Architecture of *Escherichia coli* promoters that respond to reactive nitrogen species

By

David L. Chismon

A thesis submitted to the School of Biosciences of the University of Birmingham

For the degree of
DOCTOR OF PHILOSOPHY

The University of Birmingham
Edgbaston
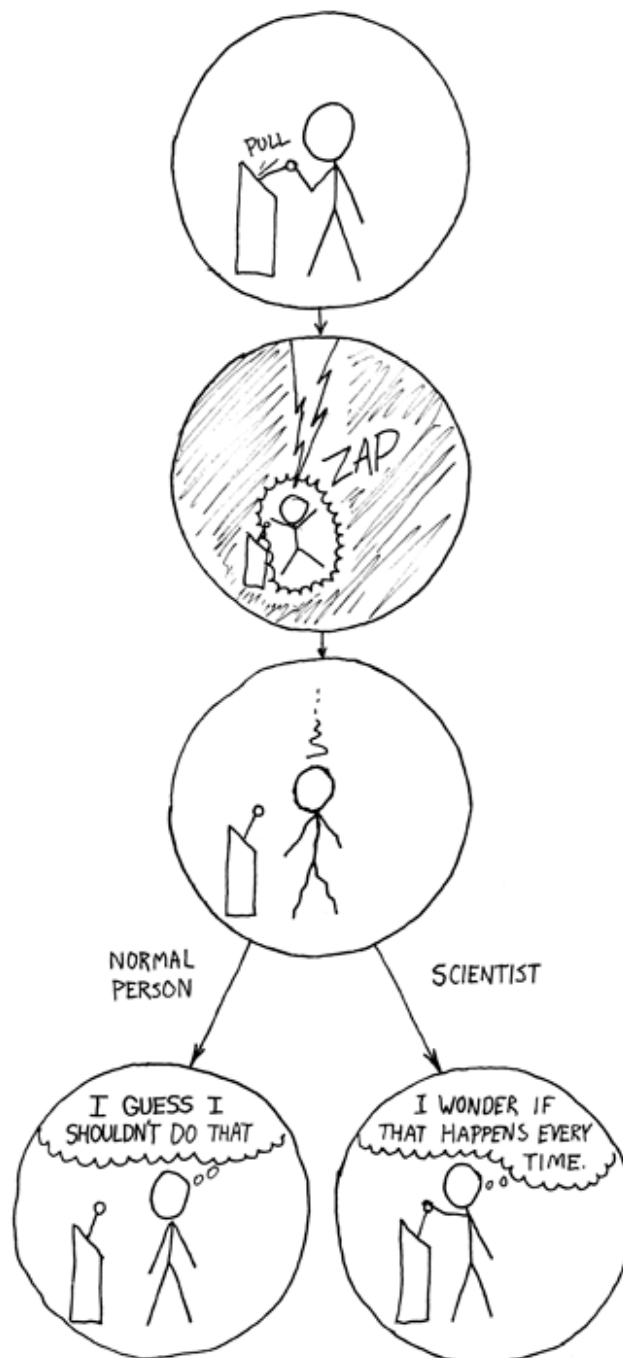Birmingham
UK
B15 2TT

Synopsis

This study examined the regulation of two genes, *hcp* and *ogt*, that are reported to be involved in protection from the effects of reactive nitrogen species in *Escherichia coli*. Biochemical techniques were used to investigate promoter activity and the effects of the transcription factors that are reported to regulate expression of the *hcp* and *ogt* genes, namely NarL / NarP, FNR and NsrR. Transcription activation by NarL was then studied using semi-synthetic promoters.

The *hcp* gene was found to be positively regulated by FNR and negatively regulated by NsrR. Contrary to previous reports, NarL and NarP were found to have little direct effect on expression of *hcp*, however, an indirect effect of NarL was detected. This study demonstrates that the indirect regulation of *hcp* expression by NarL is related to repression by NsrR and suggests that NarL is involved in the generation of reactive nitrogen species.

The *ogt* gene was confirmed to be activated by NarL independently of FNR. Studies focussed on characterising the different classes by which NarL / NarP can activate promoter activity independently of FNR. NarL was found to activate by class I, II and III mechanisms, whilst NarP was capable of class II activation only.

Activation by NarL was studied and a library of alanine substitutions in the carboxyl terminal domain (CTD) of the α subunit of RNA polymerase was used to demonstrate direct interaction between NarL and RNA polymerase. NarL, the CTD of NarL, NarXL, NarP and the CTD of NarP were expressed from plasmids and transcription activation studied in cells lacking chromosomal *narL* and *narP*. Full-length NarL activated by Class I, II and III mechanisms, whilst the CTD of NarL activated by class II mechanisms only. Both the full-length NarP and the CTD of NarP activated by class II mechanisms only.

Dedicated to my family and to Esther

Acknowledgements

Table of contents

Table of figures

Table of tables

Chapter 1: Introduction

1.1. *Escherichia coli* and its environment

Escherichia coli is a gram-negative bacterial species that is able to colonise the human digestive tract and is also able to survive outside of the body in soil and water (Campbell *et al.*, 2001). Although many strains of *E. coli* are harmless, even beneficial to man, pathogenic strains have been identified that can cause fatal disease (Freter and Abrams, 1972). *E. coli* is considered by many to be a model organism for study, due to its amenability to growth in a laboratory and the existence of closely related pathogenic and non-pathogenic strains (Cronan, 2001, Welch *et al.*, 2002). As such, much bacterial research focuses on *E.* coli making it one of the best-understood organisms, although much remains unsolved (Blattner *et al.*, 1997).

E. coli is a facultative anaerobe and is therefore able to survive in the presence or absence of oxygen. *E. coli* can grow on a variety of carbon sources and adapts to various environmental conditions. It is thought that the ability of *E. coli* to survive in such varying conditions is a product of its lifestyle as a commensal (Cases *et al.*, 2003). During its passage through the digestive system, it will encounter variation in oxygen availability, nutrient availability, acidity and risks interaction with the host immune system (Jones *et al.*, 2007). Once outside the body, it will encounter oxygen, UV, varying availability of nutrients and water, all of which it has to be able to survive long enough to colonise a new host.

During respiration, energy is released during the transfer of electrons from an electron donor such as NADH (reduced nicotinamide adenine dinucleotide) through a chain of electron carriers to a terminal electron acceptor (Richardson, 2000). A key factor that determines the energy available to an organism as a result of this reaction is the difference between the midpoint redox potentials of the electron donor and the terminal electron acceptor. The midpoint potential of the NADH/NAD+ pair is -0.32 V, therefore, the more positive the midpoint potential of the terminal electron acceptor and its reduced product, the more energy

will available to the organism. Oxygen is the most powerful oxidant used by bacteria and the midpoint potential of the oxygen/H_2O pair is +0.82 V. *E. coli* is able to utilise oxygen as a terminal electron acceptor and therefore grows rapidly in aerobic conditions. In the absence of molecular oxygen, *E. coli* is able to utilise alternative electron acceptors such as nitrate (NO_3^-), which is the most powerful oxidant that *E. coli* can utilise in the absence of oxygen. (Bonnefoy and Demoss, 1994). Figure 1.1 illustrates the reduction of nitrate to ammonia. When nitrate is reduced by nitrate reductases, it yields nitrite (NO_2^-) in a reaction with a midpoint potential of +0.43 V. Nitrite is then reduced further to ammonia (NH_4^+) in a reaction with a midpoint potential of +0.34 V. Although energy producing, the reduction of nitrate and nitrite also yields toxic by-products, the reactive nitrogen species (RNS) nitric oxide (NO) and dinitrogen trioxide (N_2O_3) (Weiss, 2006). Successful growth in the absence of oxygen therefore requires that *E. coli* be able to utilise the alternative electron acceptors nitrate and nitrite, and be able to mitigate the threat posed by the toxic by-products.

1.2. Reactive nitrogen species

Reactive nitrogen species, such as NO, are powerful mutagens that are able to induce bacteriostasis, cell death or produce mutations in the DNA (Fang, 2004). RNS can cause damage by different pathways. RNS are able to interact with proteins and cause nitrosation of peptides (Sedgwick, 1997). RNS are also reported to methylate DNA indirectly. It is thought that RNS nitrosylate an as-yet unidentified molecule, which in turn methylates DNA upon contact (Taverna and Sedgwick, 1996). Methylation of DNA poses a serious risk to the cell as it can block replication or introduce lethal inter-strand crosslinks (Noll *et al.*, 2006). NO has also been shown to interact with DNA directly, deaminating bases so that they pair incorrectly (Spek *et al.*, 2002). This can introduce mutations into the genome following replication. RNS can also react with the iron-sulphur clusters at the active sites of many enzymes, producing dinitrosyl-iron-complexes (DNIC) (Toledo *et al.*, 2008). The inactivation of proteins containing iron-sulphur clusters is thought to be one of the major disruptive effects of NO.
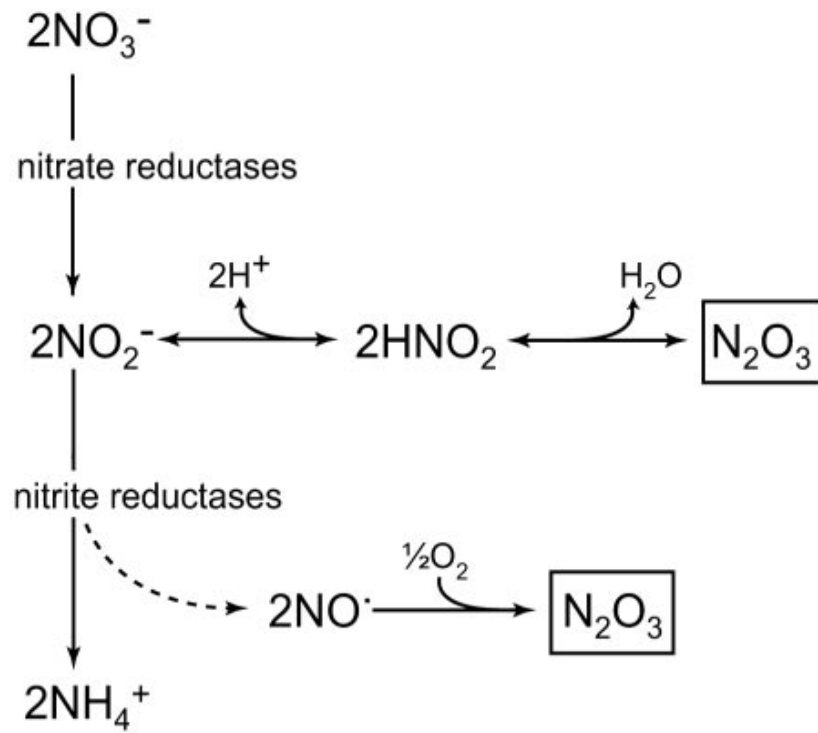
Figure 1.1 Generation of reactive nitrogen species during nitrate reduction

From Weiss *et al.* (2006). Production of the mutagenic nitrosating agents NO and N_2O_3 during nitrate/nitrite metabolism in *E. coli*. NO is a minor by-product of nitrite reductase activity.

Iron-sulphur containing proteins are involved in many processes, from the catalysis of reactions to the regulation of transcription. One important protein that is damaged by RNS is IlvD, the producer of branched chain amino acids (Ren *et al.*, 2008). If the iron-sulphur cluster of IlvD is nitrosylated, bacteria are unable to grow unless provided with branched chain amino acids.

There are multiple theories as to the primary producer of endogenous RNS during nitrate / nitrite metabolism in *E. coli*. Corker and Poole (2003) reported that the periplasmic nitrite reductase Nrf is responsible for NO generation. Conversely, Weiss *et al.* (2006) reported that the nitrite reductase Nir produces NO during nitrite metabolism, whilst the nitrate reductase NarG has also been implicated (Metheringham and Cole, 1997). As well as endogenous sources of RNS, bacteria may also encounter external sources of RNS, such as that released by neutrophils and activated macrophages as a part of the immune response (Fang, 2004).

NO can be detoxified both aerobically and anaerobically. The flavohaemoglobin Hmp oxidises NO to NO_3^- in aerobic conditions and reduces NO to N_2O in anaerobic conditions (Laver *et al.*, 2010). The periplasmic nitrite reductase Nrf is also able to reduce NO anaerobically, as is the cytoplasmic flavorubredoxin NorV (Poock *et al.*, 2002, Gardner *et al.*, 2002). *E. coli* is thought to contain as-yet unidentified enzymes that can detoxify NO, as cells lacking Hmp, Nrf and NorV remain able to reduce NO (Claire Vine, University of Birmingham, personal communication).

E. coli is also able to repair the damage done by RNS to DNA. *E. coli* contains the DNA repair enzymes Ada and Ogt (see below) that directly repair methylated DNA (Potter *et al.*, 1987). Damage such as crosslinks and mis-incorporated nucleotides are repaired by the base excision repair system (reviewed by Grossman and Kovalsky, 2001). The protein YtfE has recently been identified as responsible for repairing damaged iron-sulphur clusters (Justino *et al.*, 2007).

Justino *et al.* (2007) demonstrated that *ytfE* mutant cells had increased susceptibility to reactive nitrogen and oxygen species compared to wild type cells. They also found that cells lacking YtfE were found to be unable to repair damaged iron-sulphur clusters and that when purified YtfE was incubated *in vitro* with damaged iron-sulphur cluster containing proteins, it was able to repair those clusters. The exact mechanism by which YtfE repairs damaged clusters is not yet known although it is thought to involve a di-iron centre (Todorovic *et al.*, 2008). YtfE is therefore the first di-iron protein proven to be involved in the repair of iron-sulphur clusters.

1.3. HCP – the Hybrid Cluster Protein

The Hybrid Cluster Protein (HCP) was initially isolated from *Desulfovibrio vulgaris* as a contaminant (Pierik *et al.*, 1992). HCP has since been found to be produced by a diverse range of bacterial species that exist in a variety of environments (Rodionov *et al.*, 2005). HCP is of particular interest to researchers as early work indicated that HCP contains 2 iron-sulphur clusters, one of which is thought to be novel (van den Berg *et al.*, 2000).

The amino acid sequence of HCP shows no homology to any protein except carbon monoxide dehydrogenases. The crystal structure reveals that the 2 iron-sulphur clusters are located in proximity to each other and hydrophobic solvent channels provide access to them (Cooper *et al.*, 2000). The presence of such channels indicates that the substrate for HCP may be a soluble gas (Montet *et al.*, 1997). Sequence studies indicate that there are 3 classes of HCP (van den Berg *et al.*, 2000): Class 1 HCPs that are found in strictly anaerobic bacteria; Class 2 HCPs that have a different spacing of the residues that hold the iron-sulphur cluster and are found in facultative anaerobes and class 3 HCPs that resemble class 1 HCPs but contain a deletion of residues and are found in hyper-thermophiles and archaea. The differences between the classes could indicate that HCP has subtly different functions in the 3

groups, or more likely indicates adaptations for the environment of the bacteria, such as stability in the presence of oxygen.

Despite intensive study and the availability of a high-resolution structure, the biological function of HCP has yet to be identified. It been established that HCP in *E. coli* is optimally expressed in anaerobic conditions and with nitrate or nitrite present in the growth medium (Filenko *et al.*, 2005). This led to the suggestion that HCP may be involved in nitrate / nitrite respiration, possibly by detoxifying dangerous by-products of such respiration. It was noted that carbon monoxide dehydrogenases, which are similar in structure to HCP, can be modified to become hydroxylamine reductases (Wolfe *et al.*, 2002). Hydroxylamine, an RNS, was reported to be reduced by HCP, although this theory was later discounted, as the optimum pH for the reaction was higher than would be encountered in the cell (pH 9). HCP was also identified by a study of genes induced by RNS in *Salmonella enterica* (Kim *et al.*, 2003). The study indicated that HCP expression was induced not only by RNS in the growth medium, but also when cells were challenged with activated macrophages. However, the study failed to identify a role for HCP in protection from the host immune system although the authors hypothesized that it might be involved in NO detoxification. Subsequently an NO sensitive transcription factor, NsrR (see below), was found to regulate expression of HCP (Filenko *et al.*, 2007). This corroborates the hypothesis that HCP is involved in protection from RNS. An alternative hypothesis is that HCP is involved in protection from reactive oxygen species (ROS). The results of Almeida *et al.* (2006) demonstrated that HCP expression is induced by hydrogen peroxide and that deletion of OxyR, a hydrogen peroxide sensitive transcription factor, increased HCP expression.

1.4. DNA repair methyltransferases – Ogt and Ada

Ogt and Ada are the two DNA repair methyltransferases (MTases) of *E. coli,* which act as acceptors of methyl and alkyl groups from methylated or alkylated DNA (Potter *et al.*, 1987).

Ada was initially identified in a screen for determinants of DNA MTase activity. While Potter and colleagues were attempting to isolate the gene encoding Ada, they encountered a second, smaller, DNA fragment that encoded an MTase. The second fragment was initially thought to have arisen by partial duplication of *ada* but was found to encode a distinct MTase, Ogt. Ada is both an MTase and a transcription factor that regulates its own expression, and has been shown to be the main regulator in the adaptive response to alkylating / methylating agents (Teo *et al.*, 1986). Ogt was initially reported to be constitutively expressed, although it has since been shown to be induced by nitrate (Squire *et al.*, 2009).

Both Ogt and Ada are able to repair methyl-guanine and methyl-thymine bases, whilst Ada is also able to repair methyl-phosphotriesters (see Figure 1.2). This difference in function of Ogt and Ada is reflected in the domain structure. Ada has 2 distinct domains that possess different functionalities: the NTD contains the DNA binding region and determinants of methyl-phosphotriester MTase activity and the CTD contains determinants of methyl-guanine MTase activity (Shevell and Walker, 1991). Ogt is a smaller protein and is similar to the CTD of Ada (Potter *et al.*, 1987).

