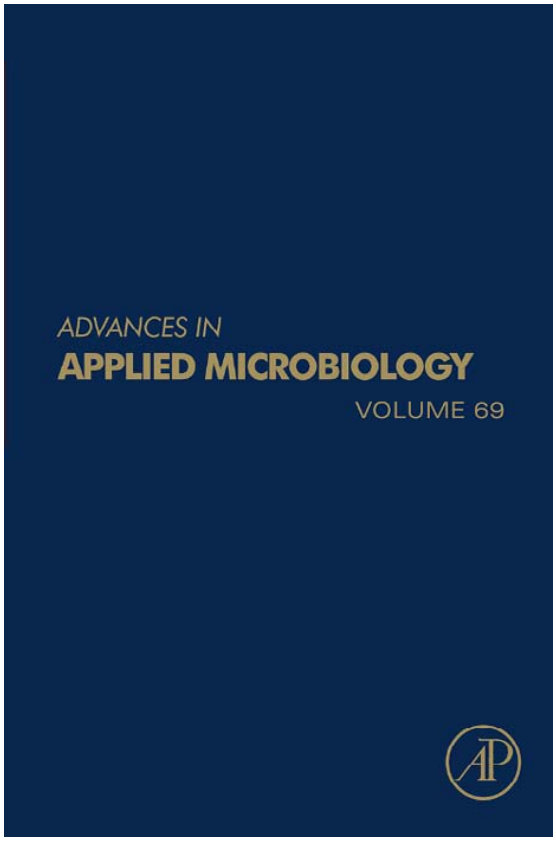
Provided for non-commercial research and educational use only.
Not for reproduction, distribution or commercial use.

This chapter was originally published in the book ADVANCES IN APPLIED MICROBIOLOGY, Vol. 69, published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues who know you, and providing a copy to your institution's administrator.



All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:
http://www.elsevier.com/locate/permissionusematerial

From: Paul A. Hoskisson and Se´bastien Rigali, Variation in Form and Function: The Helix-Turn-Helix Regulators of the GntR Superfamily.
In Allen I. Laskin, Sima Sariaslani and Geoffrey M. Gadd editor:
ADVANCES IN APPLIED MICROBIOLOGY, Vol. 69,
Burlington: Academic Press, 2009, pp.1-22.
ISBN: 978-0-12-374824-9
© Copyright 2009 Elsevier Inc.
Academic Press.

Author's personal copy

CHAPTER 1

Variation in Form and Function: The Helix-Turn-Helix Regulators of the GntR Superfamily

Paul A. Hoskisson* and **Sébastien Rigali†**

Contents

Abstract

One of the most abundant and widely distributed groups of Helix-turn-helix (HTH) transcription factors is the metabolite-responsive GntR family of regulators (>8500 members in the Pfam database; Jan 2009). These proteins contain a DNA-binding HTH domain at the N terminus of the protein and an effector-binding and/or

* Strathclyde Institute of Pharmacy and Biological Science, University of Strathclyde, Royal College Building, George Street, Glasgow, United Kingdom
† Centre d'Ingénierie des Protéines, Université de Liège, Institut de Chimie B6, Sart-Tilman, B-4000 Liège, Belgium

Advances in Applied Microbiology, Volume 69
ISSN 0065-2164, DOI: 10.1016/S0065-2164(09)69001-8

© 2009 Elsevier Inc.
All rights reserved.

Author's personal copy

oligomerisation domain at the C terminus, where upon on binding an effector molecule, a conformational change occurs in the protein which influences the DNA-binding properties of the regulator resulting in repression or activation of transcription. This review summarises what we know about the distribution, structure, function and classification of these regulators and suggests that they may have a future role in biotechnology.

...endless forms most beautiful and most wonderful have been and are being, evolved.

Charles Darwin, 1859

I. INTRODUCTION

As bacteria sense different micro-environments, they modify their gene expression appropriately to enable them to respond to the prevailing conditions. Often the signal sensed within the cell is a metabolic intermediate, and these are sensed by many classes of helix-turn-helix (HTH) transcription factor, through which they modulate gene expression. The effector molecules bound by these proteins are often related catabolic substrates, substrates and/or intermediates of the pathway controlled by the transcription factor.

One of the most abundant groups of HTH bacterial metabolite-responsive transcription factors is the GntR family of regulators (>8500 members in the Pfam database; Jan 2009). These multi-domain transcription factors are widely distributed throughout the bacterial world where they play a fundamental role in modulation of gene expression to respond appropriately to the environment context.

This review aims to bring together and summarise our current thinking on GntR regulators, their structure, function, evolution, and how they may be exploited in biotechnology.

II. HELIX-TURN-HELIX DNA-BINDING PROTEINS

The identification of a tri-helical domain and its critical role in DNA binding within the bacteriophage Lambda proteins, cI and cro and the *lac* operon repressor, LacI, were early advances in the pioneering work of Matthews and co-workers (Ohlendorf *et al.*, 1982, 1983) and Sauer *et al.* (1982). The importance of helix two and helix three of the domain led to the identification of what became known as the HTH motif. The third α-helix is often referred to as the 'recognition' helix, which fits within

Author's personal copy

the major groove of the DNA mediating the protein–DNA interaction (Aravind *et al.*, 2005). Ohlendorf *et al.* (1983) and Sauer *et al.* (1982) suggested, through extensive sequence analysis and secondary structure analysis, that this domain was present in several DNA-binding bacterial activators and repressors, and they hypothesised that these domains descended from a common ancestor.

Throughout the 1980s and 1990s, extensive sequencing, the emergence of whole genome sequencing and experimental work confirmed the ubiquity and central role this domain played in gene regulation in both prokaryotes and eukaryotes and led to the identification of the HTH motif in all domains of life, suggesting that the HTH domain is one of the most ancient protein folds, although it appears to be most prevalent in prokaryotes (Aravind and Koonin, 1999). The development of specific algorithms for recognition of the HTH motif has become indispensible in genome annotation such as that of Dodd and Egan (1990), enabling rapid identification of HTH-containing proteins.

III. GntR REGULATORS

The HTH-containing GntR family is widely distributed throughout the bacteria where they regulate many diverse biological processes. It was named GntR after the first member identified, the *Bacillus subtilis* repressor of the gluconate operon (Haydon and Guest, 1991; Prosite Family PS50949; Pfam family: PF00392). GntR regulators are often located on the chromosome adjacent to the genes that they control, which in many cases allows insight into the metabolites that they may bind. There are however many examples where this is not the case, and identifying their cognate ligands remains a significant barrier to understanding their function.

In general, these proteins contain a DNA-binding HTH domain at the N terminus of the protein and an effector-binding and/or oligomerisation domain at the C terminus (Fig. 1.1). Upon binding an effector molecule at the C-terminal domain, a conformational change occurs in the protein which influences the DNA-binding properties of the regulator resulting in repression or activation of transcription. The DNA-binding domain is conserved throughout the GntR family yet the regions outside the DNA-binding domain are more variable; however, this is not surprising given the diversity of molecules that they bind, and this feature is used to define the GntR-like sub-families (Rigali *et al.*, 2002). Despite the large number of GntR-like regulators identified there are few examples where their effector molecules are known and the complete regulatory circuitry elucidated. Knowledge of this is of particular importance where GntR-like regulators control genes of unknown biochemical function and can

Author's personal copy

FIGURE 1.1 Schematic representation of a GntR protein. Indicates the N-terminal helix-turn-helix DNA-binding domain and the longer C-terminal effector-binding/oligomerisation domain (E-b/O).

provide information of their cellular function and will enable these processes to be built into modelling frameworks in terms of using systems biology approaches. GntR-like regulators are known to control many fundamental cellular processes such as motility (Jaques and McCarter, 2006), development (Hoskisson *et al.*, 2006), antibiotic production (Hillerich and Westpheling, 2006), antibiotic resistance (Truong-Bolduc and Hooper, 2007), Plasmid transfer (Reuther *et al.*, 2006) and virulence (Casali *et al.*, 2006; Haine *et al.*, 2005). In all these cases the exact ligand regulating gene expression through these proteins is unknown.

