–2344.

Pohlmann, A., Cramm, R., Schmelz, K., and Friedrich, B. 2000. A novel NO-responding regulator controls the reduction of nitric oxide in *Ralstonia eutropha*. Mol. Microbiol. 38: 626–638.

Postma, P.W., Lengeler, J.W., Jacobson, G.R. 1993. Phosphoenolpyruvate: carbohydrate phosphotransferase systems of bacteria. Microbiol. Rev. 57: 543–594.

Reitzer, 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.

Reizer, J., and Saier, M.H. Jr. 1997. Modular multidomain phosphoryl transfer proteins of bacteria. Curr. Opin. Struct. Biol. 7: 407–415.

Rippe, K., von Hippel, P.H., and Langowski, J. 1995 Action at a distance: DNA–looping and initiation of transcription. Trends Biochem. Sci. 20: 500–506.

Rombel, I., North, A., Hwang, I., Wyman, C., and Kustu, S. 1998. The bacterial enhancer-binding protein NtrC as a molecular machine. Cold Spring Harb. Symp. Quant. Biol. 63: 157–166.

Saier, M.H. Jr., and Reizer, J. 1994. The bacterial phosphotransferase system: new frontiers 30 years later. Mol. Microbiol. 13: 755–764.

Schlensog, V., and Bock, A. 1990. Identification and sequence analysis of the gene encoding the transcriptional activator of the formate hydrogenlyase system of *Escherichia coli*. Mol. Microbiol. 4: 1319–1327.

Schmitz, G., Nikaido, K., and Ames, G.F. 1988. Regulation of a transport operon promoter in *Salmonella typhimurium*: identification of sites essential for nitrogen regulation. Mol. Gen. Genet. 215: 107–117.

Siebold, C., Flukiger, K., Beutler, R., and Erni, B. 2001. Carbohydrate transporters of the bacterial phosphoenolpyruvate: sugar phosphotransferase system (PTS). FEBS Lett. 504: 104–111.

Stock, A.M., Robinson, V.L.., and Goudreau, P.N. 2000. Two-component signal transduction. Annu. Rev. Biochem. 69: 183–215.

Stormo, G.D. 2000. DNA binding sites: representation and discovery. Bioinformatics. 16: 16–23.

Studholme, D.J., and Buck, M. 2000. The biology of enhancer-dependent transcriptional regulation in bacteria: insights from genome sequences. FEMS Microbiol. Lett. 186: 1–9.

Studholme, D.J., Buck, M., and Nixon, B.T. 2000. Identification of potential σ^{N}-dependent promoters in bacterial genomes. Microbiology. 146: 3021–3023.

Taylor, B.L.., and Zhulin, I.B. 1999. PAS domains: internal sensors of oxygen, redox potential, and light. Microbiol. Mol. Biol. Rev. 63: 479–506.

Tchieu, J.H., Norris, V., Edwards, J.S., and Saier, M.H., Jr. 2001. The complete phosphotransferase system in *Escherichia coli*. J. Mol. Microbiol. Biotechnol. 3: 329–346.

Tse-Dinh, Y.C., Qi, H., and Menzel, R. 1997. DNA supercoiling and bacterial adaptation: thermotolerance and thermoresistance. Trends Microbiol. 5: 323–326.

Wu, S.Q., Chai, W., Lin, J.T., and Stewart, V. 1999. General nitrogen regulation of nitrate assimilation regulatory gene *nasR* expression in *Klebsiella oxytoca* M5al. J. Bacteriol. 181: 7274–7284.

Xu, H., and Hoover, T.R. 2001. Transcriptional regulation at a distance in bacteria. Curr. Opin. Microbiol. 4: 138–44.

Yamamoto, T., Sashinami, H., Takaya, A., Tomoyasu, T., Matsui, H., Kikuchi, Y., Hanawa, T., Kamiya, S., and Nakane, A. 2001. Disruption of the genes for ClpXP protease in *Salmonella enterica* serovar Typhimurium results in persistent infection in mice, and development of persistence requires endogenous gamma interferon and tumor necrosis factor alpha. Infect Immun. 69: 3164–3174.