Modified bases will often mispair, causing mutations following replication. For example, whilst guanine pairs with cytosine, O6-methyl-guanine pairs with thymine. In many cases, mutations that arise will be silent and therefore harmless, while some mutations may disrupt a vital gene. Mutations that are more serious may arise from methylated / alkylated bases because the mismatch repair system is believed to interact with the replication machinery at such bases, producing a lethal double strand break. Even a single double strand break can be fatal to *E. coli* if not repaired (Lawley and Phillips, 1996). Ogt and Ada are able to transfer methyl and alkyl groups from the modified base onto a cysteine residue. Ogt and Ada have been described as suicidal methyltransferases because the transfer reaction is irreversible and
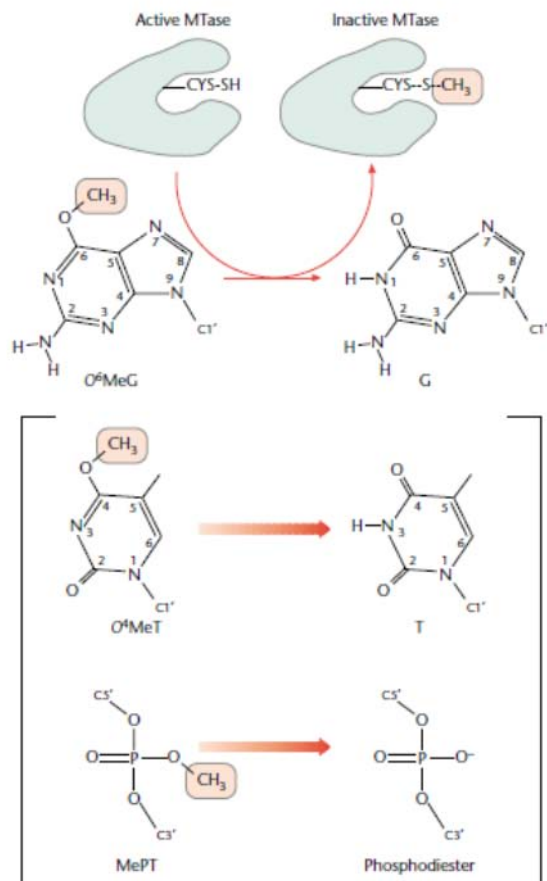
	Reaction catalysed by:
$O^6MeG \rightarrow G$	Ogt / Ada
$O^4MeT \rightarrow T$	Ogt / Ada
MePT $\rightarrow$ Phosphodiester	Ada

Figure 1.2 DNA repair catalysed by Ogt and Ada

From Memisoglu and Samson (2001). The transfer of methyl groups from DNA by DNA repair MTases O^6-methyl-guanine, O^4-methyl-thymine and methyl-phosphotriesters (S-diastereomer only) lesions in doublestranded DNA are recognized by DNA repair MTases and the inappropriate methyl group is transferred to a cysteine residue in the active site of the MTase protein. Methyl transfer inactivates the MTase and so these DNA repair proteins have been called suicide enzymes.

the transfer of a methyl group prevents further MTase activity by the protein. Following transfer of a methyl group, Ogt and Ada are believed to be broken down by proteases.

1.5. Regulation of gene expression

Like all organisms, *E. coli* regulates gene expression in response to stimuli. By sensingvariation in its environment and responding accordingly, *E. coli* is able to adapt to changes in its situation, which is evidenced by an *in silico* study that concluded that regulation of gene expression showed signs of having been "strongly selected for to enable rapid adaptation to environmental conditions" (Cases *et al.*, 2003). Regulation may occur at any stage in gene expression. However, the expression of the majority of genes is believed to be regulated at the stage of transcription initiation by modulating the affinity of RNA polymerase for a promoter or controlling the formation of the elongation complex (covered in detail below). Gene expression can also be regulated by factors that affect mRNA stability, which dictates how long a transcript persists in the cell and is available for translation. RNA can also adopt a secondary structure that may be influenced by the environment and can promote or hinder translation. Regulation of gene expression, particularly that at the stage of transcription initiation, ensures that minimal cellular resources are wasted on the unnecessary production of transcripts and ensuing proteins (Stoebel *et al.*, 2008).

1.6. The bacterial RNA polymerase

E. coli contains a single DNA-dependent RNA polymerase (RNAP) that is responsible for all transcription in the cell. Figure 1.3, Panel A shows the structure of a bacterial RNAP core enzyme and shows that it is a multi-subunit enzyme, with subunits $\beta\beta'\alpha_2\omega$ comprising the core enzyme (Zhang *et al.*, 1999). Studies have shown that the 'crab-claw' structure of RNAP is conserved from bacteria to man, whilst sequence, structural and functional studies also indicate that individual subunits of the bacterial RNAP show significant homology to relevant proteins in all 3 domains of life (Darst, 2001, Allison *et al.*, 1985).
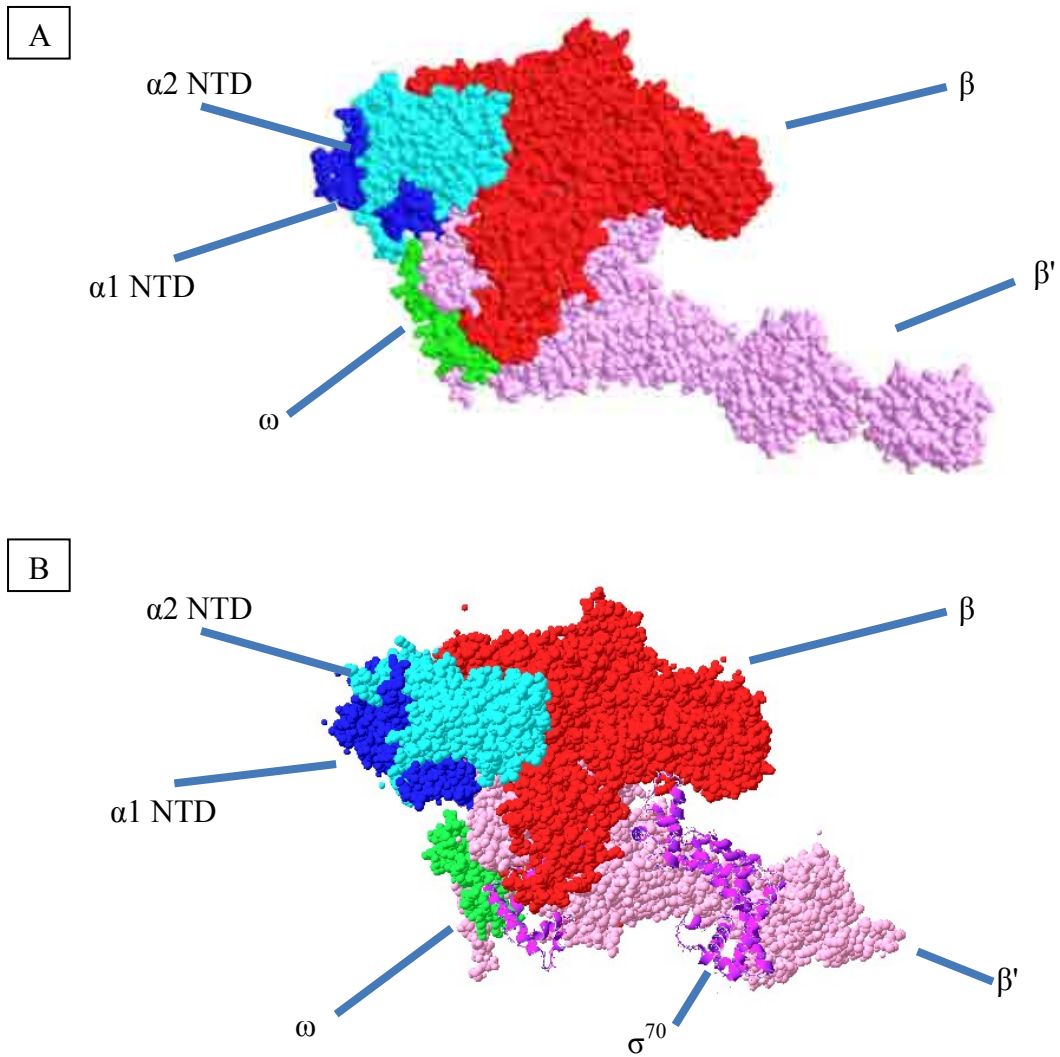
Figure 1.3 Structure of RNA polymerase core enzyme and holoenzyme

Panel A shows the 3D structure of RNAP core enzyme from *T. aquaticus* coloured by subunit. Subunits are also labelled. Structure from Campbell *et al.* (2005), PDB ID 1ynj
Panel B shows the 3D structure of RNAP-σ^{70} holoenzyme coloured by subunit. Sigma 70 is depicted as a ribbon. Structure from Vassylyev *et al.* (2002), PDB ID 1iw7

The active site of bacterial RNAP is formed by the interaction of the large β (beta) and β' (beta prime) subunits and contains a universally conserved Mg^{2+} ion at the centre (Zhang *et al.*, 1999). The detailed structure of *T. aquaticus* RNAP also reveals channels, formed by the interaction of β and β'subunits, that permit DNA access to the active site during transcriptionand provide an exit for the nascent RNA chain (Murakami and Darst, 2003).

The two chemically identical α (alpha) subunits each contain a distinct amino terminal domain (NTD) and a carboxyl terminal domain (CTD) that are joined by a flexible linker (Blatter *et al.*, 1994). The αNTD interacts heavily with either β or β', and is the initiator for RNAP assembly, which occurs in the order $2\alpha \rightarrow \alpha_2 \rightarrow \beta\alpha_2 \rightarrow \beta\beta'\alpha_2 \rightarrow \rightarrow \beta\beta'\alpha_2\sigma$ (Ebright and Busby, 1995, Igarashi *et al.*, 1991). The αCTD contains a DNA binding determinant that recognises UP elements at certain promoters (Estrem *et al.*, 1999).

The ω (omega) subunit is conserved throughout all domains of life although has been found to be non-essential during slow growth and *in vitro* transcription experiments (Burgess, 1969, Gentry *et al.*, 1991). Sequence, structural and functional studies indicate that the bacterial ω subunit is similar to the eukaryotic RNAP subunit RBP6, which chaperones the assembly of RNAP and stabilises the overall structure (Minakhin *et al.*, 2001)

The RNAP core enzyme is competent for transcript elongation but is unable to recognise promoter regions specifically. To bind DNA, the core enzyme must incorporate the σ (sigma) subunit, becoming the holo enzyme (Burgess *et al.*, 1969). Figure 1.3, panel B shows the structure of the *Thermus thermophilus* RNA polymerase holoenzyme. The σ subunit is a multi-domain protein that contains DNA binding determinants that allow σ to bind core promoter elements (Burgess and Anthony, 2001). The majority of bacterial σ factors show structural similarity to the primary *E. coli* σ factor, σ^{70}. The primary, or 'housekeeping', σ factor in a species is responsible for directing RNAP to transcribe genes needed for growth of the cell under standard conditions (Wosten, 1998). Bacteria also contain alternative σ factors

that direct RNAP to different classes of promoters, which are typically grouped functionally. An example of an alternative σ factor is the *E. coli* $σ^H$, which directs RNAP to transcribe genes required for the stress response (Hengge, 2009). However, it has been noted that there is significant functional overlap between the genes transcribed by RNAP containing alternative σ factors and those transcribed by $σ^{70}$ containing RNAP (Wade *et al.*, 2006).

1.7. Overview of transcription

Transcription is a process that comprises 4 steps: promoter binding; RNA chain initiation / promoter escape; transcript elongation and transcript termination (Uptain *et al.*, 1997). An overview of transcription initiation is shown in Figure 1.4. Figure 1.4, Panel A, depicts RNAP bound at a promoter in what is known as the closed complex (RP_c). The DNA helix is bound by contacts with the σ subunit and is positioned around the surface of polymerase. RP_c then undergoes a series of conformational changes to become the open complex (RP_o), during which the DNA is partially unwound and the two strands separated to form a bubble that includes the transcription start site and -10 element (Figure 1.4, Panels B and C). It is hypothesised that σ is responsible for separating the strands: The region surrounding the transcription start site is typically AT-rich and, as well as having weaker inter-strand contacts than a GC-rich region, is expected to cause the DNA to 'kink' placing one strand in proximity to conserved residues in region 2.3 of σ. During the change to RP_o, sigma domain 1.1, which blocks the DNA entry channel in the RP_c, is repositioned so that DNA can enter the core of polymerase and the template DNA strand is repositioned so that it is in close proximity to the RNAP active site (Panel C) (Vassylyev *et al.*, 2002, Murakami *et al.*, 2002).

Free NTPs enter through the secondary channel and RNAP synthesises short transcripts in a process called 'abortive initiation' (Hsu, 2002). It is thought that this is because a loop in σ region 3.2 blocks the RNA exit channel and prevents transcripts from exiting polymerase (Murakami and Darst, 2003). Eventually, a transcript of >12 nt is synthesised, which
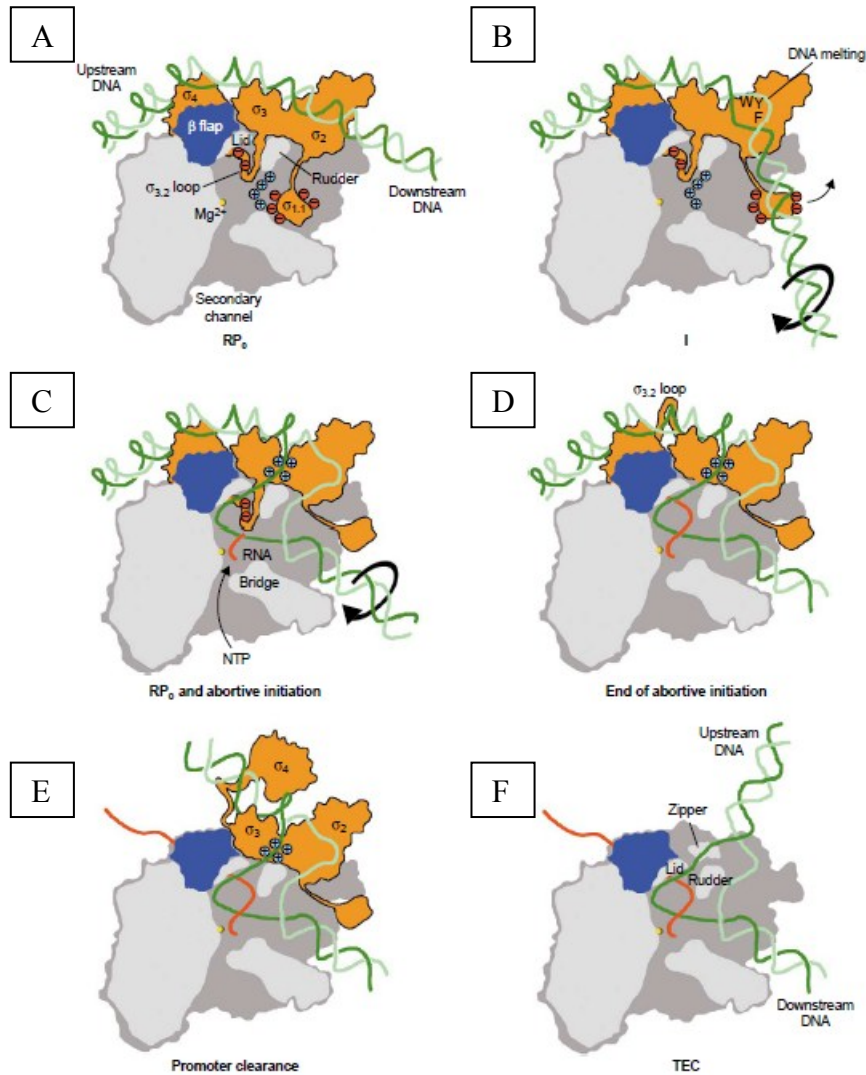
Figure 1.4 Overview of transcription initiation

From Murakami and Darst (2003). Structural transitions during the steps of transcription initiation. Cross sectional views of the RNAP holoenzyme are shown (β flap, blue; σ, orange; rest of RNAP, gray; catalytic Mg^2, yellow sphere), promoter DNA (template strand, dark green; non-template strand, light green; -10 and -35 elements, yellow) and the RNA transcript (red) at the closed complex (RP_c) (panel A), intermediate (I) (Panel B), open complex (RPo) and abortive initiation (panel C), end of abortive initiation (panel D), promoter clearance (panel E) and transcription elongation complex (panel F) stages of transcription initiation. The view is looking down on top of the β subunit, but with most of β removed, revealing the inside of the RNAP active site channel.

repositions the loop in σ region 3.2 so that the channel is no longer blocked (Figure 1.4, panel D). RNAP then clears the promoter and commences transcription as the elongation complex (Figure 1.4, panels E and F). It is thought that in order to clear the promoter, σ must dissociate from polymerase and the repositioning of σ by the nascent RNA strand may sufficiently weaken the interaction of σ and polymerase so that dissociation occurs. However, a growing body of work suggests that it is not necessary for σ to dissociate for transcription to occur and that some polymerase molecules may retain σ throughout transcription (Abbondanzieri *et al.*, 2005, Mooney *et al.*, 2005).

RNAP is thought to transcribe from the DNA template one base at a time using a 'Brownian ratchet' (Herbert *et al.*, 2008). In this model, RNAP fluctuates at each position between the pre- and post-translocated position until the incorporation of an NTP stabilises it. Rates of nucleotide incorporation have been reported to be around 100 nt / second *in vivo* (Vogel and Jensen, 1994). RNA polymerases are processive enzymes, meaning that if it releases the RNA transcript before it has completed the transcript, it is unable to rebind the transcript and continue (Uptain *et al.*, 1997). Transcriptional pausing and arrest are therefore mechanisms for RNA polymerase to overcome problems such as low NTP availability by simply pausing until transcription is once again possible. At some pause sites, RNAP backtracks along the DNA, which removes the 3′ end of the RNA transcript from the active site. Elongation factors GreA and GreB bind to RNAP and stimulate cleavage of the backtracked portion of the transcript, allowing RNAP to re-attempt transcription of the region (Borukhov and Severinov, 2002). Transcriptional pausing and arrest can also be induced by other factors such as regulatory proteins and mRNA secondary structure, and are recognised points of regulation for the expression of certain genes (Artsimovitch and Landick, 2000).