There are many cases where GntR-like regulators are not located next to genes that they control (orphan regulators), or without their effectors they are activators of gene expression elsewhere in the genome. One well-studied example is FadR, the fatty acid metabolism regulator in *Escherichia coli*, where it is known to negatively control 12 genes or operons and activate transcription of at least three genes when a fatty acid precursor is bound (DiRusso *et al.*, 1993; See section VIII).

The identification of the small molecules that bind to these regulators has traditionally been difficult and has mainly relied on gene context and bioinformatics to identify possible effector molecules. This area remains a significant challenge to researchers in this field and urgently requires novel methods to aid identification of effector molecules.

IV. DISTRIBUTION OF GntR REGULATORS

Examination and analysis of GntR regulator distribution throughout completely sequenced genomes demonstrate some interesting trends in terms of their abundance and may give clues to how an organism is distributed in a particular ecological niche or the kind of plasticity it experiences within its natural environment.

There are 8561 GntR regulators in the Pfam database (PfAM family GntR: PF00392: Finn *et al.*, 2008). The bulk of these (8561 sequences) are found in 764 bacterial taxa indicating that this protein fold has been widely adopted as a regulatory mechanism. Examination of taxonomic distribution of these regulators throughout the bacteria demonstrates a wide distribution; however, the predominant phyla (from current

Author's personal copy

sequences available in Pfam) are the Proteobacteria, Firmicutes and the Actinobacteria (Fig. 1.2). Detailed examination of the distribution within well-characterised species (Fig. 1.3) shows an interesting trend, not only with increasing genome size, but also with ecological niche. The trend suggests that organisms that live in complex, highly variable environments such as soil (e.g. *Streptomyces*, *Burkholderia*, *Rhizobium*) have a larger complement of the metabolite-responsive GntR regulators than obligate intracellular parasites and endosymbionts (e.g. *Chlamydia* and *Buchnera*). This trend is reinforced even within genera with *Mycobacterium smegmatis* having a complement of about 60 GntR regulators (Vindal *et al.*, 2007), where all these have been lost in the obligate intracellular pathogen *Mycobacterium leprae* during the extensive gene decay observed in this species (Cole *et al.*, 2001). These data indicate that, whilst there is a trend to increase metabolite-responsive regulators in the genome to enable rapid response to changing conditions in complex environments, this is lost when a stable niche is occupied and this requirement ameliorated.

There are twelve GntR regulators known in the Archaea. Two known from Eukaryotes, one from the sea anemone (*Nematostella vectensis*) and one from *Trichomonas vaginalis*; however, the exact functions of these are unknown. The two known GntRs in viruses are both in bacteriophages,

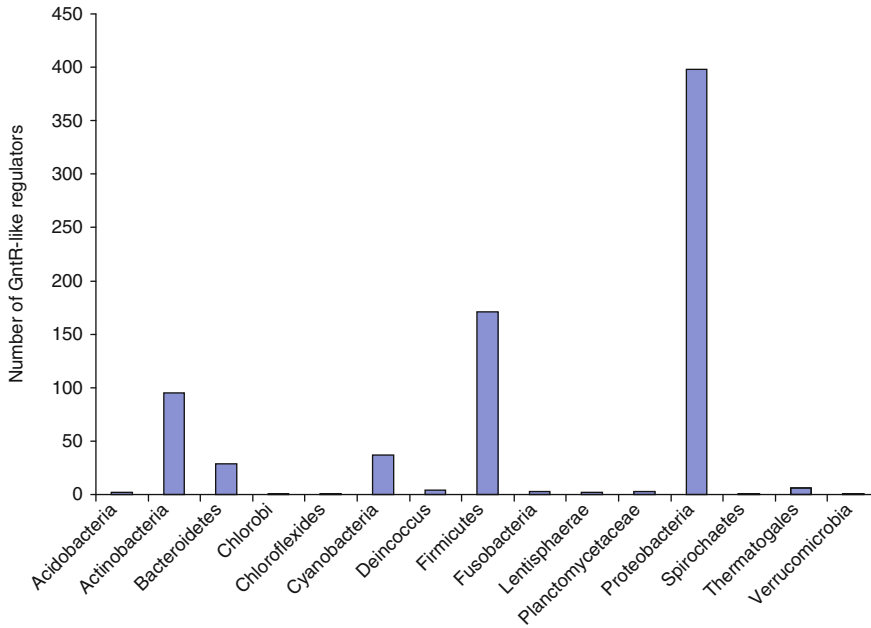


FIGURE 1.2 Distribution of GntR proteins throughout the bacterial Phyla. Please see text in Section IV. Data were taken from the sequences deposited in the Pfam database.

Author's personal copy

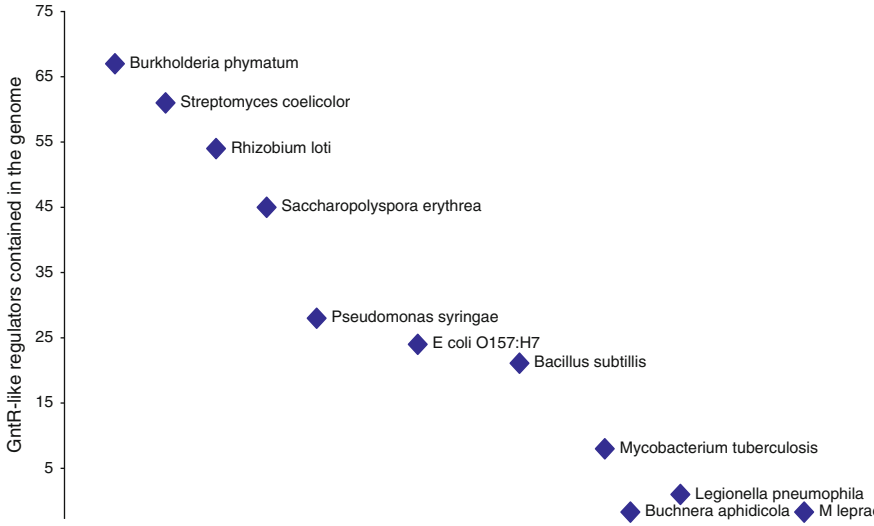


FIGURE 1.3 Distribution of GntR regulators in selected bacterial whole genomes. See text in Section V.

one in the *Streptomyces* phage ϕC31 and one in an enterophage ϕp27. Whilst the function of these within the bacteriophages is unknown, it is likely that they have been acquired from host strains.

V. STRUCTURE AND CLASSIFICATION OF GntR REGULATORS

Haydon and Guest (1991) first described the GntR family based on a common sequence at the N terminus of the proteins. They showed that a highly conserved 69-amino acid N-terminal region, containing a predicted HTH motif was conserved (Fig. 1.4). Further analysis of the domain using Pfam has indicated that the HTH domain can be refined to an average of 62.2 amino acids within the GntR domain (Finn *et al.*, 2008). Whilst overall sequence identity in the N-terminal HTH domain is low, the prediction of secondary structure is highly conserved with the three α-helices, characteristic of the HTH domain being apparent (Fig. 1.4). Despite the abundance of GntR sequences in the databases there are few crystal structures available to fully examine structure/function relationships at a detailed level.