Transcription continues until it is terminated by one of two methods: intrinsic termination or factor dependent termination. Intrinsic termination is reported to be the mechanism of

termination for ~50% of genes and is characterised by a 3′-terminal DNA sequence that encodes a hairpin loop, followed by a run of T residues (Lesnik *et al.*, 2001). Such a structure is thought to induce termination by disrupting contacts between RNAP and the nucleic acids being processed. The run of T residues is reported as introducing a transcriptional pause exactly at the point of termination and thereby allows the hairpin to form (Gusarov and Nudler, 1999). Factor dependent termination is where an additional protein induces termination. There are currently 2 known host termination factors: Rho, which acts to terminate transcription of many genes; and MFD, which terminates transcription when polymerase is stalled by DNA lesions (Nudler and Gottesman, 2002). Rho binds to extended sites in the RNA transcript and then migrates along the transcript until it reaches polymerase (Jin *et al.*, 1992) but is not able to terminate transcription without interaction with NusG (Sullivan and Gottesman, 1992). It has been recently suggested that during coupled transcription / translation, NusG is interacts with the ribosomal protein NusE. Once the ribosome has reached the end of an operon, NusG is released, and is able to interact with Rho to terminate transcription (Burmann *et al.*, 2010, Proshkin *et al.*, 2010). Crucial transcripts that are not occupied by ribosomes, such as rRNAs, are protected from Rho-dependent termination by an anti-termination complex comprising NusA, NusB, NusE and NusG (Condon *et al.*, 1995).

1.8. Promoter recognition by RNAP

Having integrated a σ subunit and become the holoenzyme, RNAP is thought to locate promoters by moving randomly about the chromosome using a combination of migrating along the DNA and changing strands (von Hippel and Berg, 1989). A σ^{70}-dependent promoter can contain up to 4 core types of element that are recognised by the holoenzyme: the -10 element; the extended -10 element; the -35 element and the UP element. Figure 1.5 shows the RNAP holoenzyme interacting with the core promoter elements.
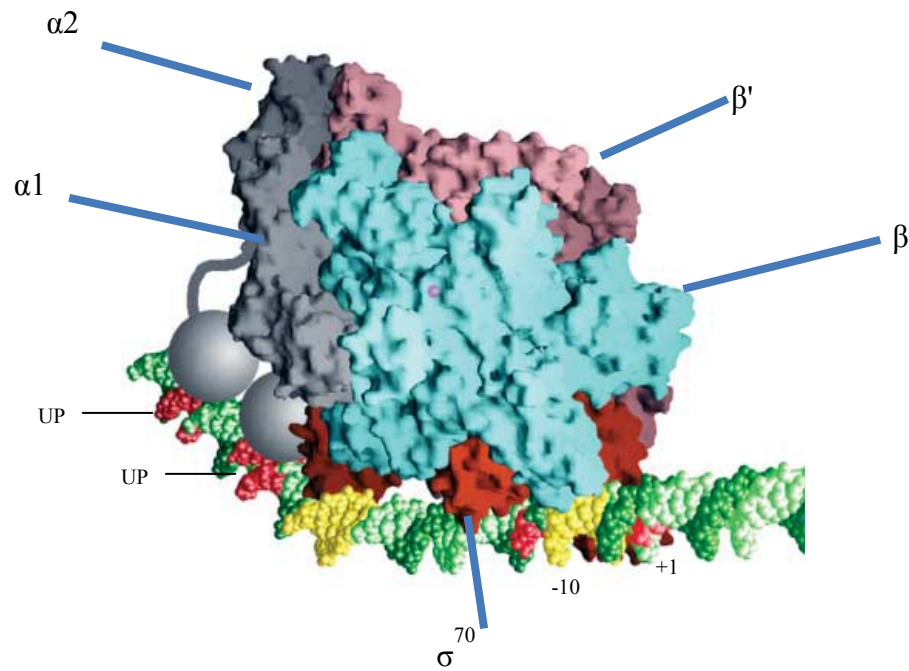
Figure 1.5 RNA Polymerase bound at a promoter

From Browning *et al.* (2004). A model based on crystallographic studies of the initial docking of the RNA polymerase holoenzyme to a promoter. The DNA strands are shown in green, with the −10 and −35 elements highlighted in yellow and the TGn extended −10 and the UP elements highlighted in red. RNA polymerase is shown with the β and β′ subunits coloured light blue and pink, respectively, αNTDs are coloured grey and the different domains of σ are coloured red. Grey spheres labelled I and II, represent the domains of αCTD that bind to the promoter. The RNA polymerase active site is denoted by the Mg^{2+} ion (magenta).

The -10 element, sometimes referred to as the Pribnow box, is a hexamer with the consensus sequence TATAAT and is located ~10 bp upstream of the transcription start site (Pribnow, 1975). High resolution structural studies indicate that the -10 element is recognised by a helix-turn-helix DNA binding determinant in region 2.4 of σ (Murakami *et al.*, 2002). The extended -10 element, if present in a promoter, is found immediately upstream of the -10 element and has a consensus sequence TGn, where 'n' is any base. The extended -10 element is typically found at promoters with non-consensus -35 elements and is believed to improve transcription from otherwise weak promoters (Mitchell *et al.*, 2003). Studies indicate that the extended -10 element is recognised by region 3 of σ^{70}, which supports its classification as a distinct element from the -10 element (Sanderson *et al.*, 2003).

The -35 element is a hexamer with the consensus sequence TTGACA and is located ~35 upstream of the transcription start (Rosenberg and Court, 1979). Numerous studies have shown that the spacing between the -10 and -35 elements is crucial for promoter activity, and that the optimum distance between the 2 sites is 17 bp. The -35 element is recognised by a helix-turn-helix domain in region 4.2 of σ^{70} (Russell and Bennett, 1982, Hidalgo and Demple, 1997).

At certain promoters, RNAP makes contacts with the promoter through the αCTD in addition to those contacts made through σ^{70}. The αCTD recognises an A / T rich sequence called an UP element that lies upstream of the -35 element and is around 20 bp long (Estrem *et al.*, 1999). UP elements are found at a variety of promoters although are often found at promoters for rRNAs and tRNAs (Gourse and Gaal, 2000). This is likely to be because such genes need to be expressed at high levels and UP elements can improve transcription of a gene by between 10 and 100-fold (Ebright and Busby, 1995).

The strength of a promoter is generally considered to be a function of the resemblance of its core promoter elements to the consensus sequences for such elements (Hawley and

McClure, 1983). Therefore, a promoter with core promoter elements that closely resemble the consensus elements will be highly active (Rhodius and Mutalik, 2010). It should be noted however, that no natural promoter contains core elements that perfectly match the consensus sequences as such a promoter would bind RNAP tightly and prevent it from clearing the promoter and initiating transcription (Ellinger *et al.*, 1994).

1.9. Regulation of transcription initiation

The majority of genes in *E. coli* are not constitutively expressed. Instead, their expression is regulated, commonly during transcription initiation. Transcription initiation is regulated by transcription factors, which are able to bind DNA and up-regulate or down-regulate transcription.

Different transcription factors regulate the expression of different numbers of operons. Figure 1.6 shows an overview of the transcriptional regulatory network of *E. coli*. A small number of transcription factors, such as CRP, have been designated 'global regulators' because they regulate large amounts of the genome. Martinez-Antonio and Collado-Vides (2003) proposed that there are 7 such regulators in *E. coli* that regulate in response to stimuli such as aerobiosis, anaerobiosis or the presence of glucose. The remaining transcription factors fall into 2 loose groups, a 'middle management' layer that regulate the expression of multiple genes but are not considered global transcription factors and 'local' transcription factors that regulate the expression only one or two operons.

The activity of transcription factors themselves can be regulated by 4 mechanisms: binding of a ligand; covalent modification of the factor; sequestration of the factor or expression level of the factor. Transcription factors that bind a ligand can be activated or deactivated by the ligand in question. An example would be the *lac* repressor, LacI. When LacI binds its ligand allolactose, it becomes unable to recognise its DNA site and thus regulates expression in response to lactose (Lewis, 2005). Transcription factors can also be
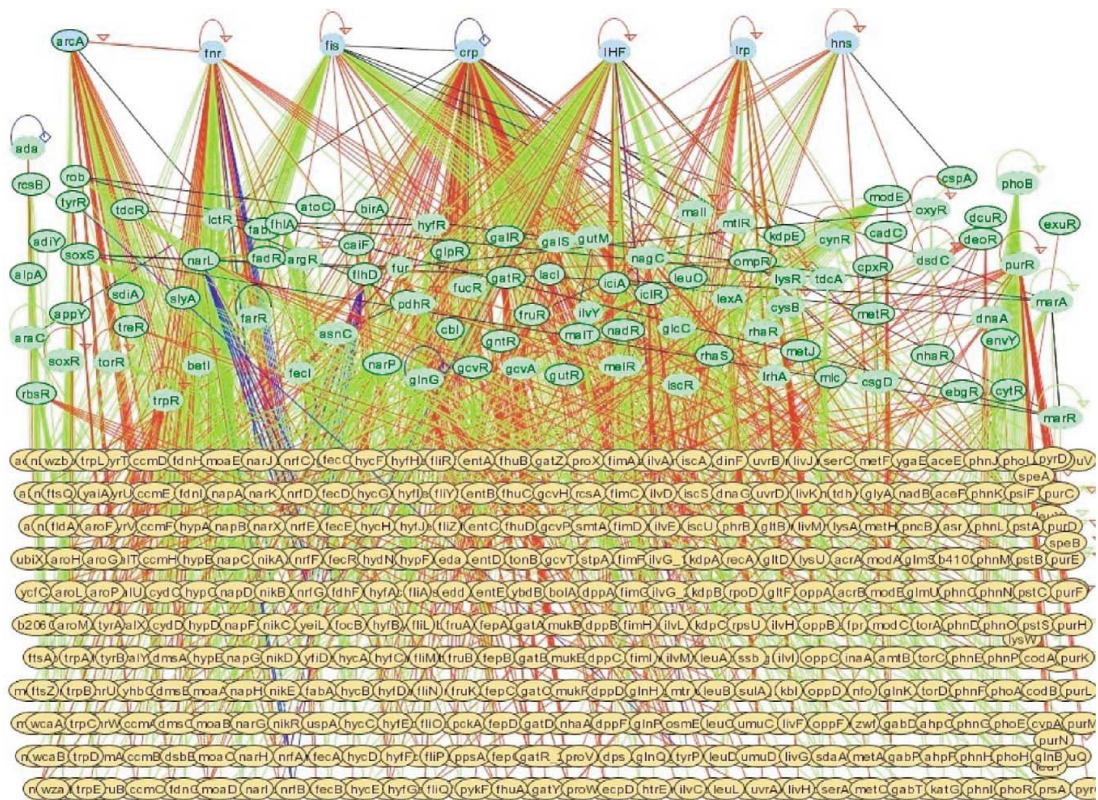
Figure 1.6 Global and local transcription factors in *E. coli*

From Martinez-Antonio and Collado-Vides, (2003). Overview of the transcriptional regulatory network of *E. coli*. Regulated genes are shown as yellow ovals, TFs are shown as green ovals and TFs considered to be global regulators are shown as blue ovals. The green lines indicate activation, red lines indicate repression and dark blue lines indicate dual regulation (activation and repression).

covalently modified by chemicals or other proteins, which cause a conformational change that activates or de-activates the factor. In some cases, it can be difficult to distinguish between activation of a transcription factor by ligand binding and activation of a factor by covalent modification. For example, NsrR (covered in detail below) is deactivated by NO modification of an iron-sulphur cluster. Regulation of NsrR by NO could be described as either ligand interaction or covalent modification. Transcription factors can be sequestered in the cytoplasm by other proteins, which prevents them from being able to regulate (Plumbridge, 2002). Finally, the activity of some transcription factors (such as SoxS) is regulated by the expression level of the factor itself (Demple, 1996).

The majority of transcription factors in bacteria are one-component systems (Ulrich *et al.*, 2005), meaning that the transcription factor does not require another protein to activate it in response to a stimulus. However, there are a number of two-component systems (TCS) in bacteria that play important roles in regulation (Oshima *et al.*, 2002). A TCS is a system comprised of a sensor kinase protein and a response regulator protein (Hoch, 2000). The sensor kinase protein is often transmembrane and interacts with its ligand present in the periplasm. When the sensor kinase protein interacts with its ligand, it auto-phosphorylates on its cytoplasmic region. The phosphorylated sensor kinase then phosphorylates its cognate response regulator and therefore activates it by covalent modification. The response regulator typically undergoes a conformational change once phosphorylated, which makes it competent to bind DNA. TCSs are therefore an effective method for regulating in response to an extracellular ligand. Research suggests that despite the fact that two-component systems function by similar mechanisms, there is little biologically significant cross talk between non-related systems (Bijlsma and Groisman, 2003). Many sensor kinase proteins aid the prevention of crosstalk by dephosphorylating the response regulator in the absence of the stimulating ligand. Thus, if the response regulator is activated by a non-cognate sensor kinase,

it will be rapidly dephosphorylated by its own sensor kinase and have little effect on gene expression.

1.10. Simple activation of transcription

There are 3 basic mechanisms by which transcription factors can induce transcription. The first, activation by contacting RNAP / recruitment of RNAP to a promoter, can be divided into 3 subgroups: Class I; Class II and Class III activation. Figure 1.7, panel A, depicts class I activation. In class I activation, a transcription factor binds to a DNA site upstream of the -35 element and recruits RNAP to the promoter through contacts with the αCTD. Because the αCTD is connected to the αNTD via a flexible linker, the DNA site to which a class I activator binds can be many bp upstream of the transcription start. An example of a class I activated promoter is that for the *lac* operon. At the *lac* promoter, the transcription factor CRP binds to a site centred at position -61.5 relative to the transcription start site (Zhou *et al.*, 1994). From this site, CRP contacts the αCTD, which recruits polymerase to the promoter and initiates transcription.

Figure 1.7, panel B depicts class II activation. In class II activation, a transcription factor binds to a site that overlaps the -35 element and recruits polymerase to the promoter. By binding to such a site, the transcription factor is positioned in-between the αCTD and polymerase and it able to make multiple contacts with polymerase. A transcription factor activating by a class II mechanism can contact the αCTD, αNTD, upstream region of σ or even the β subunit. Some transcription factors will make a single contact while others will make multiple contacts. An example of class II activation is at the *nfo* operon promoter where the SoxS transcription factor binds to a site centred at -37 (Li and Demple, 1994).

Figure 1.7, panels C and D, depicts 2 derivatives of class III activation, which can be viewed as a hybrid of class I and II activation or as two instances of class I activation. At the broadest level, class III activation is when 2 transcription factors bind to a promoter and
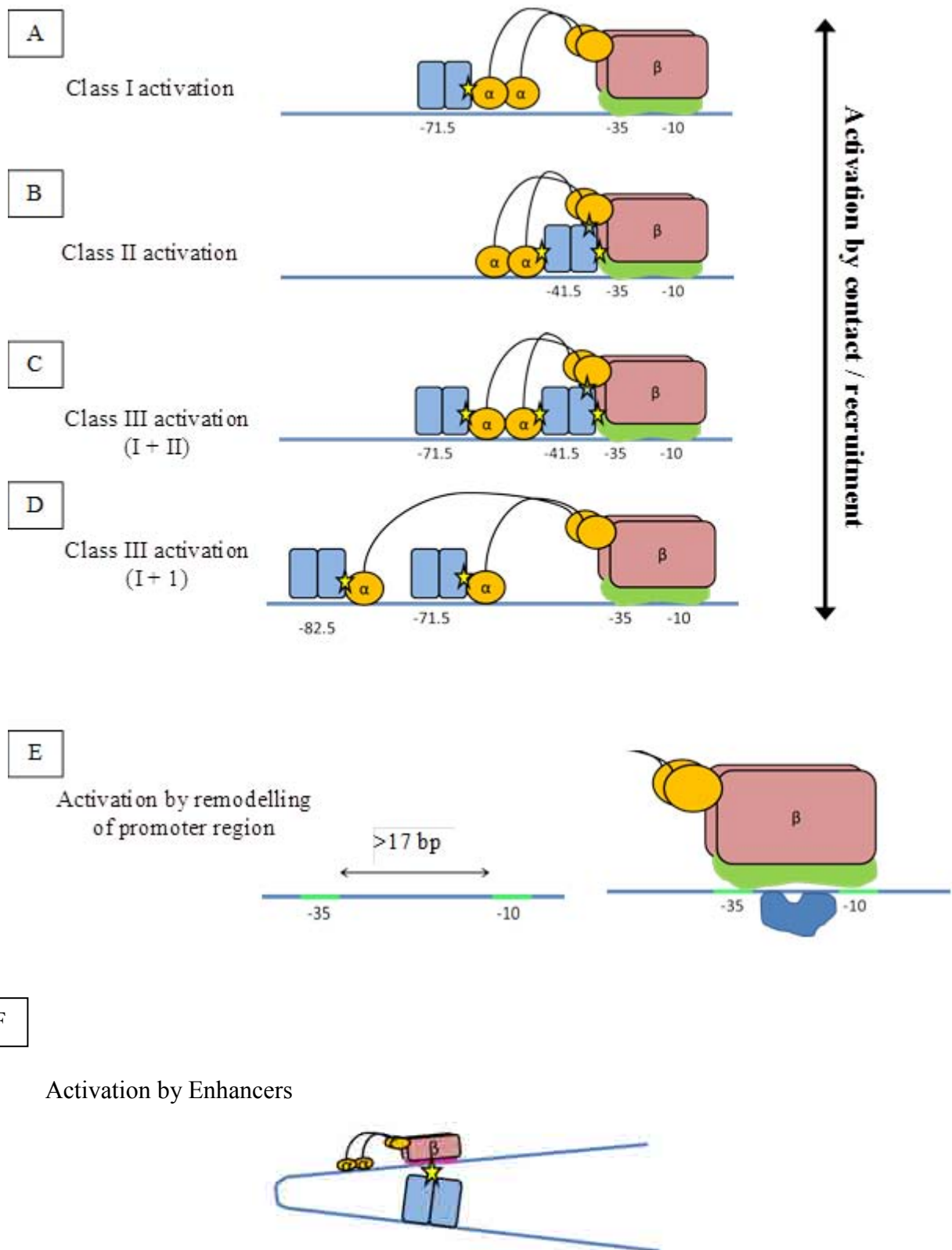
Figure 1.7 The classes of activation

Diagrams depict RNAP bound to a promoter and activated by various classes of activation. Blue rectangles denote a transcription factor. The RNAP α subunits are depicted as yellow circles connected by a flexible linker. σ^{70} is depicted as a green shape incorporated into the RNAP, σ^{54} is depicted as a purple shape incorporated into RNAP and yellow stars indicate possible interactions between the transcription factor and polymerase subunits.

contact polymerase to up-regulate promoter activity. Class III activation can manifest in 2 forms: co-dependent class III activation, in which both transcription factors are required for activity of the promoter and synergistic class III activation, in which binding of the promoter by the second transcription factor merely enhances promoter activity beyond that activated by the first (Scott *et al.*, 1995). Class III activation therefore allows for multiple possibilities in regulation: The two transcription factors involved can be multiple molecules of the same factor or can be different factors, which allows a promoter to be regulated in response to two different stimuli (Joung *et al.*, 1994, Joung *et al.*, 1993). The sites to which the transcription factors bind can either both be upstream of the -35 element or one will overlap the -35 element whilst the other is upstream. An example of class III activation is that of the *ansB* promoter, activity of which is activated by dimers of CRP and FNR binding to DNA sites at positions -91.5 and -41.5 respectively (Scott *et al.*, 1995).

The second mechanism by which transcription can be activated is by promoter remodelling. Some promoters contain core promoter elements that closely resemble the consensus sequences for such elements, but with sub-optimal distances between those elements, which prevents RNAP from binding to them (see Figure 1.7, panel E). Transcription factors of the MerR family are able to bind multiple DNA sites near the core promoter elements and distort the chromosome so that RNAP can recognise the promoter (Summers, 1992, Brown *et al.*, 2003). An example of such a factor is Mta, a transcription factor from *B. subtilis* that binds a symmetrical sequence in the 19 bp spacer of its own promoter and that of another gene. From this site it is able to manipulate the DNA to bring the -35 and -10 elements into correct alignment, thereby activating transcription (Baranova *et al.*, 1999).

The final mechanism of activation depicted in Figure 1.7, panel F, is activation by enhancer binding proteins (EBPs). RNAP-σ^{54} holoenzyme binds to promoters but is unable to

form the open complex and initiate transcription because of a repressive DNA structure. An EBP molecule binds to an enhancer upstream of the promoter and loops the DNA so that it can contact polymerase. Upon contacting polymerase, the EBP uses energy from ATP hydrolysis to cause changes in the RNAP-σ^{54} structure that allows open complex formation (Rappas *et al.*, 2007).

1.11. Simple repression of transcription

Transcription factors can repress activity by binding to promoters and interfering with one or more stages of transcription initiation. There are three broad mechanisms by which they are able to do so: steric hindrance; DNA bending and anti-activation. Steric hindrance is a very common mechanism of repression and is characterised by a transcription factor preventing RNAP or an activator from binding to the DNA by physically obscuring access to the DNA. A repressor typically does this by binding to a site that is near to or overlaps a core promoter element or target site for an activator. An example of steric hindrance is regulation of the *gcd* gene by CRP. CRP binds to a DNA site centred at position -31.5, which prevents RNAP from binding the promoter (Izu *et al.*, 2002).

Repression by DNA bending is another common mechanism of repression. A transcription factor binds to a site or multiple sites upstream of the promoter and distorts the structure of the DNA in such a way that transcription initiation is impaired. An example of repression by DNA bending is regulation of the *nirB* promoter by IHF and Fis. Transcription of *nirB* is repressed by IHF and Fis bound to DNA sites centred at positions -88 and +23 / -97 / -142 respectively. By binding to these sites, IHF and Fis are hypothesized to contort the *nirB* promoter region into a structure that prevents transcription (Browning *et al.*, 2000).

Repression by anti-activation is when a transcription factor binds to a promoter near an activator and interferes with its function. In contrast to the above mechanisms of repression, which are widespread, anti-activation following promoter binding by the repressor is rare.

Two examples of note are that of CytR, which negatively regulates activity of several promoters by binding to a DNA site in between 2 CRP sites and that of FNR, which has been reported to interfere with activation by another molecule of FNR at the *yfiD* promoter (Green and Marshall, 1999). CytR anti-activates by binding to its DNA site and then interacting with the nearby CRP molecules and preventing them from activating transcription (Holt and Senear, 2010). It should be noted that anti-activation, where a protein interferes with an activator, is a more widespread mechanism of regulation, but since the vast majority of anti-activators do not bind DNA, they are not classified as transcription factors.

1.12. Complex regulation of transcription

Although some promoters are regulated by a single transcription factor that functions by one of the mechanisms described above, the expression of 49% of genes is regulated by multiple transcription factors (Martinez-Antonio and Collado-Vides, 2003). This is thought to allow the fine-tuning of transcription responses to stimuli and ensure that genes are only expressed when it is beneficial for the cell to do so.

For example, a promoter may be positively regulated by a transcription factor in response to stimulus A, but repressed by a different transcription factor in response to stimulus B. Such regulation allows the *lac* operon promoter to be activated in response to the absence of glucose by the transcription factor CRP, but repressed in response to the absence of lactose by LacI (Lewis, 2005). This makes sense for the cell, as there is little benefit in producing proteins to utilise lactose if there is a better carbon source (glucose) available and / or lactose is not present.

Transcription factors are known to bind to some DNA sites preferentially over others. Most transcription factors have a consensus binding sequence, i.e. a DNA site to which they optimally bind. However, not all examples of a DNA site for a transcription factor will perfectly match the consensus sequence and this gives rise to a hierarchy of target sites for a

transcription factor: Sites that better resemble the consensus sequence will be bound preferentially over those that poorly resemble the consensus sequence. A hierarchy of target sites allows the activity of a transcription factor to be analogue as opposed to digital, i.e. simply 'on' or 'off'. While individual molecules of a transcription factor will be active or inactive, a cell will contain hundreds or thousands of molecules of a particular transcription factor. Therefore, in lower concentrations of the stimulus for that factor, only a subset of molecules will be active, which allows for concentration dependent regulation of gene expression. For example, the *yfiD* promoter contains 2 DNA sites for the transcription factor FNR, which is active in the absence of oxygen (Marshall *et al.*, 2001). In micro-aerobic conditions, small amounts of FNR will be active and will bind preferentially to the site at position -40.5 relative to the *yfiD* transcription start, which better resembles the consensus FNR binding sequence than does the site at position -93.5. When bound solely to the DNA site at position -40.5, FNR activates transcription of the *yfiD* gene. However, in the total absence of oxygen, there is enough active FNR that the second site at position -93.5 will also be occupied and FNR then represses activity of the promoter.

The combined effect of multiple factors regulating a single promoter and concentration dependent binding of DNA sites can make understanding complex regulation of a promoter difficult. Many promoters will not simply be active or inactive and transcriptional activity will instead be a function of the interplay of the various transcription factors combined with the inherent activity of the core promoter elements. An example of the problem posed by such complexity is that of regulation of the *nrfA* promoter, to which up to 8 transcription factors have been shown to be involved (Browning *et al.*, 2010). In the face of such complexity, a reductionist, 'bottom up' approach is often taken. Hence, the actions and roles of individual transcription factors will be studied, often in the absence of other factors, and then the models combined to attempt to understand total regulation of the promoter.

The transcription factors that have been studied in this work are FNR, NarL / NarP and NsrR, and are introduced below.

1.13. FNR

FNR is the master regulator of anaerobic adaption in *E. coli* and regulates the transcription of over 100 genes in response to the absence of oxygen. Because of the number of genes that FNR regulates, it is considered to be one of the 6 global transcription factors of *E. coli*. FNR shows significant similarity to the global transcription factor CRP (Shaw *et al.*, 1983). However, because of the sensitivity of FNR to oxygen, no high-resolution structures have been obtained (Yan and Kiley, 2009). Alignment of the amino acid sequence of FNR to that of CRP (the structure of which has been determined) suggests that FNR has a distinct CTD and NTD. The CTD contains a DNA binding determinant that closely resembles that of CRP and the NTD contains an iron-sulphur cluster, through which FNR senses oxygen.

It has been suggested that FNR protein is synthesised and loaded with a 4Fe-4S cluster by the IscS protein (Schwartz *et al.*, 2000). In such a form, FNR is able to bind to DNA and regulate transcription. Interaction with molecular oxygen causes the iron-sulphur cluster to oxidise into a 3Fe-4S cluster, rendering FNR unable to bind DNA (Crack *et al.*, 2007). The 3Fe-4S cluster is then thought to degrade spontaneously to become a 2Fe-2S cluster that is found when FNR is recovered in aerobic conditions.

There is still debate as to how FNR becomes reactivated once the iron-sulphur cluster has been degraded. First, it is possible that the degradation of the cluster is reversible and that, in the absence of oxygen, the 2Fe-2S cluster can become a 4Fe-4S cluster. This is unlikely to occur unaided, and could require one of the *E. coli* iron-sulphur cluster repair proteins (Vine *et al.*, 2010). A more recent observation noted that FNR is a target for the ClpX protease implying that when FNR contains a 2Fe-2S cluster, it is simply broken down and a novel FNR molecule synthesized and loaded with a 4Fe-4S cluster (Flynn *et al.*, 2003). However,

the results of Dibden *et al.* (2005) show that FNR containing a 2Fe-2S cluster has a half-life of 45 minutes in the presence of ClpX. Dibden and colleagues also demonstrated that a molecule of FNR that had previously been deactivated by oxygen can become active again. This indicates that FNR is not merely broken down once deactivated and that the iron-sulphur cluster may be repaired or replaced. They suggest that FNR is constantly supplied with 4Fe-4S clusters by the iron-sulphur cluster assembly proteins.

NO is also able to deactivate FNR by interacting with the iron-sulphur cluster. NO disrupts the 4Fe-4S cluster in multiple stages, yielding a dinitrosyl-iron-thiol complex (DNIC) (Pullan *et al.*, 2007). NO deactivation of FNR has been demonstrated to have a regulatory effect *in vivo*. The *hmp* gene encodes an NO-detoxifying flavohemoglobin and expression of *hcp* is repressed by FNR. FNR repression of *hmp* promoter activity is relieved by NO, which suggests that FNR may function as a regulator in response to NO as well to anaerobiosis (Cruz-Ramos *et al.*, 2002).

When FNR is in its active form it is able to dimerise and bind DNA. FNR recognises a 5-4-5 inverted repeat with the consensus sequence TTGATnAAnATCAA (Jayaraman *et al.*, 1989). It has been noted that the central 4 positions, although not crucial for recognition of the site by FNR, are important for the degree of activation (Scott *et al.*, 2003). Work by Scott and colleagues showed that promoters with DNA sites for FNR that contain A / T pairs in the central 2 positions are induced an order of magnitude more efficiently than those with sites that do not contain central A / T pairs.

FNR activates transcription by a class II mechanism at the majority of promoters from a site centred at position -41.5. An initial study examining the spacing requirements of FNR reported that FNR can only activate efficiently from a class II position (Bell *et al.*, 1990). However, a later study found that FNR can activate by a class I mechanism if the -35 element is a sufficiently good match to the consensus sequence for such elements (Wing *et al.*, 1995).

The work by Wing and colleagues also showed that FNR activates by a class I mechanism most efficiently from DNA sites at positions -61.5, -71.5, -82.5 and -92.5. They also noted that activity of synthetic promoters were twice as high if FNR bound at a class II position than if FNR bound at a class I position.

Like CRP, FNR is able to activate transcription by directly contacting RNA polymerase and recruiting it to a promoter. There are 3 defined 'activating regions' that have been identified on the surface of FNR, through which FNR can contact subunits of RNA polymerase when bound to a promoter (Blake *et al.*, 2002).

1.14. NarXL and NarQP two component systems

In *E. coli* the dual two component systems NarXL and NarQP are responsible for detecting external nitrate and nitrite and regulating transcription accordingly (Stewart, 1993). The NarXL two component system is the better characterised system and consists of NarX, a sensor kinase, and NarL, a response regulator. NarQ is a sensor kinase and is homologous to NarX, whilst NarP is a response regulator that is homologous to NarL.

There is cross regulation between the NarXL and NarQP systems (Noriega *et al.*, 2009). NarQ is able to phosphorylate both response regulators NarP and NarL. NarX has a preference for phosphorylating NarL. Therefore, when nitrate is present in the periplasm, both NarX and NarQ will auto-phosphorylate. NarX will then phosphorylate NarL, and to an extent, NarP, whilst NarQ is reported to phosphorylate NarL and NarP without distinction (see Figure 1.8). When nitrite is present in the periplasm, only NarQ will auto-phosphorylate and will then phosphorylate both NarL and NarP. The presence of nitrite causes NarX to de-phosphorylate phospho-NarL. Promoters that are activated by NarL, but not NarP, will therefore have high activity when cells are grown in the presence of nitrate and low activity when cells are grown in the presence of nitrite.

NarX and NarQ are transmembrane proteins that span the cytoplasmic membrane (Stewart, 2003). Due to the transmembrane nature of NarX and NarQ, it has not been possible to obtain high resolution structures of the full length proteins. Sequence analysis reveals that they are 32% identical and 68% similar, although the cytoplasmic domains are more similar than are the periplasmic domains. However, the periplasmic domains contain a conserved region that is thought to be involved in detection of nitrate / nitrite. The cytoplasmic domains contain the residues that are responsible for phosphorylation or de-phosphorylation of partner proteins.

The crystal structure of the NarL protein reveals that NarL has two distinct domains, connected by a flexible linker (Baikalov *et al.*, 1996). The NTD contains the site of phosphorylation by NarX / NarQ and the CTD contains a helix-turn-helix DNA binding element that is similar to those of the LuxR family. The flexible linker is believed to be unordered and allows the NTD to obscure the DNA binding element of the CTD when not phosphorylated. A structure of the NarL CTD bound to DNA shows that the CTD both bends the DNA by 40° and distorts the helical structure when it binds (Maris *et al.*, 2002). The structure of NarP has not been determined. However, alignment of the amino acid sequence of NarP to that of NarL suggests that it adopts a similar structure.

NarL and NarP recognise a common DNA site, a 7-2-7 inverted repeat that has the consensus sequence TACYYMT-nn-AKRRGTA, where Y = C or T, M denotes A or C, K denotes G or T, R denotes A or G and n denotes any base (Darwin *et al.*, 1996). NarL is also able to bind single heptamers, known as half-sites (Darwin *et al.*, 1997). The crystal structure of the NarL CTD bound to DNA shows that NarL binds as a dimer to an inverted repeat. It is unknown if NarL binds to a half-site as a monomer or as a dimer, where one molecule makes non-specific contact with the DNA.

When nitrate is present:
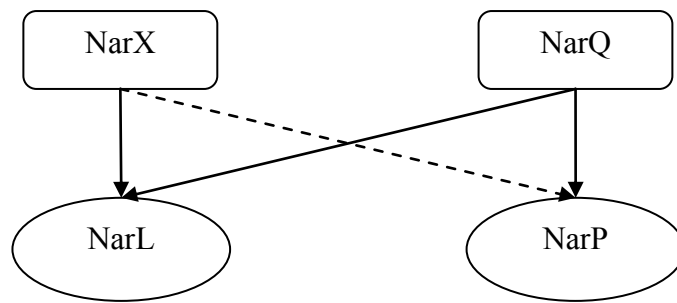
When nitrite is present:
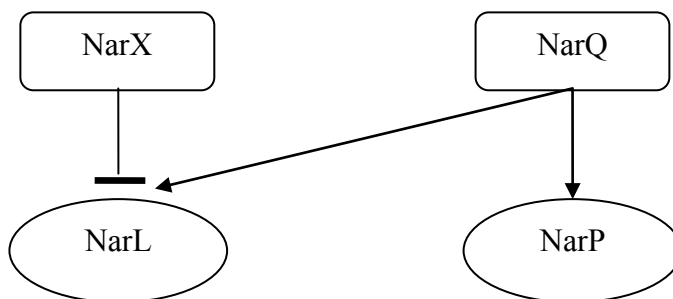
Figure 1.8 Model for NarXLQP phosphorylation

Figure is a diagram that shows the model for NarXLQP phosphorylation proposed by Noriega *et al.* (2009). The sensor kinases NarX and NarQ are depicted as rounded rectangles and sensor kinases NarL and NarP are depicted as ovals. Black arrows indicate phosphorylation, whilst a line that ends with a strong horizontal line indicates dephosphorylation.

NarL and NarP are able to repress or activate transcription depending on the position of the target DNA site relative to other elements at a promoter. If the DNA site for NarL / NarP is located in close proximity to promoter elements or activator binding sites, then NarL / NarP will repress promoter activity by steric hindrance. An example is that of expression of the *dcuS* gene, which is repressed by NarL binding to a site that overlaps the -35 region of the promoter (Goh *et al.*, 2005).

The majority of promoters that are activated by NarL are also activated by FNR (Constantinidou *et al.*, 2006). Co-operative positive regulation by FNR and NarL is logical for genes that encode proteins involved in nitrate utilisation. Because oxygen is a more powerful oxidiser than is nitrate, it is beneficial to the cell to only express nitrate utilisation genes in the absence of oxygen (when FNR is active) and in the presence of nitrate (when NarL is active). At promoters for such genes, FNR typically binds to a DNA site centred at or near position -41.5 and NarL binds to a DNA site centred further upstream. This organisation suggests that FNR and NarL may together activate by a class III mechanism. However, unlike FNR, NarL has not been shown to interact directly with RNA polymerase. Alternatively, it is possible that NarL is unable to activate transcription directly. At the two best understood NarL activated promoters, *nirB* and *nrfA,* NarL has been shown to overcome repressors and does not activate by contacting polymerase (Browning *et al.*, 2004, Browning *et al.*, 2006).

1.15. NsrR

E coli contains 2 transcription factors that sense NO and regulate gene expression in response: NorR and NsrR. NorR senses NO directly and regulates the expression of itself and the NO-detoxifying enzymes NorVW (D'Autreaux *et al.*, 2005). NsrR also senses NO directly and is thought to regulate the expression of around 60 genes (Partridge *et al.*, 2009). NorR is therefore considered to be a specific regulator, whilst NsrR is thought by some to be a global regulator. An attempt was made to identify the NsrR regulon using transcriptomics and

revealed 20 genes that were apparently repressed by NsrR and 22 that were apparently activated by NsrR (Filenko *et al.,* 2007). Genes repressed by NsrR were genes such as those for the periplasmic nitrate reductase *napF* and the periplasmic nitrite reductase *nrfA,* the latter of which has been confirmed by single promoter study (Browning *et al.*, 2010). However, there have been no reports of activation by NsrR from single promoter studies. A more recent ChIP-chip study of chromosomal NsrR binding revealed that NsrR binds to 62 sites in or near 5′ non-coding regions (Partridge *et al.*, 2009). The results of the ChIP-chip study generally supported the results of the previous transcriptomics study and suggested that NsrR may also regulate genes involved in motility and metabolism.