Haydon and Guest (1991) also noted that there was extensive variation in the C-terminal domain suggesting heterogeneity in the effector molecules that they bind. Interestingly analysis of all full-length GntR-like sequences in the Pfam database indicates that on average the N-terminal

Author's personal copy

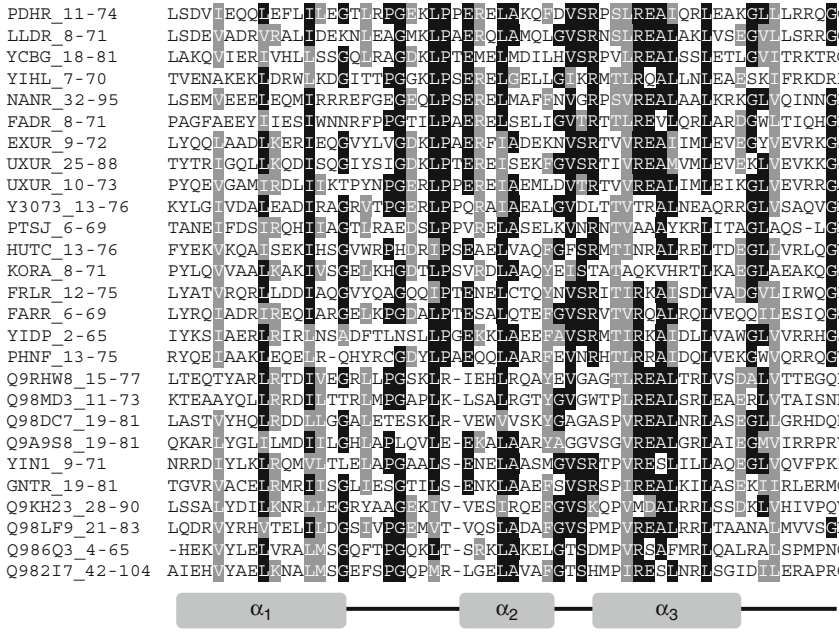


FIGURE 1.4 Alignment of the HTH domain of GntR regulators. This alignment demonstrates conservation of sequence structure within the HTH domain. Alignment was performed using the Seed alignment from Pfam (Finn *et al.*, 2008) used to generate the GntR family hidden Markov Model; the alignment was performed in ClustalW (Larkin *et al.*, 2007) with the residues coloured using Boxshade (http://www.ch.embnet.org/software/BOX_form.html). Secondary structural predictions were performed using the ProteinPredict webserver (Rost *et al.*, 2004) and were checked against the 3D structure of FadR.

domain accounts for 21.8% of the amino acid sequence, yet homology across the whole protein is around 30% reinforcing the concept of extensive C-terminal heterogeneity.

Despite only limited knowledge of two operator sequences at the time, Haydon and Guest (1991) noted that the recognition sequence for the GntR and HutC regulators was identical at seven of the eleven residues. These observations indicated that there are three interacting components of the GntR regulator: the DNA-binding domain, the so-called effector-binding and/or oligomerisation domain (E-b/O) and the *cis*-acting operator sequence. Rigali *et al*. (2002) exploited this idea through extensive analysis of the C-terminal domain of 270 GntR sequences, the N-terminal DNA-binding domain and the operator site. This led to the first extensive work on the family after its initial designation by Haydon and Guest (1991) and the formation of four sub-families within the GntR regulators,

Author's personal copy

and subsequent work led to the designation of a further three: the AraR, DevA and PlmA sub-families (Franco *et al.*, 2006; Hoskisson *et al.*, 2006; Lee *et al.*, 2003).

Due to the abundance of sequences in the databases, the sub-division of HTH DNA-binding proteins has become an issue in bioinformatics with regard to informing genome annotators. There are relatively few studies where extensive analysis of a protein and its operators has been attempted to inform on protein function (Brown *et al.*, 2003; Busenlehner *et al.*, 2003; Korner *et al.*, 2003; Maddocks and Oyston, 2008; Molina-Henares *et al.*, 2006; Weickert and Adhya, 1992).

The classification of the GntR sub-families was based on alignment of the C terminus and secondary structural predictions to reveal distinct structural topologies within each sub-family (Rigali *et al.*, 2002). Extensive analysis was verified by comparison of the predicted topology with that of the known crystal structure of FadR (van Aalten *et al.*, 2000a) which added confidence to the findings. Comparison of the predicted secondary structures is shown in Fig. 1.5 and illustrates the diversity and length differences in each sub-family described so far.

The most abundant GntR sub-family is FadR, named after the fatty acid biosynthesis and degradation regulator of the same name. This sub-family accounts for approximately 40% of GntR regulators, with the C terminus averaging 160 amino acids and consisting of six or seven α-helices. The crystal structure of the C-terminal of FadR of *Escherichia coli* is known (van Aalten *et al.*, 2000a,b, 2001) and served as a validation of the secondary structural predictions of Rigali *et al.* (2002). FadR is an acyl-CoA-responsive member of the GntR family (van Aalten *et al.*, 2001). The regulator exhibits an unusual protein fold overall. The winged helix-turn-helix (wHTH) is fused to a seven α-helix bundle which has crossover topology, containing a large internal cavity required for binding acyl-CoA (Raman and DiRusso, 1995; van Aalten *et al.*, 2000a,b). One unclear aspect of this study was how an effector-binding domain, located 30 Å from the DNA-binding domain, can affect transcription. van Aalten *et al.* (2001) elucidated this through showing, that upon binding acyl-CoA, the protein backbone undergoes dramatic conformational shifts, which results in a 7.2-Å movement of the DNA recognition helix preventing DNA binding and subsequent transcriptional repression.

Recently a cluster within the FadR sub-family was identified that appears to have evolved in Gram-positive organisms for citrate utilisation, with the authors inferring that this lineage arose through E-b/O domain replacement in an ancestral protein (Blancato *et al.*, 2008).

The second sub-family is the HutC grouping, which represents a highly diverse family in terms of effector molecules and processes regulated and accounts for around 30% of GntR regulators. The HutC sub-family C terminus consists of α-helices and β-sheets and averages 170

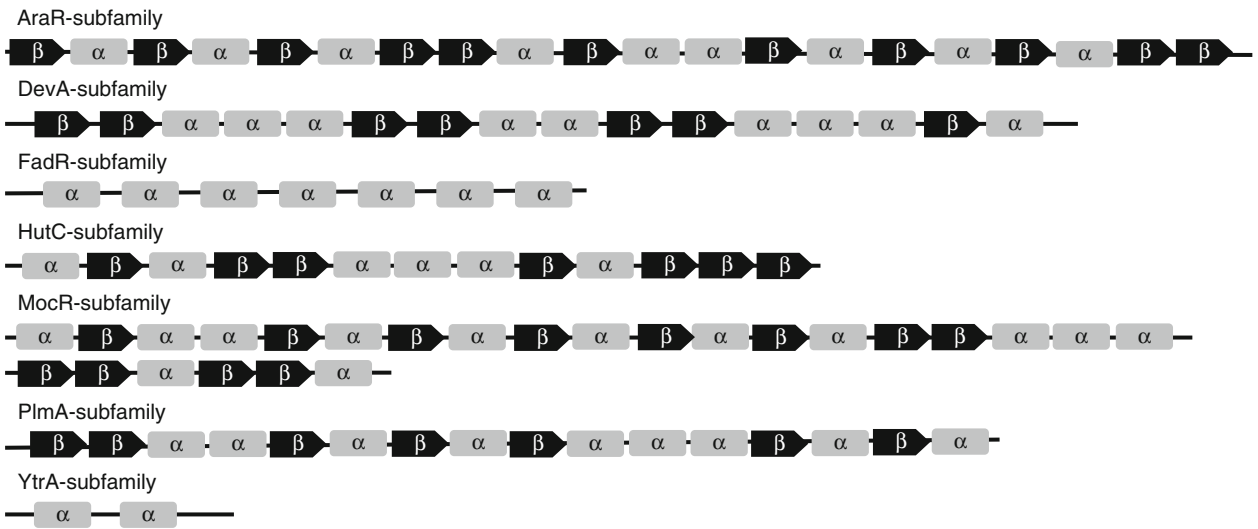


FIGURE 1.5 Structural representation of the C-terminal domains of the GntR sub-families. Secondary structural predictions were performed using the ProteinPredict webserver (Rost *et al.*, 2004) and were checked against the 3D structure of FadR. The α-helices and β-sheets are not to scale and just represent the structural arrangement of these domains.