Structural studies of *E. coli* NsrR have been impeded by the oxygen labile nature of the protein. Hence, no structure of NsrR has been reported, although NsrR has been identified as a member of the RRF2 protein family on the basis of sequence similarity. An alignment of the amino acid sequence of NsrR to that of IscR, a better characterised member of the RRF2 family, indicates that NsrR has 2 distinct domains. An NTD that contains a helix-turn-helix DNA binding element and a CTD that contains an iron-sulphur cluster, through which NsrR is thought to sense NO.

It is unclear what form of iron-sulphur cluster the *E. coli* NsrR contains as both 4Fe-4S and 2Fe-2S clusters have been reported in different bacterial strains (Tucker *et al.*, 2008, Yukl *et al.*, 2008). As with FNR, NO is thought to nitrosylate the iron-sulphur cluster of NsrR, resulting in a DNIC. NsrR containing a DNIC is unable to bind DNA and regulate transcription.

NsrR recognises an 11-1-11 inverted repeat with the consensus sequence AAGATGCYTTTnAAARGCATCTT, where Y = C or T and R = A or G (Partridge *et al.*, 2009). NsrR is thought to bind its site as a dimer, although a recent study indicates that NsrR regulates the expression of many genes that contain only a half-site in their regulatory region.

It is not known whether NsrR binds such sites as a monomer or a dimer, with the second molecule making non-specific contact with the DNA. Many DNA sites for NsrR are located close to core promoter elements or activator binding sites. This suggests that NsrR represses activity by steric hindrance.

1.16. Outline of thesis

This thesis is concerned with the architecture of promoters that respond to reactive nitrogen species in *E. coli*. The products of the *hcp* and *ogt* genes of *E. coli* have been implicated in protection from RNS and expression of both genes has been previously shown to be up-regulated by nitrate. Furthermore, NarL, a nitrate-dependent regulator (see above), has been shown to regulate expression of both genes.

Chapter 3 focuses on regulation of the *hcp* gene. Previous work by Nina Filenko and colleagues had shown that expression of *hcp* is positively regulated by FNR, the master anaerobic regulator of *E. coli* and NarL / NarP homologous nitrate / nitrite dependent regulators (Filenko *et al.*, 2005). Filenko later demonstrated that the NO-sensitive repressor NsrR represses expression of *hcp*. The regulation of *hcp* was investigated using promoter fragments fused to *lacZ*. Derivatives of promoter fragments were constructed to probe the relative contributions of the different transcription factors and core promoter elements. *In vitro* protein binding to DNA fragments was also studied.

Chapter 4 focuses on regulation of the *ogt* gene. Previous work by Derrick Squire and colleagues had shown that expression of *ogt* is activated by NarL independently of FNR. NarL activates transcription from the wild type *ogt* promoter by binding to 2 DNA sites (Squire *et al.,* 2009). Promoter derivatives and promoter::*lacZ* fusions were used to study the positions from which NarL and NarP can activate transcription.

Chapter 5 focuses on the mechanisms of activation by the nitrate / nitrite responsive transcription factors NarL and NarP. Plasmids containing RNAP α subunit carrying alanine substitutions in the CTD were used to examine whether there is direct interaction between NarL and RNAP at the *ogt* promoter. NarL and NarP were expressed from plasmids as either full-length constructs or the CTD alone, and the ability to activate transcription was assayed.

Chapter 2: Materials and Methods

2.1. Suppliers

Chemicals were purchased from Sigma-Aldrich or Fisher Scientific unless otherwise stated. Radionucleotides were purchased from Perkin Elmer. Oligodeoxyribonucleotides were synthesised by Alta Biosciences, University of Birmingham. Restriction endonucleases, DNA ladders, Calf alkaline phosphatase, Phusion polymerase, T4 DNA ligase and T4 polynucleotide kinase were purchased from New England Biolabs. Biomix Red DNA polymerase, Biotaq DNA polymerase and dNTP mix were purchased from Bioline. TransformAid and 6 X orange loading dye were purchased from Fermentas. Glycogen was purchased from Invitrogen and DNase I was supplied by Roche. All enzymes / kits were used according to the manufacturer's instructions, and in the buffers provided. All chemicals were stored as instructed by the manufacturer. Purified FNR and NarL protein were kindly donated by Dr. Douglas Browning, University of Birmingham.

2.2. Buffers, solutions and reagents

Solutions for use in bacterial growth or DNA manipulations were autoclaved for 20 minutes + 10 minutes per 100 ml at 120°C and 15 psi, or filter-sterilised using 0.2 μm filters.

2.2.1. Buffers for Gel electrophoresis of DNA

Stock 5 X TBE buffer was purchased from National Diagnostics and contained 0.445 M Tris base, 0.445 M boric acid (pH 8.3) and 2 mM Na_2EDTA. Stock TBE was diluted to 0.5 X in distilled water for use as running buffer in agarose gel electrophoresis and to 1 X for use in polyacrylamide gel electrophoresis. A stock of 7.5% polyacrylamide solution was prepared containing 125 ml of a 30% (w/v) acrylamide, 0.8% bisacrylamide stock solution (ProtoGel - National Diagnostics), 100 ml 5 X TBE, and 20 ml glycerol; and was diluted with distilled water to 500 ml. Sequencing polyacrylamide gels were prepared using the SequaGel Sequencing System, which was purchased from National Diagnostics and used according to

the manufacturer's instructions. A solution containing 10% (v/v) methanol and 10% (v/v) acetic acid was used to fix sequencing gels following electrophoresis.

2.2.2. Buffers for extraction and purification of nucleic acids

TE Buffer was prepared in distilled water and contained 10 mM Tris-HCl and 0.5 mM EDTA (pH 8.0). TES buffer was prepared in distilled water and contained 50 mM NaCl, 60 mM Tris (pH 8.0) and 5 mM EDTA

2.2.3. Buffers used in the preparation of competent cells

Buffer TFB1 was prepared in sterile distilled water and contained 100 mM RbCl, 50 mM MnCl$_2$, 30 mM potassium acetate, 10mM CaCl$_2$ and 15% glycerol (v/v). The pH of the buffer was adjusted to 5.8 and the buffer was filter sterilized prior to being stored at -4 °C. Buffer TFB2 was also prepared in sterile distilled water and contained 10 mM MOPS, 10 mM RbCl, 75 mM CaCl$_2$ and 15% glycerol (v/v). The pH was measured and (if necessary) adjusted to 6.8. Buffer TFB2 was filter sterilized and stored at -4 °C.

2.2.4. Buffers and solutions used in β-galactosidase assays

Z-buffer was prepared in distilled water and contained 75 g/l KCl, 0.25 g/l, MgSO$_4$.7H$_2$O, 8.53 g/l Na$_2$HPO$_4$ and 4.87 g/l NaH$_2$PO$_4$.2H$_2$O. Z-buffer was sterilized and stored at room temperature. Immediately prior to use, Z-buffer was supplemented with 2.70 ml/l β-mercaptoethanol. A 13 mM ONPG (o-nitrophenyl-β-D-galactopyranoside) solution was prepared by dissolving 3.92 g/l of ONPG in Z-buffer.

2.2.5. Buffers used in the purification of cell lysate

Wash buffer used in the purification of cell lysate contained 20 mM Tris-HCl (pH 8.0), 5% glycerol, 1 mM DTT, 200 µg/ml phenylmethylsulfonyl fluoride and 4 µg/ml pepstatin.

2.2.6. Buffers used in electrophoretic mobility shift assays (EMSA)

DNA for use in EMSAs was radiolabelled with a 3000 Ci/mmol preparation of [γ^{32}P]-

ATP (Perkin-Elmer). Unincorporated nucleotides were removed by passing through a Sephadex G-50 column. Sephadex G-50 stock was prepared by adding 5 g Sephadex G-50 (Pharmacia Biotech) into 100 ml TE buffer, which was autoclaved. The suspension was then washed three times in 150 ml TE buffer and finally suspended in 50 ml TE buffer. Sephadex suspension was stored at 4°C. Hepes-gluatamate buffer was prepared as a 10 X stock, which contained 200mM HEPES (pH 8.0), 50 mM $MgCl_2$, 500 mM potassium glutamate and 10 mM dithiothreitol. HEPES dilution buffer contained 10% (v/v) HEPES glutamate buffer and 10% (v/v) BSA. FNR binding buffer was prepared as a 10 X stock, which contained 1 M potassium glutamate, 10 mM EDTA, 100 mM K_2HPO_4, 100 mM KH_2PO_4 (pH 7.5) and 500 µM DTT. Buffer B1 contained 20 mM potassium Epps (pH 8.0), 125 mM potassium glutamate, 5% glycerol (w/v) and 1.5 mM DTT.

2.2.7. *Buffers used in DNase I footprinting*

DNA for use in DNase I footprints was radiolabelled with a 7000 Ci/mmol preparation of $[\gamma^{32}P]$-ATP (Perkin-Elmer). Stocks of FNR binding buffer, buffer B1, HEPES-glutamate buffer and HEPES dilution buffer were prepared as described in section 2.2.6. DNase I stop solution contained 0.3 M sodium acetate (pH 7.0) and 10 mM EDTA. Following the footprinting reaction, samples were mixed with a gel loading buffer that contained 95% (v/v) high quality deionized formamide, 20 mM EDTA, 0.05% bromophenol blue and 0.05% xylene cyanol FF. Samples were separated on a sequencing gel.

2.3. Growth media

2.3.1. *Liquid media*

All liquid media were prepared using distilled water and autoclaved for 20 minutes + 10 minutes per 100 ml at 120°C and 15 psi, and stored at room temperature. Medium was cooled to room temperature before the addition of supplements such as antibiotic. LB medium contained 20 g/l tryptone, 10 g/l yeast extract and 10 g/l NaCl. Minimal salt medium contained

4.5 g/l KH$_2$PO$_4$, 10.5 g/l K$_2$HPO$_4$, 1 g/l (NH$_4$)$_2$SO$_4$, 0.5 g/l sodium citrate, 0.1 g/l; MgSO$_4$.7H$_2$O, 1 nM sodium selenate, 1 nM ammonium molybdate and 1 ml/l sulphur-free trace metal solution (Cole *et al.*, 1974). SOC broth contained: 2% (w/v) tryptone (pancreatic digest of casein), 0.5% (w/v) yeast extract, 8.6 mM NaCl, 2.5 mM KCl, 20 mM MgSO$_4$ and 20 mM glucose. Purchased pre-prepared from Sigma-Aldritch. Stock solutions of sodium nitrate and sodium nitrite were prepared in distilled water and sterilised by autoclave treatment. Nitrate-supplemented media contained a final concentration of 20 mM sodium nitrate and nitrite-supplemented media contained a final concentration of 2.5 mM sodium nitrite.

2.3.2. *Solid media*

Solid media were made up using distilled water and autoclaved for 20 minutes + 10 minutes per 100 ml at 120°C and 15 psi. Plates were poured immediately, allowed to cool and then were stored at 4°C. Supplements were added prior to pouring once media had cooled. Immediately prior to use, plates were dried in a 60°C oven for 10 minutes. LB agar contained: 20 g/l tryptone, 10 g/l yeast extract, 10 g/l NaCl, 10 g/l agar. MacConkey-Lactose agar contained 50 g/l MacConkey lactose agar (Difco)

2.3.3. *Antibiotics*

All stock solutions were made up in distilled water unless otherwise specified and sterilised using 0.2 µm filters. Ampicillin stock was prepared containing 50 mg/ml ampicillin and was stored at -20°C. Kanamycin stock was prepared containing 10 mg/ml kanamycin and was stored at -20°C. Tetracycline stock was prepared containing 20 mg/ml tetracycline in methanol and was stored at -20°C.

Antibiotics were added to medium to select for strains carrying antibiotic resistance markers either chromosomally or on a plasmid. The following final concentrations were used: 100 µg/ml ampicillin, 50 µg/m kanamycin and 20 µg/ml tetracycline.

41

2.4. Bacterial strains

The bacterial strains used in this study are listed in Table 2.1. Aseptic technique was used at all stages of bacterial growth. Strains were maintained as glycerol stocks at -80°C in LB medium with 15% glycerol. Before use, strains were streaked onto an LB plate and incubated at 37°C overnight. For overnight cultures, a conical flask was prepared with 2 X LB and relevant antibiotics and inoculated with a single colony picked with a wooden toothpick. The flask was then incubated overnight at 37 °C with 200 rpm shaking. Growth was monitored by measuring the optical density at 650 nm (OD_{650}) using a Helios Gamma spectrophotometer.

2.5. Plasmids

Plasmids used in this study are listed in Table 2.2. Plasmid maps are shows in figures.

2.6. Gel electrophoresis of DNA

2.6.1. Agarose gel electrophoresis

DNA fragments greater than 0.2 kb were analysed by agarose gel electrophoresis. A solution of 0.5 X TBE containing between 0.8% and 1.5% agarose was prepared and boiled until the agarose was fully dissolved and then allowed to cool to below 60°C. To permit visualisation of the DNA fragments, Sybrsafe (Invitrogen) was added to the cooled agarose solution to a final concentration of 1 X. The gel was poured using a gel casting plate. DNA samples were mixed with orange loading buffer (Fermentas) and gels were run in 0.5 X TBE buffer at 100 V for 30 to 40 minutes. Gels containing fragments to be excised were imaged using a visible spectrum trans-illuminator (Clare Chemical). In all other cases, a UV trans-illuminator was used.

2.6.2. Polyacrylamide gel electrophoresis

Small (100 mm x 70 mm x 1.5 mm) polyacrylamide gels were used to image DNA fragments smaller than 0.5 kb. Such polyacrylamide gels contained 7.5% (w/v) stock acrylamide, 4% glycerol and 1 X TBE, and were polymerised by adding 0.01 volumes of 10%

Table 2.1 Bacterial strains used in this work

Strain	Description	Origin
JCB387	*E. coli* RV Δ*nir* Δ*lac*	Page *et al.* (1990)
JCB3875	Δ*narP*252::Tn*10*d(Cm) derivative of JCB387	Page *et al.* (1990)
JCB3883	Δ*narL* derivative of JCB387	Tyson *et al.* (1994)
JCB3884	Δ*narL* Δ*narP*252::Tn*10*d(Cm) derivative of JCB387	Tyson *et al.* (1994)
JCB387 Δ*nsrR*	Δ*nsrR* derivative of JCB387	Chismon *et al.* (2010)
JCB3911	Δ*fnr* Cm^R derivative of JCB387	Dr Douglas Browning, University of Birmingham, UK
JCB38849S	Tet^s derivative of JCB387 Δ*narL* Δ*narP*253::Tn*10*dCm Δ*him*A452::Tn*10*dTc isolated as a fusaric acid resistant colony	Chismon *et al.* (2010)
JCB38849S Δ*nsrR*	Δ*nsrR* derivative of JCB38849S	Chismon *et al.* (2010)

Table 2.2 Plasmids used in this work

Plasmid	Description	Origin
pRW50	Low copy-number *lac* expression vector used for cloning EcoRI-HindIII fragments as transcriptional fusions to *lacZ*. Used in all β-Galactosidase assays in this study unless otherwise stated. Encodes tetracycline resistance.	Lodge *et al.* (1992)
pRW224	Derivative of pRW50 that allows cloning of fragments as translational fusions to *lacZ*.	Islam *et al.*(2010)
pSR	High copy-number plasmid for general cloning. Encodes ampicillin resistance.	Kolb *et al.* (1995)
pNF383	pAA182 *lac* expression plasmid containing the *E. coli hcp-hcr* regulatory region. Encodes ampicillin resistance.	Filenko *et al.* (2005)
pSTBlue-1	High copy-number general purpose cloning vector. Encodes ampicillin resistance.	Novagen
pGIT9	pSTBlue-1 containing *E. coli nsrR*.	Bodenmiller and Spiro (2006)
pJW15 Δ100	pAA121 high copy-number plasmid with an EcoRI-HindIII fragment containing the *melR* promoter and coding region, with an NsiI site engineered at codons 1 and 2 of *melR* and the first 100 codons of *melR* deleted. Encodes ampicillin resistance.	Kahramanoglou *et al.* (2006)
pLG339	Low copy-number cloning vector. Encodes kanamycin and tetracycline resistance.	Stoker *et al.* (1982)
pLG339 ΔRS	Derivative of pLG339 with a deletion of sequence between the EcoRI and SalI sites thereby disrupting the tetracycline resistance gene.	Belyaeva *et al.*(2000)
pDLC1	Derivative of pJW15 Δ100 carrying the NsiI-HindIII fragment NarL-CTD.	This work
pDLC2	Derivative of pJW15 Δ100 carrying the NsiI-HindIII fragment NarL-Full.	This work
pDLC3	Derivative of pLG339 carrying the BamHI-EcoRI fragment *be*NarL-CTD.	This work
pDLC4	Derivative of pLG339 carrying the BamHI-EcoRI fragment *be*NarL-Full.	This work
pDLC5	Derivative of pLG339 with a BamHI-EcoRI fragment carrying the *narXL* operon and regulatory region.	This work
pDLC6	Derivative of pJW15 Δ100 carrying the NsiI-HindIII fragment NarP-CTD	This work
pDLC8	Derivative of pJW15 Δ100 carrying the NsiI-HindIII fragment NarP-CTD138	This work
pDLC9	Derivative of pJW15 Δ100 carrying the NsiI-HindIII fragment NarP-CTD148	This work
pDLC14	Derivative of pLG339 with a BamHI-EcoRI fragment carrying the *narP* gene and regulatory region.	This work
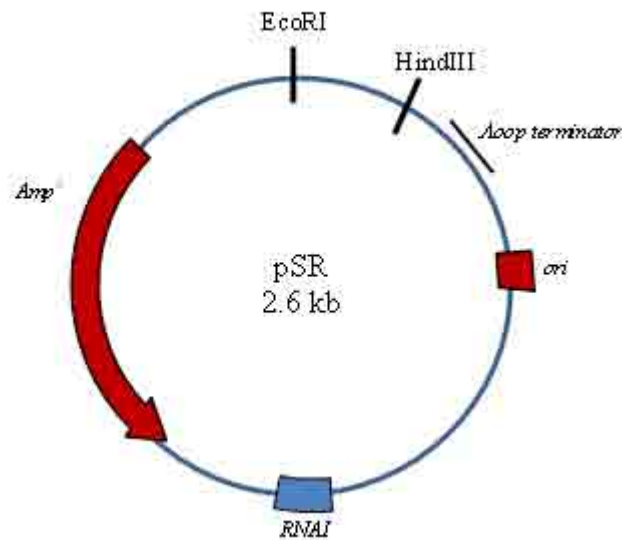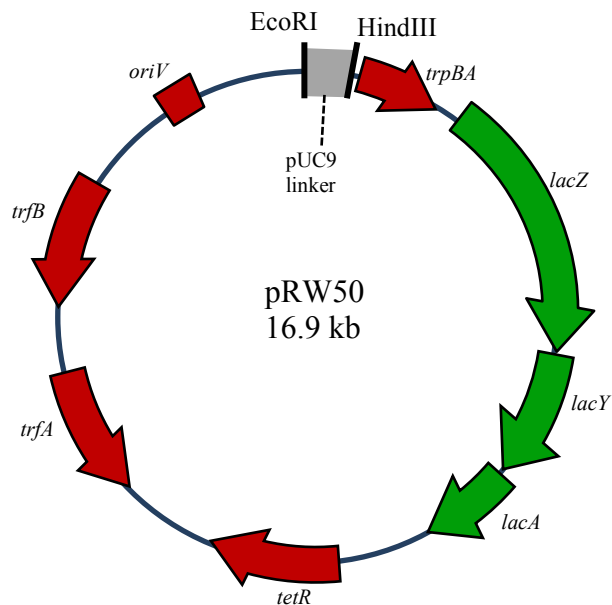
Figure 2.1 Maps of plasmids pSR and pRW50

Panel A shows a map of the general cloning plasmid pSR. EcoRI-HindIII fragments are cloned upstream of the λ*oop* terminator to ensure transcripts are of a defined length. Also shown are the origin of replication (*ori*), ampicillin resistance gene (*Amp^R*) and *RNA1* gene.

Panel B shows a map of the *lac* fusion vector pRW50. EcoRI-HindIII fragments are cloned into the pUC9 linker region (shown in grey) as transcriptional fusions to the *trpBA* and *lacZYA* operons. Also shown is the origin of replication (*oriV*), plasmid replication genes (*trfA* and *trfB*) and tetracycline resistance gene (*tet^R*)

A
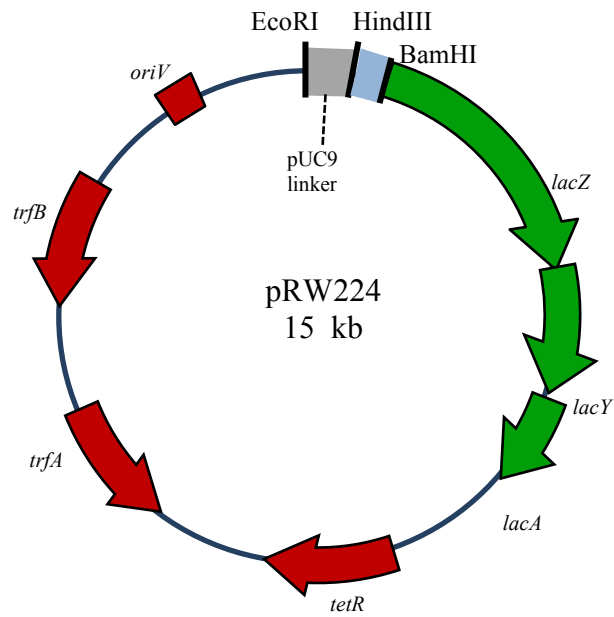
B

	EcoRI	pUC9 linker	HindIII

ATTCGCGAGAGCCTTGAGTCCACGCTAGATCTGAATTCCCGGGGATCCGTCGACCTGCAGCCAAGCTTAATGGAG

galE   BamHI

CGAATTATGAGAGTTCTGGTTACCGCCAAGCTCCGGGATCCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCT

Figure 2.2 Map and partial sequence of plasmid pRW224

Panel A shows a map of the pRW50 derivative, pRW224. The *trpBA* operon of pRW50 is deleted and the cloning region modified to allow fragments to be cloned upstream of the *lacZYA* operon as transcriptional or translational fusions. EcoRI-HindIII fragments utilise the *galE* Shine-Dalgarno sequence (shaded blue), whilst EcoRI-BamHI fragments are translational fusions.

Panel B shows a partial sequence of pRW224 which shows the region surrounding the EcoRI and   Bam
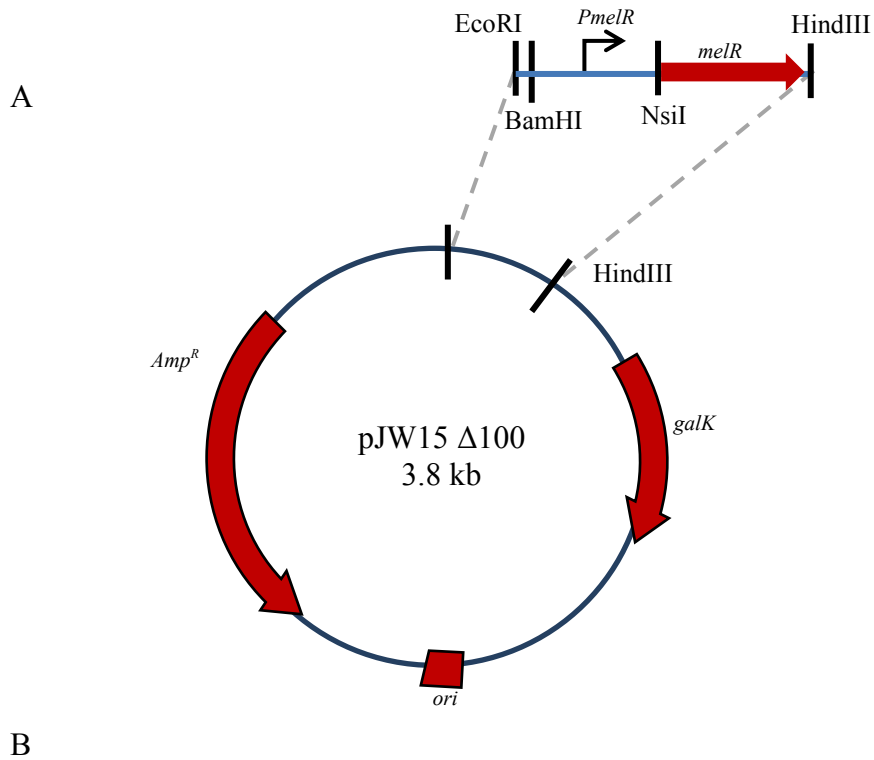
A

B

EcoRI   BamHI   CRP

GAATTCCCGGGGATCCCGGGAAGAGGGTGAAAACCGTGCTCCCACTCGCAGTCATCCTCCCTCACTCCTGCCATA

                                                                     -10

              +1   NsiI

ATTCTGATATTCCAGGAAAGAGAGCCATCCATGCATCTGTTTCTCTCCTGGCCGCTGGATAAAGACCTGATTAAC

                              SD   *melR* Δ1-99

Figure 2.3 Map and partial sequence of plasmid pJW15 Δ100

Panel A shows a map of plasmid pJW15 Δ100, a derivative of pJW15 with codons 1 through 99 of *melR* deleted. Map also shows the ampicillin resistance gene (*Amp*^{R}), the origin of replication (*ori*) and the *galK* gene.

Panel B shows a partial sequence of pJW15 Δ100. A binding site for CRP is shown in red and the *melR* -10 element, transcription start (+1) and Shine-Dalgarno (SD) sequence are underlined in bold and labelled. Alternate *melR* codons are underlined.
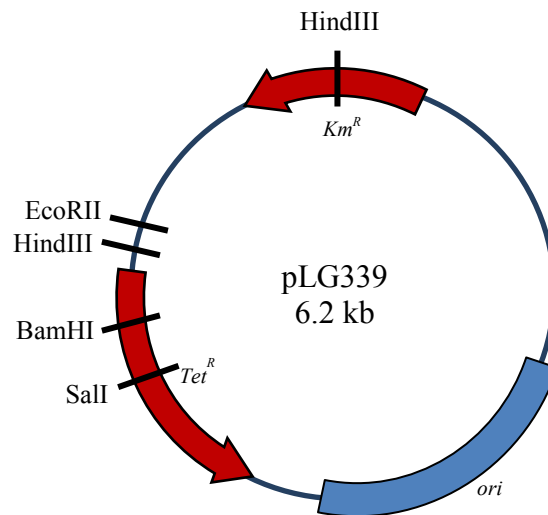
Figure 2.4 Map of pLG339

Figure shows a map of the low copy-number cloning plasmid pLG339. The origin of replication (*ori*) is shown in blue and the tetracycline (*Tet^R*) and kanamycin (*Km^R*) genes are shown in red.

A

EcoRIEcoRI   *PmelR*   *narL*   HindIII

BamHI   NsiI

EcoRI

HindIII

Amp^R

pDLC1
3.8 kb

galK

ori

B

EcoRI   BamHI   CRP

GAATTCCCGGGGATCCCGGGAAGAGGGTGAAAACCGTGCTCCCACTCGCAGTCATCCTCCCTCACTCCTGCCATA
-10

+1   NsiI

ATTCTGATATTCCAGGAAAGAGAGCCATCCATGCATGCCACTACTGAGCGCGATGTTAACCAGTTAACCCCACGC
SD   *narL* Δ1-145

GATATTCTCAAGCTGATTGCCCAGGGTTTGCCGAACAAGATGATTGCCCGCCGCCTGGATATCACCGAAAGCACA

GTAAAAGTGCACGTCAAGCACATGCTGAAGAAAATGAAGCTCAAGTCTCGCGTGGAAGCAGCGGTATGGGTGCAT

CAGGAGCGCATTTTCTGAAAGCTT
HindIII

Figure 2.5 Map and partial sequence of pDLC1

Panel A shows a map of pDLC1, a derivative of pJW15 Δ100. The *melR* coding region from pJW15 Δ100 is deleted and replaced with DNA encoding the CTD of NarL such that expression of the NarL CTD is driven by the *melR* promoter.

Panel B shows the sequence of the EcoRI-HindIII fragment carried by pDLC1. A binding site for CRP is shown in red and the *melR* -10 element, transcription start (+1) and Shine-Dalgarno (SD) sequence are underlined in bold and labelled. Alternate *narL* codons are underlined.

49

EcoRI   BamHI   CRP

GAATTCCCGGGGATCCCGGGAAGAGGGTGAAAACCGTGCTCCCACTCGCAGTCATCCTCCCTCACTCCTGCCATA

                                                                            -10

             +1   NsiI

ATTCTGATATTCCAGGAAAGAGAGCCATCCATGCATAATCAGGAACCGGCTACTATCCTGCTGATTGACGATCAC

                           SD   narL
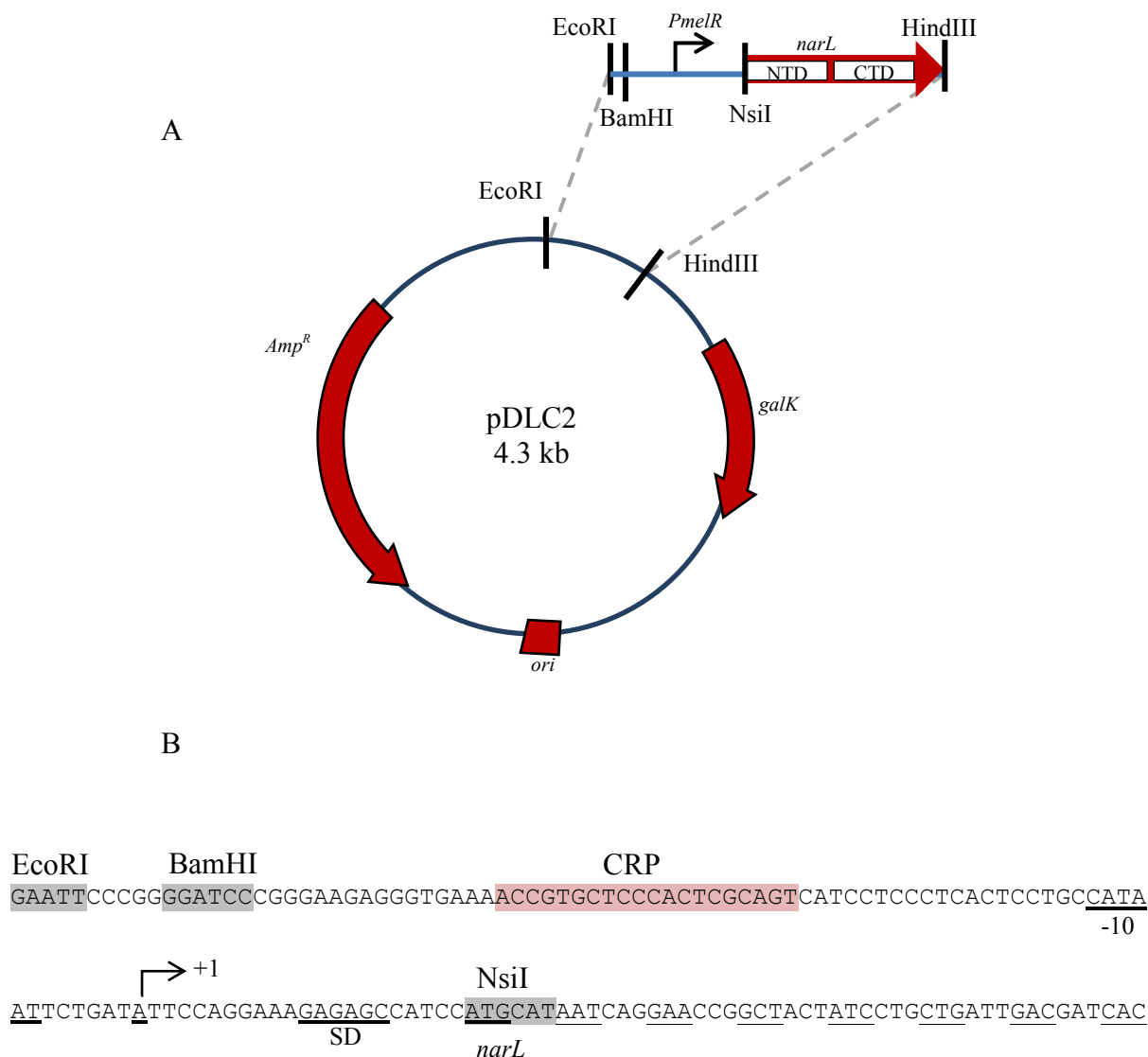
Figure 2.6 Map and partial sequence of pDLC2

Panel A shows a map of pDLC2, a derivative of pJW15 Δ100. The melR coding region from pJW15 Δ100 is deleted and replaced with the *narL* gene such that expression of *narL* is driven by the *melR* promoter.

Panel B shows a partial sequence of the EcoRI-HindIII fragment carried by pDLC2. A binding site for CRP is shown in red and the *melR* -10 element, transcription start (+1) and Shine-Dalgarno (SD) sequence are underlined in bold and labelled. Alternate *narL* codons are underlined.