Author's personal copy

amino acids in length. Named after the Histidine utilisation operon regulator, members of this family have been shown to regulate processes as diverse as amino acid uptake and plasmid transfer (Allison and Phillips, 1990; Kendell and Cohen, 1988). Three-dimensional crystal structural and analysis of the HutC family member, PhnF from *E. coli* (the regulator of transport and biodegradation of phosphonates) (Gorelik *et al.*, 2006), confirmed earlier bioinformatic work of Aravind and Anantharaman (2003) that the C-terminal domain had homology to the chorismate lyase fold, comprising a six-stranded antiparallel β-sheet, a two-stranded parallel β-sheet and four short α-helices. There is a large cavity present on the surface of the PhnF C terminus which may represent a putative effector molecule-binding pocket which shows conservation within other HutC family members (Gorelik *et al.*, 2006).

The MocR sub-family is remarkable for the average length of the C terminus being 350 amino acids and the presence of good homology to the class I aminotransferase enzymes. These enzymes act as dimeric proteins to catalyse the transfer of an amino group to an acceptor molecule such as a keto acid (Ko *et al.*, 1999). Interestingly this raises the possibility of the C terminus of MocR-like regulators being catalytic or that homology is linked to the dimerisation of these proteins and may suggest the recruitment of such a domain to facilitate protein–protein interaction (Rigali *et al.*, 2002).

The YtrA sub-family contains the shortest E-b/O domain, which consists of only two α-helices and often appears to be associated with ATP-binding cassette (ABC) transporters. The homology observed between this group is low (Rigali *et al.*, 2002), which may indicate further sub-division of this family may occur. Interestingly the shortened E-b/O in this sub-family, averaging only 50 amino acids, may not be efficient for binding effector molecules, but may still be effective for dimerisation (Rigali *et al.*, 2002; Yoshida *et al.*, 2000). It is interesting to hypothesise that given the shortened C terminus of the YtrA sub-family, the binding of small molecules may impair dimerisation resulting in loss of DNA binding.

The DevA sub-family of GntR regulators has only been found in streptomycetes (Hoskisson *et al.*, 2006). DevA-like proteins are most similar in the N-terminal domain to the HutC sub-family; however, the C-terminal domain has a novel topology (Fig. 1.5). Phylogenetic analysis and BLAST results confirmed that DevA and its relatives (orthologues in *Streptomyces avermitilis* and *S. scabies*) form a novel GntR sub-family due to the distinct C-terminal topology. The exact role of these proteins has yet to be elucidated; however, DevA is required for correct development in *S. coelicolor* (Hoskisson *et al.*, 2006).

The PlmA sub-family of GntR regulators represent a minor grouping that were first identified in playing a role in plasmid maintenance in the filamentous Cyanobacterium *Anabaena* sp. Strain PCC7120 (Lee *et al.*, 2003).

Author's personal copy

This sub-family has a novel domain topology when compared to other GntR sub-families (Fig. 1.5) and appears to be confined to the Cyanobacteria. Alignments of the N-terminal appear to cluster this sub-family closely with the MocR and YtrA sub-families suggesting that despite their similarity in the DNA-binding domain, the E-b/O domain may have been replaced during evolution.

The AraR sub-family is another minor grouping (Franco *et al.*, 2006; Lee *et al.*, 2003). The GntR regulator AraR from *B. subtilis* is responsible for the control of a range of carbon catabolic genes. The N-terminal domain is classically GntR-like; however, the C-terminal domain has extensive homology to the LacI/GalR regulator family, reinforcing the paradigm of modular repressors in a one-component regulatory system. This protein has reasonably well-characterised binding sites (Mota *et al.*, 1999, 2001) which results in a high-level repression by co-ordinately binding two in-phase operators causing DNA looping. The second results from single operator binding which is autoregulatory and results in low levels of transcription.

The emergence of further GntR sub-families is highly likely and preliminary screening of bacterial genomes for novel C-terminal domains fused to a GntR-like DNA-binding domain has already revealed three new effector-binding domain topologies (Rigali, Unpublished). The structure/function relationship of these new GntR domains is currently under investigation in order to highlight whether these chimeric anomalies have emerged within strains in response to a specific need or a particular environmental context.

VI. DNA BINDING, OPERATOR SEQUENCES AND REGULATION

The HTH DNA-binding domain is the best characterised of all transcription factors, exhibiting significant structural and functional versatility despite its simple structural scaffold (Aravind *et al.*, 2005). Most prokaryotic HTH transcriptional regulators bind DNA as homodimeric proteins either stable in solution or dimerisation occurring on binding (Raman *et al.*, 1997). Studies on FadR (Raman *et al.*, 1997; van Aalten *et al.*, 2000a, 2001) demonstrated the protein binds DNA as a dimer through interaction of specific regions of the C-terminal domain, as does the crystal structure of PhnF (Gorelik *et al.*, 2006). Surface-enhanced laser desorption ionisation (SELDI) mass spectrometry analysis of DevA also revealed a likely dimeric arrangement of this protein (Jakimowicz and Hoskisson, Unpublished). These data suggest that this is a common arrangement of these proteins; however, this must also be considered along with the operator-binding site. Operator-binding sites can be in the form of

Author's personal copy

inverted repeats or as directed repeats (Rigali *et al.*, 2002). The directed repeat arrangement of the operator would impact on the dimerisation mode and also on the arrangement of the operator site, which would lack the classical palindromic sequence. This lack of a symmetrical operator is the likely arrangement for TraR (Rossbach *et al.*, 1994), AphS (Arai *et al.*, 1999), BphS (Watanabe *et al.*, 2000) and FucR (Hooper *et al.*, 1999).

Rigali *et al.* (2004) suggested that sterical constraints on the HTH domain from the binding of effector molecules ultimately impact on the accommodation of the HTH in *cis*-acting elements. Considering this, alignment of the known symmetrical operator sequences by Rigali *et al.* (2002) showed that the family binds a palindromic operator comprising 5′-$(N)_y$GT$(N)_x$AC$(N)_y$-3′ with the number of residues (y) and the nature of the central residues (x) varying. Interestingly the operator sites of the HutC, FadR and YtrA sub-families show degrees of family identity suggesting a common ancestry. The FadR consensus is 5′-t.GTa.tAC.a-3′ and HutC sub-family consensus is 5′-GT.ta.AC-3′, which indicates high levels of conservation in the HTH domain, which is also demonstrated in the phylogenetic tree presented in Rigali *et al.* (2002). The YtrA sub-family shows only 5′-GT.AC3′ identity over a much larger palindromic sequence; this diversity observed in the YtrA operator is also reflected in the looser phylogenetic association of the family members. The remaining sub-families do not have enough or any experimentally confirmed binding sites to draw conclusions on operator sites. However given the negative autoregulatory nature of many of these proteins, insight can be gained from analysis of the upstream region of the GntR regulator. This approach was used by Rigali *et al.* (2004) in *Streptomyces coelicolor*. Using a systematic analysis of the HutC sub-family members in the *S. coelicolor* genome a consensus sequence was identified for this grouping (5′-GT-$N_{(1)}$-TA-$N_{(1)}$-AC-3′) which was then searched against the genome. The presumed target genes are then identified through appropriate location of the *cis*-acting site and a list of candidate genes for testing experimentally is compiled. Compilation of a weight matrix refined the consensus sequence and the most likely GntR regulator could then be identified. Rigali and coworkers found this strategy worked extremely well, enabling insight into the regulon of DasR through experimental confirmation of the *in silico* predictions. This work identified genes of the phosphotransferase system specific for the uptake of *N*-acetylglucosamine, and electrophoretic mobility assays confirmed their promoters were bound by DasR. Indicating that such *in silico* approaches can identify targets for experimental work. Further development of this method led to the inception of the PREDetector tool (Hiard *et al.*, 2007) which allowed the weight matrix to be constructed for all bacterial genomes to help elucidate the regulons of a range of DNA-Binding proteins.

Author's personal copy

The identification and characterisation of GntR regulators by many authors has led to increasing numbers of binding sites being characterised and identified. Interestingly the majority of these studies have demonstrated the negative autoregulatory nature of the GntR regulator. It is not unprecedented that these proteins can also act as activator proteins. In *E. coli* the FadR regulator is known to regulate 12 genes or operons, repressing genes encoding enzymes for the catabolism of fatty acids (including *fadL*, *fadD*, *fadE*, *fadBA* and *fadH*) and activating the genes of the anabolic fatty acid pathway (DiRusso and Nyström, 1998; Henry and Cronan, 1991). Indeed it is fascinating from a physiological point of view that this metabolic-responsive regulator acts in a positive and negative way to maintain poise within lipid metabolism through binding of lipid intermediates. Another member of the FadR sub-family exhibits this dual regulation role; NorG from *Staphylococcus aureus* appears to act as a repressor of an ABC-like transporter involved in cell wall autolysis and a direct activator of drug efflux proteins (Truong-Bolduc and Hooper, 2007). Interestingly transcriptional activation by GntR regulators has been observed in the citrate regulator CitO in *enterococcus* (Blancato *et al.*, 2008), antibiotic biosynthesis in *Serratia* (Fineran *et al.*, 2005), Taurine utilisation via TauR in *Rhodobacter* (Wiethaus *et al.*, 2008). Whether activator function or activator/repressor function is widespread within this family is unknown and at this point it is impossible to elucidate with reliability the differences in positively and negatively regulated operator sites. Yet it does indicate the flexibility of this modular metabolite-responsive regulator framework and sets scene for further investigation.

VII. EVOLUTION OF GntR REGULATORS

Several families of HTH-containing transcription factors are conserved throughout the bacteria and archaea. Despite the high levels of horizontal gene transfer between these groups, the distribution the GntR regulators suggests that a pan-bacterial and pan-archaeal distribution may represent a lineage that can be traced back to the last universal common ancestor, indicating that this domain is extremely ancient (Aravind *et al.*, 2005). The recruitment of sensor domains to the HTH domain has occurred frequently throughout evolution and has resulted in the diversity of functions known for this group of proteins. The sub-family separation observed for the GntR regulators indicates that domains may have been swapped or fused, creating chimeric proteins able to respond to novel metabolites and regulate cellular processes in response to changing conditions. Although there is little direct evidence for evolution of GntR regulators in this manner we can infer several scenarios from genome comparisons and sequence analysis.

Author's personal copy

The presence of multiple E-b/O domains that have fused to a common HTH domain accounts for the observed sub-families in Rigali *et al*. (2002), with the increase in each lineage likely to be the result of horizontal transfer and gene duplication events (Hoskisson, unpublished; Rigali *et al.*, 2002). This scenario would account for the higher levels of sequence similarity amongst sub-family members.

Interestingly the fusion of the HTH domain can be explained with two examples: the presence of aminotransferase-like proteins in the MocR sub-family and the presence of enzyme-like protein folds in HutC sub-family members. Fusion of genes can occur through gene fission, horizontal gene transfer or gene decay (Suhre and Claverie, 2004), so it is clear that adjacent genes can easily become fused with HTH domains. The recruitment of protein–protein interaction domains through this route may also explain the frequency of dimeric interactions in this and other HTH-containing families. Analysis of the HutC GntR-like protein in *Pseudomonas* by Aravind and Anantharaman (2003) using iterative BLAST searching on the C-terminal E-b/O domain recovered the chorismate lyase protein (UbiC) of *E. coli*. Structural analysis *in silico* revealed that this specific protein fold is widespread and is likely to have evolved for ligand binding and evolved into the enzyme chorismate lyase, whilst another version was recruited to the HTH domain and diversified to interact with various ligands from which it can mediate gene expression through ligand binding and the HTH domain.

A second scenario is gene duplication, which is an important evolutionary force that provides an organism an opportunity to evolve new functions. One of the duplicated gene copies diverges, to acquire differential regulation, or mutations occur, followed by evolution into a gene product with a new function. In the case of oligomeric proteins, duplicate copies sometimes evolve to function as hetero-oligomers (Dickson *et al*., 2000). Duplication is also used as a mechanism to acquire a varied substrate spectrum. Thus, functional variations and differential regulation can be obtained as a result of gene duplication and provide an adaptive or fitness advantage in the natural environment. Indeed, data available for *E. coli* and *S. cerevisae* suggest that gene duplication plays a key role in the growth of gene networks (Teichmann and Babu, 2004). Classically, gene duplication is thought to enable duplicates to become specialised in different tissues or developmental stages (Ohno, 1970). Recently, we identified a novel developmental locus, *devA*, in *S. coelicolor* (Hoskisson *et al.*, 2006). It belongs to a novel sub-family of GntR regulators, of which several others have been implicated in development of *S. coelicolor* (see Section X). Interestingly, the *devA* gene is duplicated on the chromosome of all sequenced streptomycetes (with the exception of *S. griseus*; Hoskisson, Unpublished). Each paralog exhibits around 60% identity to each other, and suggests divergence following the duplication event and

Author's personal copy

represents an ideal model for studying evolution of regulatory proteins following duplication and on evolution of GntR regulators.

VIII. GntR REGULATORS IN PRIMARY METABOLISM

Given that GntR regulators respond to many and varied metabolites it is unsurprising that they have been recruited to regulate many primary metabolic processes, responding to changing metabolite concentrations to modulate genes expression. This enables the cell to respond to rapidly changing conditions, or maintain precise balance of specific metabolites. GntR regulators have been shown to play roles in maintenance of fatty acids in response to changing fatty acid concentrations (DiRusso and Nyström, 1998), amino acid catabolism (Allison and Phillips, 1990; Hänssler *et al.*, 2007; Ortuño-Olea and Durán-Vargas, 2000), organic acids (Fujita and Fujita, 1987; Lee *et al.*, 2000; Morawski *et al.*, 2000; Núñez *et al.*, 2001; Pellicer, *et al.*, 1996; Robert-Baudouy *et al.*, 1981; Shulami *et al.*, 1999), regulation of carbon catabolism (Mota *et al.*, 1999; Rigali *et al.*, 2002, 2006; Titgemeyer *et al.*, 1995) and degradation of complex organics (Arai *et al.*, 1999; Watanabe *et al.*, 2000). The range of metabolites bound by GntR regulators is reflected in the diversity of C-terminal domains, and our understanding of how these regulators accommodate and respond to different metabolites is likely to increase in the future.

IX. GntR REGULATORS IN VIRULENCE

Despite the large numbers of GntR regulators associated with metabolism it is increasingly becoming apparent that these proteins can regulate virulence in pathogenic bacteria. It is likely that these proteins regulate virulence through responding to host factors through sensing and responding to specific metabolites. Recently studies have shown that the mammalian cell entry (mce) genes of *Mycobacterium tuberculosis* are regulated by a GntR regulator, *mce1R* (Casali *et al.*, 2006). Mce1R is negatively autoregulated and a mutant is impaired in its ability to survive in murine models. Mce1R is a member of the FadR sub-family and recently this gene cluster has been implicated in cholesterol transport (Mohn *et al.*, 2008) which fits with the types of molecules bound by other members this sub-family.

A screen for mutants that were impaired in virulence in *Brucella melitensis* identified three GntR regulators (Haine *et al.*, 2005). Further work is required to elucidate the regulon of these proteins, but it opens a new avenue of research into how these proteins control virulence.

Author's personal copy

Interestingly these studies again highlight the need to develop robust methods for identifying the metabolites bound by these proteins, to understand how pathogenic organisms integrate host signals to increase survival *in vivo*.