A

PmelR   narL   EcoRI

CTD

BamHI   NsiI

HindIII

Km^R

pDLC3
6 kb

EcoRII

BamHI

SalI   Tet^R

ori

B

                                                                    +1

BamHI   CRP
GGATCCCGGGAAGAGGGTGAAAACCGTGCTCCCACTCGCAGTCATCCTCCCTCACTCCTGCCATAATTCTGATAT
                                                                    -10

                              NsiI
TCCAGGAAAGAGAGCCATCCATGCATGCCACTACTGAGCGCGATGTTAACCAGTTAACCCCACGCGAGCGCGATA
        SD   narL Δ1-145

TTCTCAAGCTGATTGCCCAGGGTTTGCCGAACAAGATGATTGCCCGCCGCCTGGATATCACCGAAAGCACAGTAA

AAGTGCACGTCAAGCACATGCTGAAGAAAATGAAGCTCAAGTCTCGCGTGGAAGCAGCGGTATGGGTGCATCAGG

                EcoRI
AGCGCATTTTCTGAGAATTC
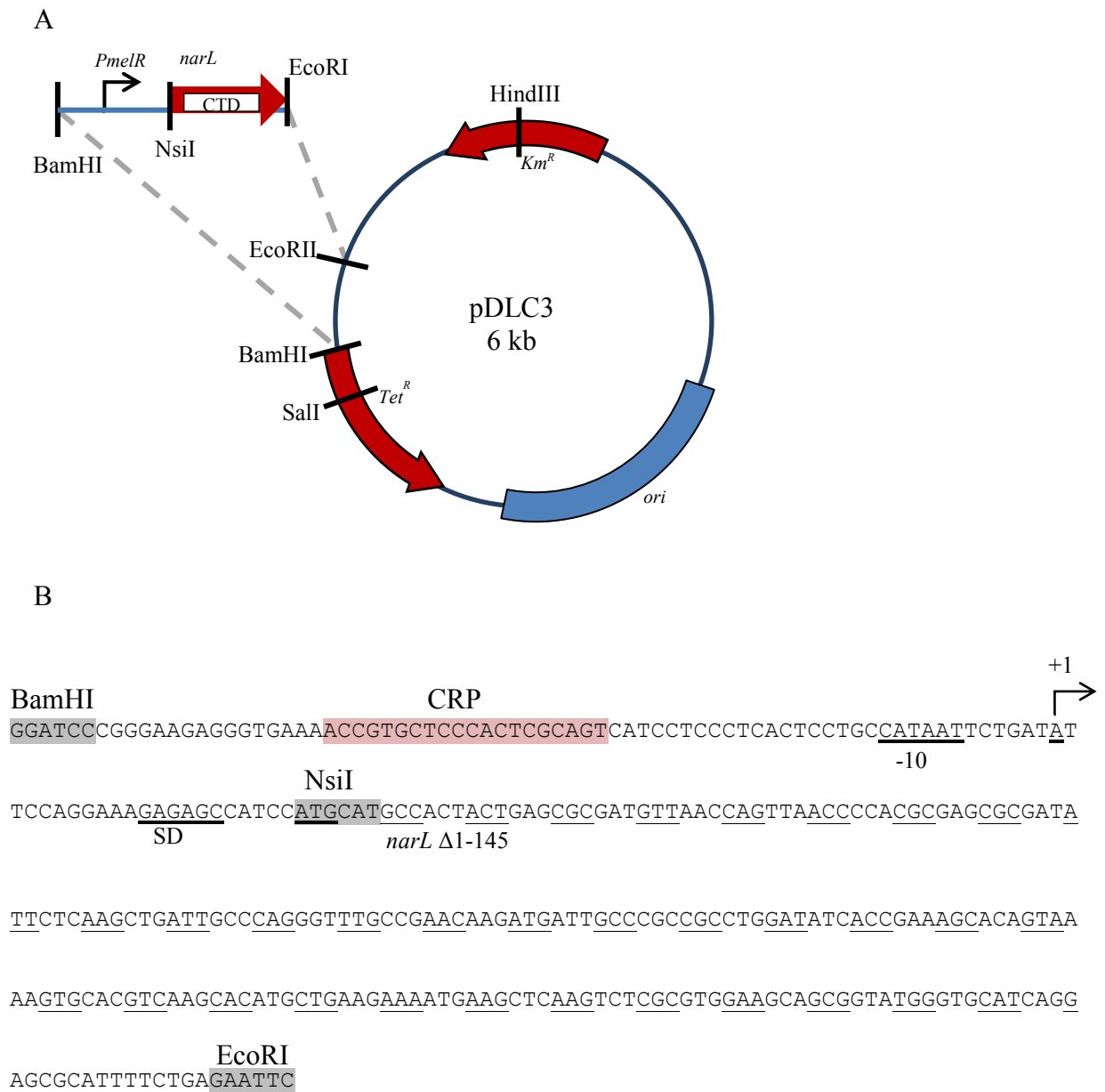
Figure 2.7 Map and partial sequence of pDLC3

Panel A shows a map of pDLC3, a pLG339 derivative. The region of the tetracycline resistance gene between the BamHI and EcoRI sites of pLG339 is replaced with the BamHI-HindIII fragment from pDLC1 that has been modified to replace the HindIII site with an EcoRI site.

Panel B shows the sequence of the BamHI-EcoRI fragment carried by pDLC3. Compare with Figure 2.5. A binding site for CRP is shown in red and the *melR* -10 element, transcription start (+1) and Shine-Dalgarno (SD) sequence are underlined in bold and labelled. Alternate *narL* codons are underlined.
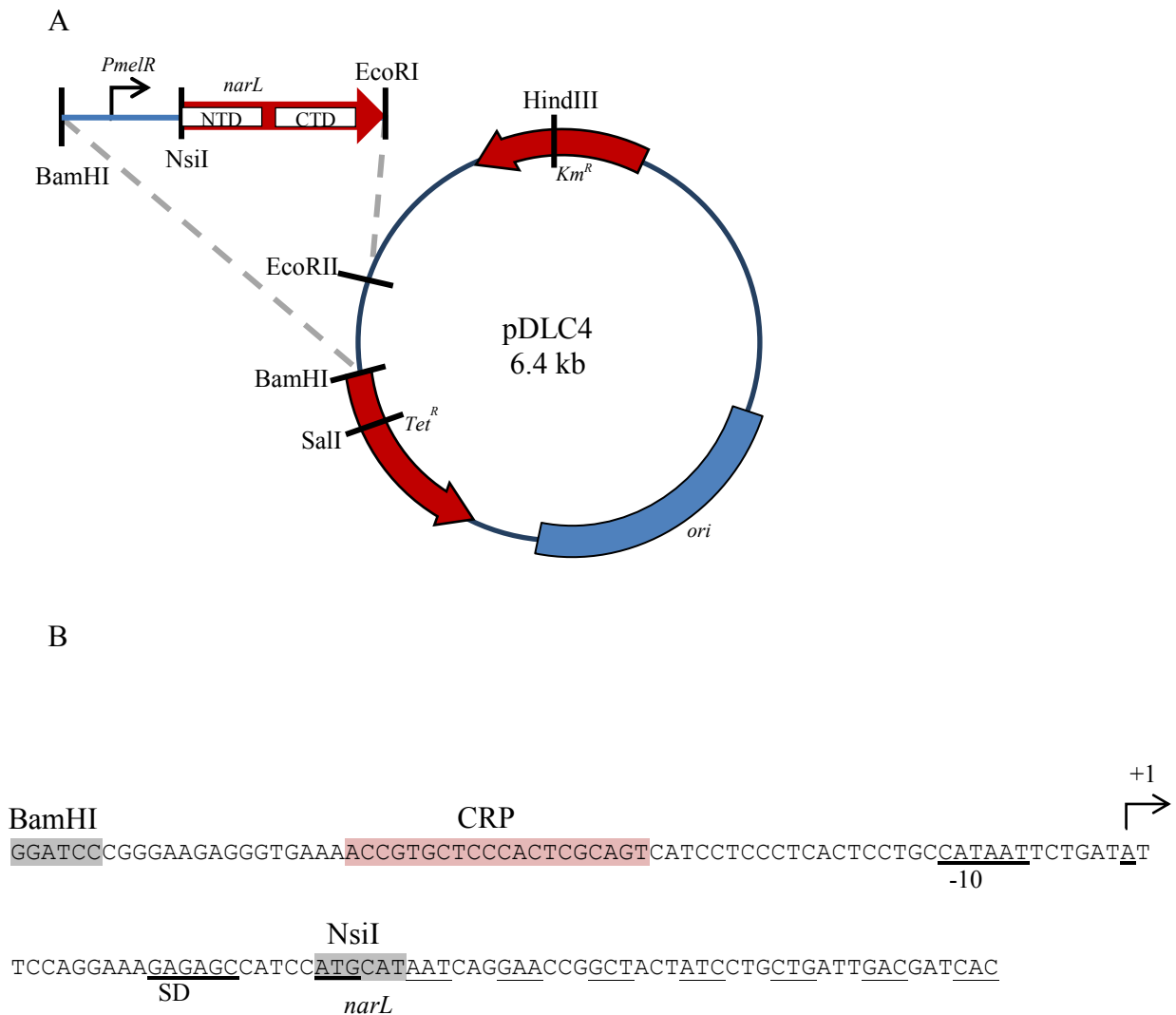
Figure 2.8 Map and partial sequence of pDLC4

Panel A shows a map of pDLC4, a pLG339 derivative. The region of the tetracycline resistance gene between the BamHI and EcoRI sites of pLG339 is replaced with the BamHI-HindIII fragment from pDLC2 that has been modified to replace the HindIII site with an EcoRI site.

Panel B shows the sequence of the BamHI-EcoRI fragment carried by pDLC4. Compare with Figure 2.6. A binding site for CRP is shown in red and the *melR* -10 element, transcription start (+1) and Shine-Dalgarno (SD) sequence are underlined in bold and labelled. Alternate *narL* codons are underlined.

A
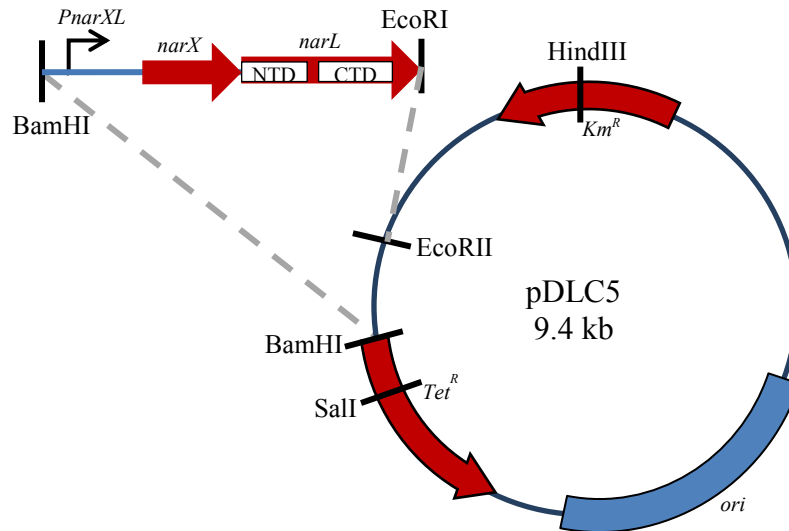
B

BamHI
GGATCCACCCATAGTGAGTACAGTGACTTCATAAAAATTATGAGATTTTTCACGGTGCTGTAAAATCCCTACCCT
                                        -35   -10

TACCGATGTAAAGCGACTAACCACACGGCAAATAAGGAGTAACTCTTTCCGGGTATGGGTATACTTCAGCCAATA

GCCGAGAATACTGCCATTCCAGAATGTATCGTCACATTCATTAAGGTTATTGCTCATTTAAAGCCTGAAGGAAGA
                                                                          SD

GGTTTACATGCTTAAACGTTGTCTCTCT
       narX

Figure 2.9 Map and partial sequence of pDLC5

Panel A shows a map of pDLC5, a pLG339 derivative. The region of the tetracycline resistance gene between the BamHI and EcoRI sites is replaced with the *narXL* operon and promoter region.

Panel B shows a partial sequence of the BamHI-EcoRI fragment carried by pDLC5. The *narXL* transcription start site and promoter elements as reported by Nohno *et al.* (1989) are underlined and labelled. Alternate codons of *narX* are underlined.
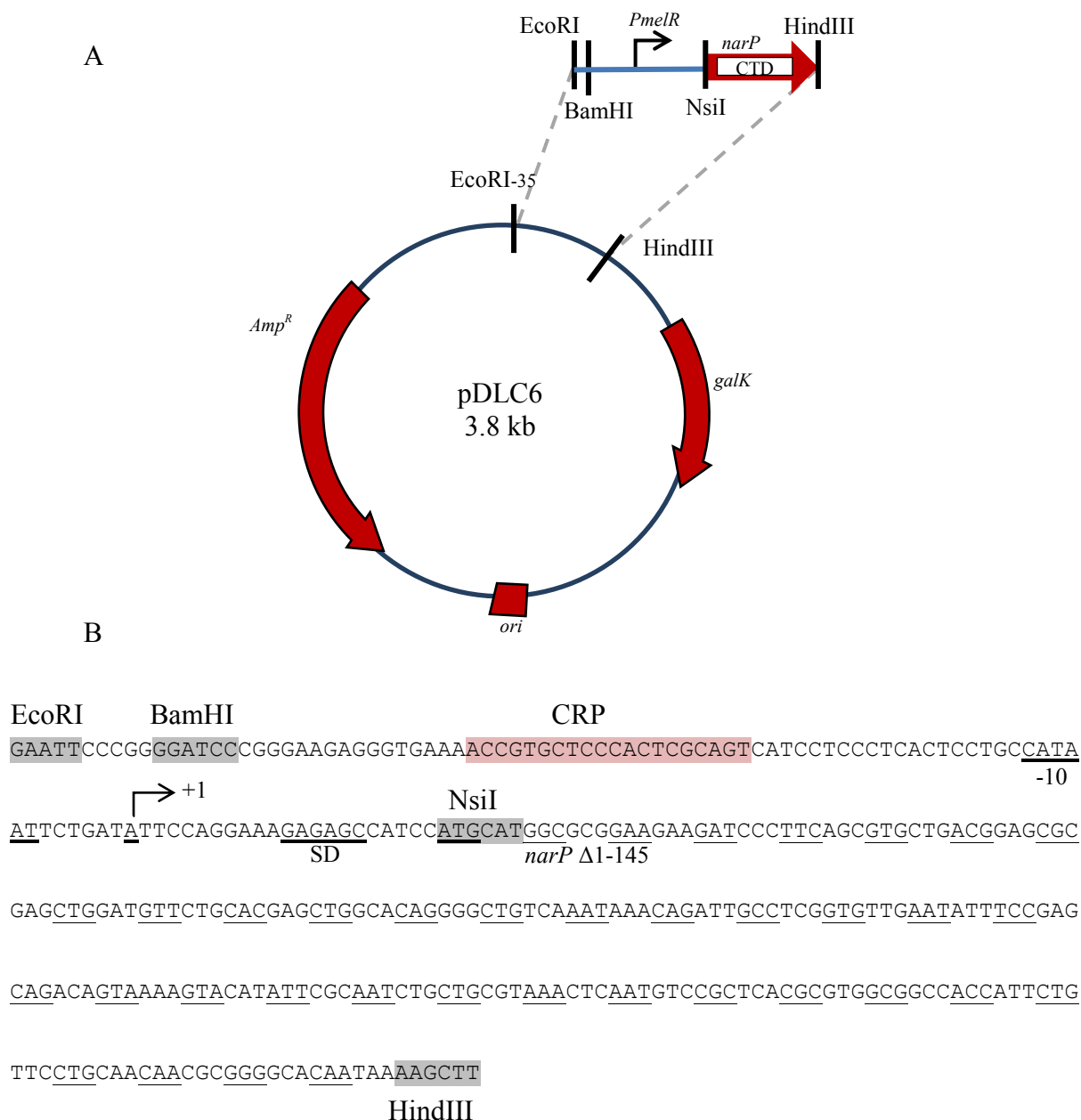
Figure 2.10 Map and partial sequence of pDLC6

Panel A shows a map of pDLC6, a pDLC1 derivative. The NsiI-HindIII fragment of pDLC1 is replaced with DNA encoding the CTD of NarP such that expression of the NarP CTD is driven by the *melR* promoter.

Panel B shows the sequence of the EcoRI-HindIII fragment carried by pDLC6. A binding site for CRP is shown in red and the *melR* -10 element, transcription start (+1) and Shine-Dalgarno (SD) sequence are underlined in bold and labelled. Alternate *narP* codons are underlined. The plasmid encodes NarP from amino acid 146.
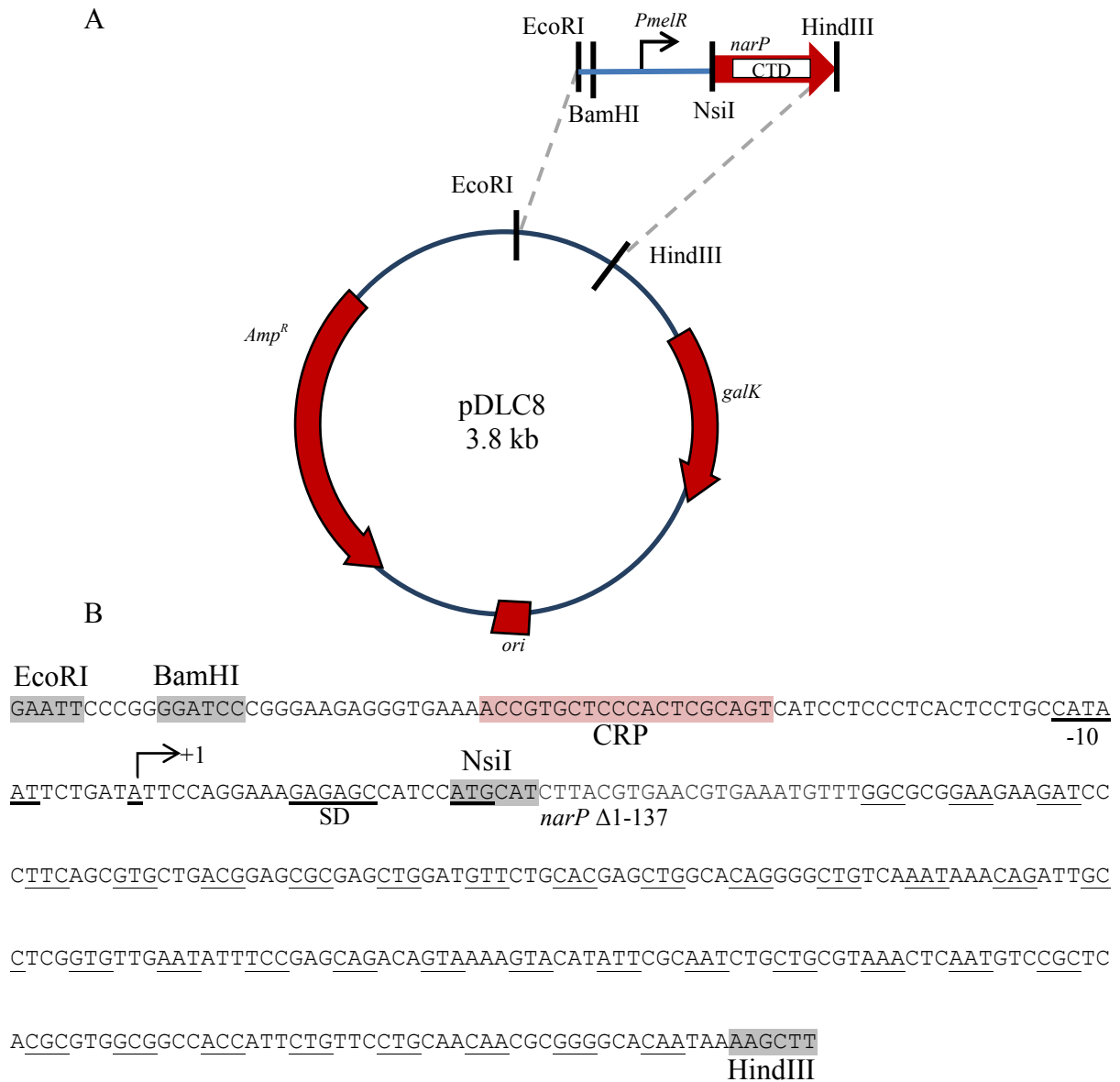
Figure 2.11 Map and partial sequence of pDLC8

Panel A shows a map of pDLC8, a pDLC1 derivative. The NsiI-HindIII fragment of pDLC1 is replaced with DNA encoding the CTD of NarP such that expression of the NarP CTD is driven by the *melR* promoter.