X. GntR REGULATORS IN *STREPTOMYCES* DEVELOPMENT AND ANTIBIOTIC PRODUCTION

Several GntR-like regulators have been implicated in development of *S. coelicolor* such as WhiH (Ryding *et al.*, 1998), *agl3R* and SE69 (Hillerich and Westpheling, 2006; Sprusansky *et al.*, 2003), yet none of their effector molecules are known. However a remarkable example of a GntR superfamily member acting as a master switch of gene expression is provided by DasR (Deficient in aerial mycelium and spore formation) whose predicted regulon includes hundreds of genes in *S. coelicolor* (Rigali *et al.*, 2008). The 'core' of the DasR regulon consists of chitinase genes, and genes involved in the transport and utilisation of aminosugars such as *N*,*N*′-diacetylchitobiose by the DasABC ATP-binding cassette transporter (Colson *et al.*, 2008) and *N*-acetylglucosamine (GlcNAc) by the phosphoenolpyruvate phosphotransferase uptake system (PTS^{GlcNAc}) (Rigali *et al.*, 2006). The importance of chitin's contribution to primary metabolism is obvious from end products of its complete catabolism via enzymes encoded by *nagA* (*N*-acetylglucosamine-6-phosphate deacetylase) and *nagB* (Glucosamine-6-phosphate isomerase), both DasR-dependent genes. These end products are acetate, ammonia, and fructose-6-phosphate, which are further metabolised via the tricarboxylic acid cycle, fatty acid metabolism, nitrogen metabolism and glycolysis. By controlling the access to the metabolites at the crossroad of these major catabolic pathways, DasR imposes itself as an essential checkpoint to sense and enable the cell to respond to the environment. In addition to the obvious exploitation of this rich carbon and nitrogen source in primary metabolism, GlcNAc has been demonstrated to be a strong nutritional signal for both morphological (sporulation) and physiological (secondary metabolites production) differentiations in *S. coelicolor*. Indeed, under rich culture conditions, the addition of GlcNAc completely blocks development (Rigali *et al.*, 2006), while on minimal media the addition of the same molecule triggers *S. coelicolor* antibiotic production (Rigali *et al.*, 2008). The inactivation of *dasR*, *dasA* and central PTS^{GlcNAc} genes results in developmental arrest of *S. coelicolor*, clearly indicating that sensing and uptake of chitin and/or cell wall and their degradation products are key elements in the decision of streptomycetes to produce reproductive aerial mycelium (Colson *et al.*, 2008; Rigali *et al.*, 2006).

Author's personal copy

Recently, in *S. coelicolor* we identified another GntR-like regulator of development, *devA*, which forms a new sub-family of these regulators (Hoskisson *et al.*, 2006). Interestingly, the *devA* gene is duplicated (*devE*) on the chromosome of all sequenced streptomycetes including the adjacent divergently transcribed genes *devC* and *devD* genes. *devA* is co-transcribed with a putative phosphatase/hydrolase, *devB*, which is of unknown function, but it would appear that tight control of *devB* during growth is required for normal development, as a mutant of *devB* is conditionally bald. *devB* is also unusual in that it is one of only a handful of developmental genes known in *Streptomyces* that does not encode a regulatory protein.

The recruitment of metabolite-responsive regulators to sporulation and antibiotic production pathways demonstrates that tight control of these developmental gene cascades is required to respond to changing environmental conditions to ensure survival in stressful or unfavourable conditions, linking primary metabolites to secondary metabolism and development.

XI. BIOTECHNOLOGY IMPLICATIONS

The tight regulation and highly responsive nature of the GntR superfamily of regulators offers an interesting opportunity in the development of inducible expression systems. The modular nature of these regulators enables the creation of chimeric proteins that respond rapidly and predictably to changing conditions, whether these are intrinsic changes in the environment or through the addition of inducer molecules. Whilst a great understanding of the molecular mechanisms of regulation is required, the three interacting components of the GntR regulator system can be exploited to facilitate the construction of vector systems containing well-characterised operator, DNA-binding domain and a metabolite-responsive effector-binding domain, whether natural or synthesised by molecular biology.

The elucidation of the GlcNAc/DasR regulon has highlighted the first complete signalling cascade for antibiotic production from nutrient sensors to antibiotic pathway-specific activators leading to a new strategy for activating pathways for secondary metabolite biosynthesis in streptomycetes, thereby offering new prospects in the fight against multi-drug resistant pathogens and cancers. Activating these pathways has proved challenging and several approaches to activate these pathways have proved unsuccessful. Rigali *et al.* (2008) demonstrated that *N*-acetylglucosamine (GlcNAc) acts as a checkpoint in the onset of secondary metabolism in several streptomycetes and a simple screen, using the addition of GlcNAc to cultures, is sufficient to induce antibiotic pathways through

Author's personal copy

derepression of DasR-regulated promoters and offers the opportunity to develop novel bioactive metabolites for the clinic in a wide range of producing strains.

XII. CONCLUDING REMARKS

The one-component regulatory proteins of the GntR family represent a highly diverse group of proteins from a structural, functional and biochemical aspect. They are widespread throughout the bacterial kingdom and regulate a wide range of processes from primary metabolism to developmental processes. The fusion of a DNA-binding domain to a diverse range of effector-binding/oligomerisation domains enables flexibility and has probably contributed to the expansion of the gene family through recruitment of new effector-binding/oligomerisation domains. This flexibility makes these proteins prime candidates for use in biotechnology creating highly effective inducible systems, responding to well-characterised metabolites in a rapid and predictable way.

ACKNOWLEDGMENTS

PAH would like to thank The Society for General Microbiology (SGM), the Scottish Universities Life Science Alliance (SULSA) and the University of Strathclyde for funding the work in his laboratory. We would also like to thank Dr Lorena Fernandez for helpful comments on the manuscript.

REFERENCES

Allison, S. L., and Phillips, A. T. (1990). Nucleotide sequence of the gene encoding the repressor for the histidine utilization genes of *Pseudomonas putida*. *J. Bacteriol.* **172,** 5470–5476.

Arai, H., Akahira, S., Ohishi, T., and Kudo, T. (1999). Adaptation of *Comamonas testosteroni* TA441 to utilization of phenol by spontaneous mutation of the gene for a trans-acting factor. *Mol. Microbiol.* **33,** 1132–1140.

Aravind, L., and Anantharaman, V. (2003). HutC/FarR-like bacterial transcription factors of the GntR family contain a small molecule-binding domain of the chorismate lyase fold. *FEMS Microbiol. Lett.* **222,** 17–23.

Aravind, L., Anantharaman, V., Balaji, S., Babu, M. M., and Iyer, L. M. (2005). The many faces of the helix-turn-helix domain: Transcription regulation and beyond. *FEMS Microbiol. Rev.* **29,** 231–262.

Aravind, L., and Koonin, E. V. (1999). DNA-binding proteins and evolution of transcription regulation in the archaea. *Nucleic Acids Res.* **27,** 4658–4670.

Blancato, V. S., Repizo, G. D., Suarez, C. A., and Magni, C. (2008). Transcriptional regulation of the citrate gene cluster of Enterococcus faecalis involves the GntR family transcriptional activator CitO. *J. Bacteriol.* **190,** 7419–7430.

Author's personal copy

Brown, N. L., Stoyanov, J. V., Kidd, S. P., and Hobman, J. L. (2003). The MerR family of transcriptional regulators. *FEMS Microbiol. Rev.* **27,** 145–163.

Busenlehner, L. S., Pennella, M. A., and Giedroc, D. P. (2003). The SmtB/ArsR family of metalloregulatory transcriptional repressors: Structural insights into prokaryotic metal resistance. *FEMS Microbiol. Rev.* **27,** 131–143.

Casali, N., White, A. M., and Riley, L. W. (2006). Regulation of the *Mycobacterium tuberculosis* mce1 operon. *J. Bacteriol.* **188,** 441–449.

Cole, S. T., Eiglmeier, K., Parkhill, J., James, K. D., Thomson, N. R., Wheeler, P. R., Honoré, N., Garnier, T., Churcher, C., Harris, D., Mungall, K., Basham, D., *et al.* (2001). Massive gene decay in the leprosy bacillus. *Nature* **409,** 1007–1011.

Colson, S., van Wezel, G. P., Craig, M., Noens, E. E., Nothaft, H., Mommaas, A. M., Titgemeyer, F., Joris, B., and Rigali, S. (2008). The chitobiose-binding protein, DasA, acts as a link between chitin utilization and morphogenesis in *Streptomyces coelicolor*. *Microbiology* **154,** 373–382.

Dickson, R., Weiss, C., Howard, R. J., Alldrick, S. P., Ellis, R. J., Lorimer, G., Azem, A., and Viitanen, P. V. (2000). Reconstitution of higher plant chloroplast chaperonin 60 tetradecamers active in protein folding. *J. Biol. Chem.* **275,** 11829–11835.

DiRusso, C. C., Metzger, A. K., and Heimert, T. L. (1993). Regulation of transcription of genes required for fatty acid transport and unsaturated fatty acid biosynthesis in *Escherichia coli* by FadR. *Mol. Microbiol.* **7,** 311–322.

DiRusso, C. C., and Nyström, T. (1998). The fats of *Escherichia coli* during infancy and old age: Regulation by global regulators, alarmones and lipid intermediates. *Mol. Microbiol.* **27,** 1–8.

Dodd, I. B., and Egan, J. B. (1990). Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. *Nucleic Acids Res.* **18,** 5019–5026.

Fineran, P. C., Everson, L., Slater, H., and Salmond, G. P. C. (2005). A GntR family transcriptional regulator (PigT) controls gluconate-mediated repression and defines a new, independent pathway for regulation of the tripyrrole antibiotic, prodigiosin, in *Serratia*. *Microbiology* **151,** 3833–3845.

Finn, R. D., Tate, J., Mistry, J., Coggill, P. C., Sammut, J. S., Hotz, H. R., Ceric, G., Forslund, K., Eddy, S. R., Sonnhammer, E. L., and Bateman, A. (2008). The Pfam protein families database. *Nucleic Acids Res. Database Issue* **36,** D281–D288.

Franco, I. S., Mota, L. J., Soares, C. M., and de Sá-Nogueira, I. (2006). Functional domains of the *Bacillus subtilis* transcription factor AraR and identification of amino acids important for nucleoprotein complex assembly and effector binding. *J. Bacteriol.* **188,** 3024–3036.

Fujita, Y., and Fujita, T. (1987). The gluconate operon gnt of Bacillus subtilis encodes its own transcriptional negative regulator. *Proc. Natl. Acad. Sci. USA* **84,** 4524–4528.

Gorelik, M., Lunin, V. V., Skarina, T., and Savchenko, A. (2006). Structural characterization of GntR/HutC family signaling domain. *Protein Sci.* **15,** 1506–1511.

Haine, V., Sinon, A., Van Steen, F., Rousseau, S., Dozot, M., Lestrate, P., Lambert, C., Letesson, J. J., and De Bolle, X. (2005). Systematic targeted mutagenesis of *Brucella melitensis* 16M reveals a major role for GntR regulators in the control of virulence. *Infect Immun.* **73,** 5578–5586.

Hänssler, E., Müller, T., Jessberger, N., Völzke, A., Plassmeier, J., Kalinowski, J., Krämer, R., and Burkovski, A. (2007). FarR, a putative regulator of amino acid metabolism in *Corynebacterium glutamicum*. *Appl. Microbiol. Biotechnol.* **76,** 625–632.

Haydon, D. J., and Guest, J. R. (1991). A new family of bacterial regulatory proteins. *FEMS Microbiol. Lett.* **63,** 291–295.

Henry, M. F., and Cronan, J. E. (1991). *Escherichia coli* transcription factor that both activates fatty acid synthesis and represses fatty acid degradation. *J. Mol. Biol.* **222,** 843–849.

Hiard, S., Maree, R., Colson, S., Hoskisson, P. A., Titgemeyer, F., van Wezel, G. P., Joris, B., Wehenkel, L., and Rigali, S. (2007). PREDetector: A new tool to identify regulatory elements inbacterial genomes. *Biochem. Biophys. Res. Commun.* **357,** 861–864.

Author's personal copy

Hillerich, B., and Westpheling, J. (2006). A new GntR family transcriptional regulator in *streptomyces coelicolor* is required for morphogenesis and antibiotic production and controls transcription of an ABC transporter in response to carbon source. *J. Bacteriol.* **188,** 7477–7487.

Hooper, L. V., Xu, J., Falk, P. G., Midtvedt, T., and Gordon, J. I. (1999). A molecular sensor that allows a gut commensal to control its nutrient foundation in a competitive ecosystem. *Proc. Natl. Acad. Sci. USA* **96,** 9833–9838.

Hoskisson, P. A., Rigali, S., Fowler, K., Findley, K., and Buttner, M. J. (2006). DevA, a GntR-like Transcriptional Regulator Required for Development in *Streptomyces coelicolor*. *J. Bacteriol.* **188,** 5014–5023.

Jaques, S., and McCarter, L. L. (2006). Three new regulators of swarming in *Vibrio parahaemolyticus*. *J. Bacteriol.* **188,** 2625–2635.

Kendell, K., and Cohen, S. (1988). Complete nucleotide sequence of the *Streptomyces lividans* plasmid pIJ101 and correlation of the sequence with genetic properties. *J. Bacteriol.* **170,** 4634–4651.

Ko, T.-P., Wu, S.-P., Yang, W. Z., Tsai, H., and Yuan, H. S. (1999). Crystallization and preliminary crystallographic analysis of the *Escherichia coli* tyrosine aminotransferase. *Acta Cryst.* **D55,** 1474–1477.

Korner, H., Sofia, H. J., and Zumft, W. G. (2003). Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: Exploiting the metabolic spectrum by controlling alternative gene programs. *FEMS Microbiol. Rev.* **27,** 559–592.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., *et al.* (2007). ClustalW and ClustalX version 2. *Bioinformatics* **23,** 2947–2948.

Lee, H. Y., An, J. H., and Kim, Y. S. (2000). Identification and characterization of a novel transcriptional regulator, MatR, for malonate metabolism in *Rhizobium leguminosarum* bv. *trifolii*. *Eur. J. Biochem.* **267,** 7224–7230.

Lee, M. H., Scherer, M., Rigali, S., and Golden, J. W. (2003). PlmA, a new member of the GntR family, has plasmid maintenance functions in *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **185,** 4315–4325.

Maddocks, S. E., and Oyston, P. C. F. (2008). Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology* **154,** 3609–3623.

Mohn, W. W., van der Geize, R., Stewart, G. R., Okamoto, S., Liu, J., Dijkhuizen, L., and Eltis, L. D. (2008). The actinobacterial mce4 locus encodes a steroid transporter. *J. Biol. Chem.* **283,** 35368–35374.

Molina-Henares, A. J., Krell, T., Guazzaroni, M. E., Segura, A., and Ramos, J. L. (2006). Members of the IclR family of bacterial transcriptional regulators function as activators and/or repressors. *FEMS Microbiol. Rev.* **30,** 157–186.

Morawski, B., Segura, A., and Ornston, L. N. (2000). Repression of *Acinetobacter vanillate* demethylase synthesis by VanR, a member of the GntR family of transcriptional regulators. *FEMS Microbiol. Lett.* **187,** 65–68.

Mota, L. J., Sarmento, L. M., and de Sá-Nogueira, I. (2001). Control of the arabinose regulon in *Bacillus subtilis* by AraR *in vivo*: Crucial roles of operators, cooperativity, and DNA looping. *J. Bacteriol.* **183,** 4190–4201.

Mota, L. J., Tavares, P., and Sá-Nogueira, I. (1999). Mode of action of AraR, the key regulator of L-arabinose metabolism in *Bacillus subtilis*. *Mol. Microbiol.* **33,** 476–489.

Núñez, M. F., Pellicer, M. T., Badía, J., Aguilar, J., and Baldomà, L. (2001). The gene yghK linked to the glc operon of *Escherichia coli* encodes a permease for glycolate that is structurally and functionally similar to L-lactate permease. *Microbiology* **147,** 1069–1077.

Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., Takeda, Y., and Matthews, B. W. (1982). The molecular basis of DNA-protein recognition inferred from the structure of cro repressor. *Nature* **298,** 718–723.

Author's personal copy

Ohlendorf, D. H., Anderson, W. F., and Matthews, B. W. (1983). Many gene-regulatory proteins appear to have a similar alpha-helical fold that binds DNA and evolved from a common precursor. *J. Mol. Evol.* **19,** 109–114.

Ohno, S. (1970). Evolution by gene duplication. Springer-Verlag, Berlin.

Ortuño-Olea, L., and Durán-Vargas, S. (2000). The L-asparagine operon of *Rhizobium etli* contains a gene encoding an atypical asparaginase. *FEMS Microbiol. Lett.* **189,** 177–182.

Pellicer, M. T., Badía, J., Aguilar, J., and Baldomà, L. (1996). glc locus of *Escherichia coli:* Characterization of genes encoding the subunits of glycolate oxidase and the glc regulator protein. *J. Bacteriol.* **178,** 2051–2059.

Raman, N., Black, P. N., and DiRusso, C. C. (1997). Characterization of the fatty acid-responsive transcription factor FadR. Biochemical and genetic analyses of the native conformation and functional domains. *J. Biol. Chem.* **272,** 30645–30650.

Raman, N., and DiRusso, C. C. (1995). Analysis of acyl coenzyme A binding to the transcription factor FadR and identification of amino acid residues in the carboxyl terminus required for ligand binding. *J. Biol. Chem.* **270,** 1092–1097.

Rigali, S., Derouaux, A., Giannotta, F., and Dusart, J. (2002). Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies. *J. Biol. Chem.* **277,** 12507–12515.

Rigali, S., Nothaft, H., Noens, E. E., Schlicht, M., Colson, S., Müller, M., Joris, B., Koerten, H. K., Hopwood, D. A., Titgemeyer, F., and van Wezel, G. P. (2006). The sugar phosphotransferase system of *Streptomyces coelicolor* is regulated by the GntR-family regulator DasR and links N-acetylglucosamine metabolism to the control of development. *Mol. Microbiol.* **61,** 1237–1251.

Rigali, S., Schlicht, M., Hoskisson, P. A., Northaft, H., Merzbacher, M., Joris, B., and Titgemeyer, F. (2004). Extending the classification of bacterial transcription factors beyond the helix-turn-helix motif as an alternative approach to discover new cis/trans relationships. *Nucleic Acids Res.* **32,** 3418–3426.

Rigali, S., Titgemeyer, F., Barends, S., Mulder, S., Thomae, A. W., Hopwood, D. A., and van Wezel, G. P. (2008). Feast or famine: The global regulator DasR links nutrient stress to antibiotic production by *Streptomyces. EMBO Rep.* **9,** 670–675.

Robert-Baudouy, J., Portalier, R., and Stoeber, F. (1981). Regulation of hexuronate system genes in *Escherichia coli* K-12: Multiple regulation of the uxu operon by exuR and uxuR gene products. *J. Bacteriol.* **145,** 211–220.

Rossbach, S., Kulpa, D. A., Rossbach, U., and de Bruijn, F. J. (1994). Molecular and genetic characterization of the rhizopine catabolism (mocABRC) genes of Rhizobium meliloti L5-30. *Mol. Gen. Genet.* **245,** 11–24.

Rost, B., Yachdav, G., and Liu, J. (2004). The PredictProtein server. *Nucleic Acids Res.* **32**(Web Server issue), W321–W326.

Reuther, J., Wohlleben, W., and Muth, G. (2006). Modular architecture of the conjugative plasmid pSVH1 from *Streptomyces venezuelae. Plasmid* **55,** 201–209.

Ryding, N. J., Kelemen, G. H., Whatling, C. A., Flardh, K., Buttner, M. J., and Chater, K. F. (1998). A developmentally regulated gene encoding a repressor-like protein is essential for sporulation in *Streptomyces coelicolor* A3(2). *Mol. Microbiol.* **29,** 343–357.

Sauer, R. T., Yocum, R. R., Doolittle, R. F., Lewis, M., and Pabo, C. O. (1982). Homology among DNA-binding proteins suggests use of a conserved super-secondary structure. *Nature* **298,** 447–451.

Shulami, S., Gat, O., Sonenshein, A. L., and Shoham, Y. (1999). The glucuronic acid utilization gene cluster from *Bacillus stearothermophilus* T-6. *J. Bacteriol.* **181,** 3695–3704.

Sprusansky, O., Zhou, L., Jordan, S., White, J., and Westpheling, J. (2003). Identification of three new genes involved in morphogenesis and antibiotic production in *Streptomyces coelicolor. J. Bacteriol.* **185,** 6147–6157.

Author's personal copy

Suhre, K., and Claverie, J.-M. (2004). FusionDB: A database for in-depth analysis of prokaryotic gene fusion events. *Nucl. Acids Res.* **32,** D273–D276.

Teichmann, S. A., and Babu, M. M. (2004). Gene regulatory network growth by duplication. *Nat. Genet.* **36,** 492–496.

Titgemeyer, F., Reizer, J., Reizer, A., Tang, J., Parr, T. R. Jr, and Saier, M. H. Jr. (1995). Nucleotide sequence of the region between crr and cysM in *Salmonella typhimurium*: Five novel ORFs including one encoding a putative transcriptional regulator of the phosphotransferase system. *DNA Seq.* **5,** 145–152.

Truong-Bolduc, Q. C., and Hooper, D. C. (2007). The Transcriptional Regulators NorG and MgrA Modulate Resistance to both Quinolones and {beta}-Lactams in *Staphylococcus aureus*. *J. Bacteriol.* **189,** 2996–3005.

van Aalten, D. M., DiRusso, C. C., and Knudsen, J. (2001). The structural basis of acyl coenzyme A-dependent regulation of the transcription factor FadR. *EMBO J.* **20,** 2041–2050.

van Aalten, D. M., DiRusso, C. C., Knudsen, J., and Wierenga, R. K. (2000a). Crystal structure of FadR, a fatty acid-responsive transcription factor with a novel acyl coenzyme A-binding fold. *EMBO J.* **19,** 5167–5177.

van Aalten, D. M., Knudsen, J., DiRusso, C. C., Kokko, T., and Wierenga, R. K. (2000b). Crystallization and X-ray diffraction studies of the fatty-acid responsive transcription factor FadR from Escherichia coli. *Acta Crystallogr. D. Biol. Crystallogr.* **56,** 469–471.

Vindal, V., Suma, K., and Ranjan, A. (2007). GntR family of regulators in *Mycobacterium smegmatis*: A sequence and structure based characterization. *BMC Genomics* **23,** 289.

Watanabe, T., Inoue, R., Kimura, N., and Furukawa, K. (2000). Versatile transcription of biphenyl catabolic bph operon in *Pseudomonas pseudoalcaligenes* KF707. *J. Biol. Chem.* **6,** 31016–31023.

Weickert, M. J., and Adhya, S. (1992). A family of bacterial regulators homologous to Gal and Lac repressors. *J. Biol. Chem.* **267,** 15869–15874.

Wiethaus, J., Schubert, B., Pfänder, Y., Narberhaus, F., and Masepohl, B. (2008). The GntR-like regulator TauR activates expression of taurine utilization genes in *Rhodobacter capsulatus*. *J. Bacteriol.* **190,** 487–493.

Yoshida, K. I., Fujita, Y., and Ehrlich, S. D. (2000). An operon for a putative ATP-binding cassette transport system involved in acetoin utilization of *Bacillus subtilis*. *J. Bacteriol.* **182,** 5454–5461.