Panel B shows the sequence of the EcoRI-HindIII fragment carried by pDLC8. A binding site for CRP is shown in red and the *melR* -10 element, transcription start (+1) and Shine-Dalgarno (SD) sequence are underlined in bold and labelled. Alternate *narP* codons are underlined. The plasmid encodes NarP from amino acid 138.
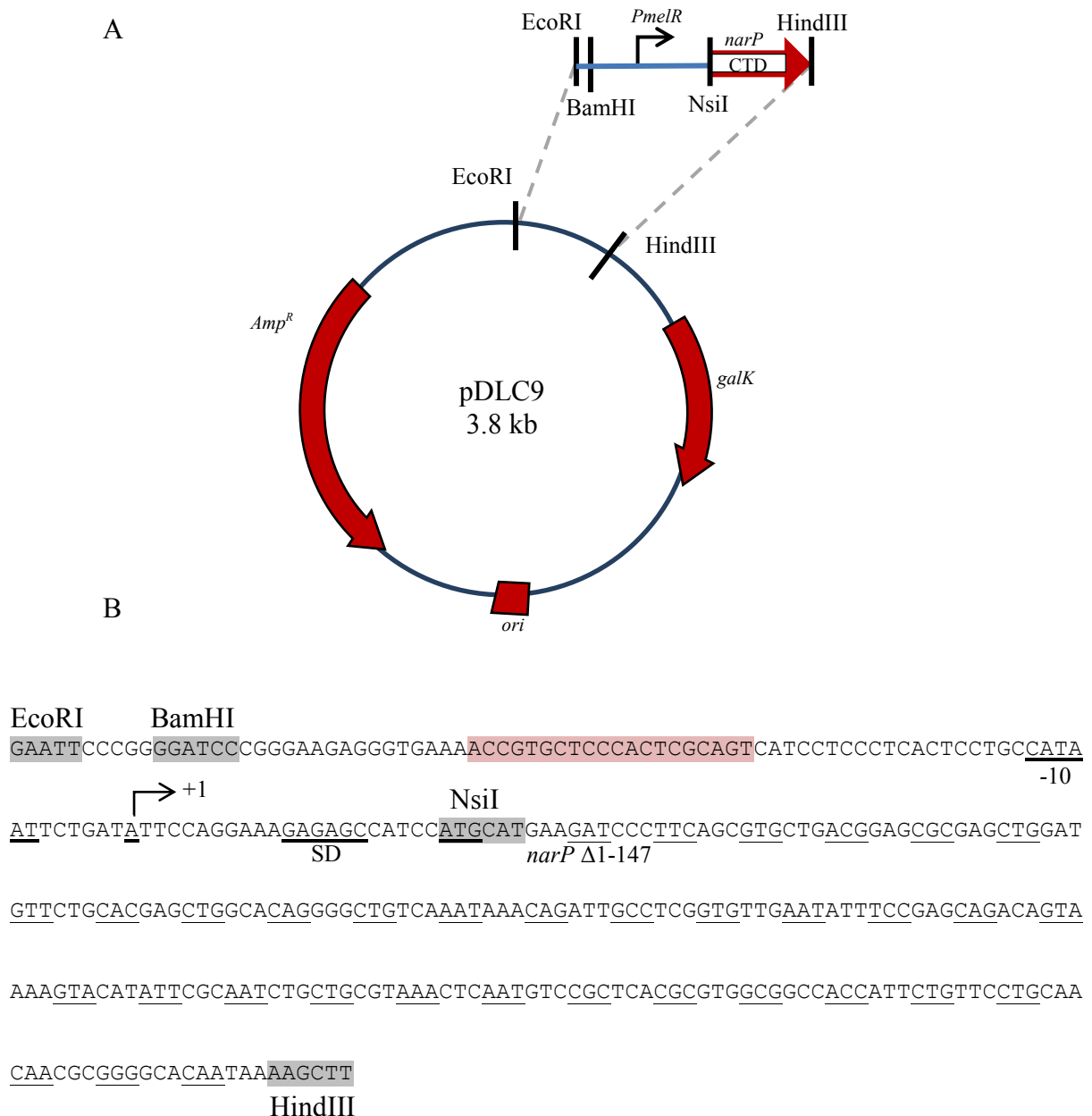
Figure 2.12 Map and partial sequence of pDLC9

Panel A shows a map of pDLC9, a pDLC1 derivative. The NsiI-HindIII fragment of pDLC1 is replaced with DNA encoding the CTD of NarP such that expression of the NarP CTD is driven by the *melR* promoter.

Panel B shows the sequence of the EcoRI-HindIII fragment carried by pDLC8. A binding site for CRP is shown in red and the *melR* -10 element, transcription start (+1) and Shine-Dalgarno (SD) sequence are underlined in bold and labelled. Alternate *narP* codons are underlined. The plasmid encodes NarP from amino acid 148.

A

B

BamHI
GGATCCTGATGCATTGAGCTTTCATCCTATGAAATTAATTGCTGTTAAAAGCATTGGGTACAGAAAATACCCATA

GCTCCATACCCGGAGTCAGTTTTTAAAAACTGTTTAAAGAAATGCACAAGTATTGTGATTGATTTTTTAGTTGTT

TTTCTTGATGAGAAGCTGATGCAAAATTCCGTCTTTATAATGAAAATGATGCCAAAGCGAACGACAAGGTTGTAG

TTTTCACTACATGTCCATACATAAAATGGG**G**TAACATTCACGCGCCTGGTAGCGTTACCAACGCTACGCTCAAAC

ATAATGATTCTAATAAAACCTCAGGAGACTACTATGCCTGAAGCA
                      SD   narP
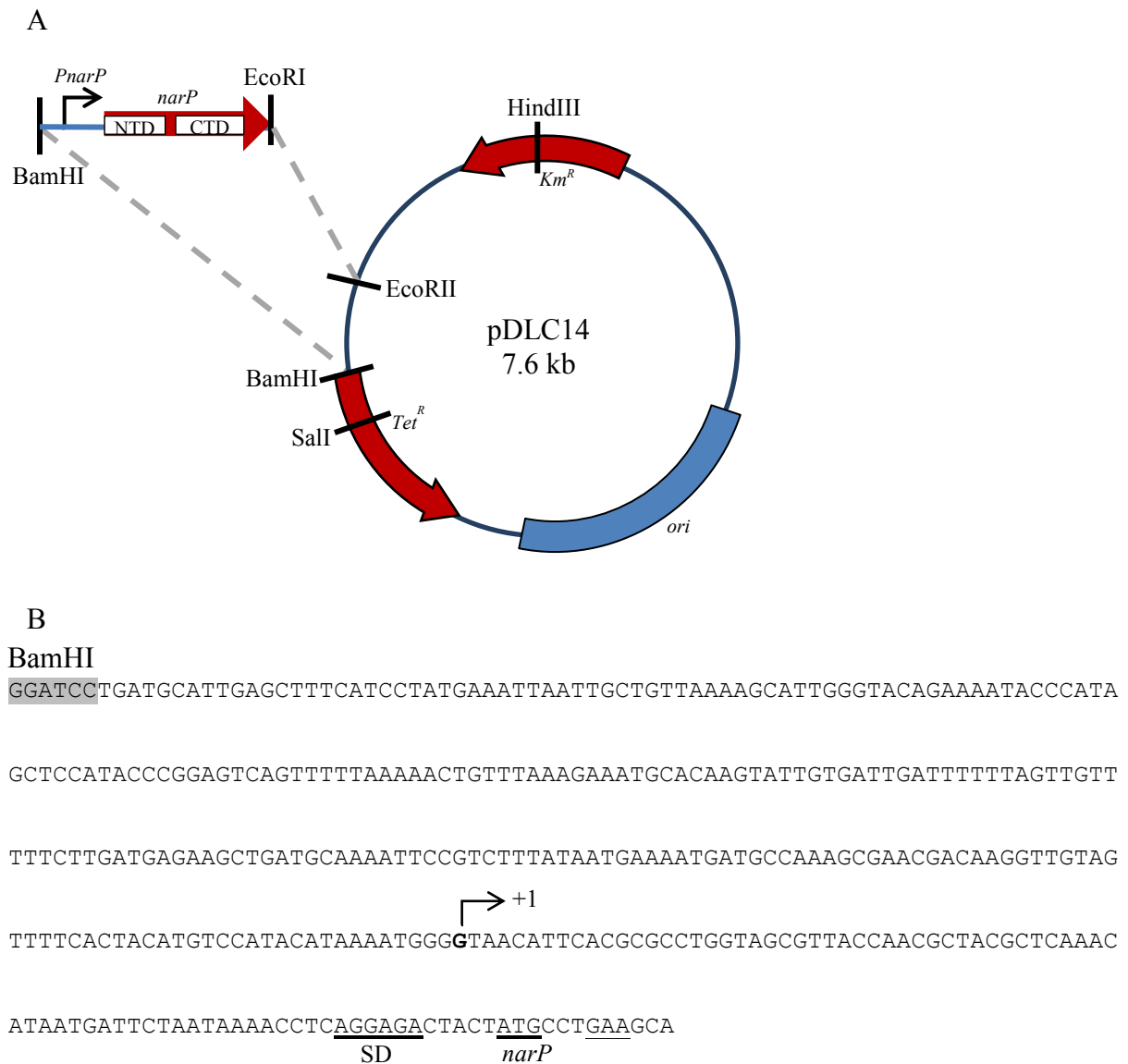
Figure 2.13 Map and partial sequence of pDLC14

Panel A shows a map of pDLC14, a pLG339 derivative. The region of the tetracycline resistance gene between the BamHI and EcoRI sites is replaced with the *narP* gene and promoter region.

Panel B shows a partial sequence of the BamHI-EcoRI fragment carried by pDLC14. The *narP* σ^{32}-dependent transcription start site as reported by Nonaka *et al.* (2006) is underlined and marked with a bent arrow. The Shine-Dalgarno sequence has been predicted on the basis of similarity to the consensus sequence for such an element. Alternate codons of *narP* are underlined.

(w/v) ammonium persulphate and 0.001 volumes TEMED (N,N,N′,N′-tetramethylethylenediamine). DNA was mixed with orange loading buffer (Fermentas) prior to loading and gels were run in 1 X TBE at 36 mA for 40 minutes. Gels were stained in 1 X TBE containing 1 X Sybrsafe (Invitrogen) and visualised using a UV trans-illuminator or a visible spectrum trans-illuminator for fragments that were to be excised.

2.6.3. *Sequencing polyacrylamide gel electrophoresis*

For analysis of DNase I protection assays, samples were separated on a large (40cm x 30 cm x 0.4 mm) 6% acrylamide sequencing gel (SequaGel). The gel was prepared according to the manufacturer's instructions and run in 1 X TBE buffer at 60 W. The gel was pre-run for 1 hour prior to loading and DNA samples were mixed with loading buffer and incubated at 90°C for 2 minutes before loading. Gels were run for up to 2 hours and then fixed in fixing solution. Gels were transferred to filter paper and vacuum dried for 30 minutes at 80°C. Gels were analysed by exposure to a Phosphor screen (Fuji) for up to 12 hours, which was then scanned using a Molecular Imager FX (BioRad). The supplied QuantityOne software was used to analyse the images produced.

2.7. Extraction and purification of nucleic acids

2.7.1. *Phenol-chloroform purification of DNA*

To remove contaminating proteins, solutions containing DNA were mixed with an equal volume of phenol-chloroform-isoamyl alcohol (25/24/1, pH 8.0) and mixed vigorously. Solutions were centrifuged for 2 minutes at 13,000 rpm to separate the sample into aqueous and organic layers. The upper aqueous layer was removed to a separate tube by pipette and 200 μl of sterile distilled water was added to the tube containing the lower, organic, layer. The tube containing the organic layer was then vigorously mixed and then centrifuged for 2 minutes at 13,000 rpm. The aqueous layer was then carefully removed by pipette and

combined with the aqueous layer that had been extracted previously. DNA was then ethanol precipitated as described below.

2.7.2. *Extraction of DNA by ethanol precipitation*

Ethanol precipitation was used to concentrate DNA solutions and to extract DNA from buffers that did not contain proteins. Samples were gently mixed with 10% sodium acetate (v/v) and 1% 1M $MgCl_2$ (v/v). Two volumes of ice-cold 100% ethanol were added to the samples, which were then incubated at -80°C for 15 minutes. DNA was sedimented by centrifugation at 4°C and 14,000 rpm for 15 minutes. The supernatant was removed and the pellet was washed with 1 ml of ice-cold 70% ethanol. The DNA was re-sedimented by centrifugation at 4°C and 14,000 rpm for 10 minutes. The supernatant was discarded and the pellet was dried under vacuum for 10 minutes. The DNA pellet was then re-suspended in an appropriate amount of sterile distilled water.

2.7.3. *Purification of DNA from solution by QiaQuick kit*

DNA was purified following PCR or digestion using a QiaQuick PCR purification kit (QIAgen). The manufacturer's instructions were followed and DNA was eluted into 50 µl buffer EB (provided in kit).

2.7.4. *Extraction and purification of DNA from agarose gels*

A scalpel was used to isolate the region of an agarose gel that contained a DNA fragment of interest. The DNA was then purified from the gel fragment using a QiaQuick Gel Extraction kit (QIAgen) according to the manufacturer's instructions.

2.7.5. *Electroelution from polyacrylamide gels*

A scalpel was used to isolate the region of a polyacrylamide gel that contained a DNA fragment of interest. A 5-10 cm length of dialysis tubing, prepared using the method of Maniatis *et al.* (1982), was clipped at one end and filled with 200 µl of 0.1 X TBE. The gel fragment was inserted into the dialysis tubing and the tubing sealed. The tube was placed in

an electroelution tank and covered with 0.1 X TBE and run at 26 mA for 30 minutes. The tubing was opened and the buffer extracted by pipette to a microcentrifuge tube. DNA was purified and extracted by phenol-chloroform treatment and ethanol precipitation (see sections 2.7.1 and 2.7.2).

2.7.6. *Small scale preparation of plasmid DNA (miniprep)*

An overnight culture of a strain containing the plasmid of interest was grown in medium containing appropriate antibiotics. Plasmid DNA was then extracted using the QIAprep Spin Miniprep kit (QIAgen) according to the manufacturer's instructions. For high copy-number plasmids, 1.5 ml of culture was processed and for low copy-number plasmids, 5 ml of culture was processed. Following purification, DNA was eluted into 50 µl buffer EB (provided in kit).

2.7.7. *Large scale preparation of plasmid DNA (maxiprep)*

The caesium chloride density gradient ultracentrifugation (maxiprep) method described in Maniatis *et al.* (1982) was used for the preparation of large amounts of highly pure plasmid DNA. An overnight culture of a strain containing the plasmid of interest was grown in medium supplemented with appropriate antibiotics. A 2500 ml culture vessel containing 500 ml growth medium (LB supplemented with appropriate antibiotics) was then inoculated with all of the overnight culture. The culture was grown at 37°C with shaking until an OD_{650} of ~0.9. The medium was then supplemented with 50 mg of chloramphenicol to retard protein synthesis but allow DNA replication to continue. The culture was incubated overnight at 37°C with shaking. The culture was sedimented in a centrifuge at 4°C and at 4,000 rpm for 30 minutes. The supernatant was discarded and the cell pellet re-suspended in 12 ml of ice-cold TES buffer. The solution was transferred to a round bottomed centrifuge tube (Oakridge style). The tube was supplemented with 2 ml cold 250 mM EDTA and 3 ml cold TES buffer containing 10 mg/ml lysozyme and was mixed and incubated for 15 minutes on ice. 2 ml of

10% SDS was added to the tube, which was then mixed and incubated on ice for a further 10 minutes. 5 ml of ice-cold 5 M NaCl was added to the tube, which was then mixed and incubated on ice overnight.

The cell debris was separated from the DNA by centrifugation at 4°C at 20,000 rpm for 2 hours. The supernatant was removed to a fresh 40 ml centrifuge tube and warmed to room temperature. DNA was precipitated by adding 15 ml of isopropanol and incubating at room temperature for 15 minutes. DNA was then sedimented by centrifugation at 20°C at 15,000 rpm for no more than 15 minutes. The supernatant was discarded and the DNA pellet was air-dried. The pellet was re-suspended in 5 ml TES buffer and 5.16 g (+/- 10 mg) of CsCl and 200 µl of ethidium bromide were added to the solution. The solution was transferred to a 10 ml quick seal tube (Beckman) and filled with TES buffer before being sealed. The sample was then centrifuged at 20°C and at 50,000 rpm for 22 hours.

Following ultracentrifugation, the tube was secured in a clamp and a hole pierced in the top of the tube with a heated needle to allow air into the tube. A second heated needle was then used to carefully extract the lower of the 2 ethidium bromide stained bands that were visible. To remove the ethidium bromide, the sample was purified with water-saturated butanol 5 times: The sample was mixed with an equal volume of water-saturated butanol and then centrifuged for 4 minutes at 13,000 rpm and the upper layer discarded. The DNA solution was then transferred to dialysis tubing and dialysed overnight in 4 l of 1X TE.

Following dialysis the DNA was purified with phenol chloroform and extracted by ethanol precipitation before being re-suspended in 600 µl 1X TE buffer and quantified on a ND1000 spectrophotometer (Thermo)

2.8. Bacterial transformations

2.8.1. *Preparation of competent cells by rubidium chloride treatment*

Rubidium chloride was used to prepare cells for transformation with plasmid DNA. A 5 ml culture of the desired strain was grown overnight and 1 ml used to inoculate 100 ml of 2X LB in a 250 ml conical flask. The flask was then incubated at 37°C with shaking until an OD_{650} of ~0.5 was reached. The flask was then placed on ice for 10 minutes and cells sedimented in a centrifuge at 4°C and at 4,000 rpm for 9 minutes. The supernatant was removed and the cell pellet re-suspended in 30 ml ice-cold TFB1 (see section 2.2.3). The cell suspension was incubated on ice for a minimum of 90 minutes and then the cells were sedimented in a centrifuge at 4°C and at 4,000 rpm for 9 minutes. The supernatant was discarded and the cell pellet re-suspended in 4 ml TFB2. The suspension was then aliquoted into micro-centrifuge tubes for future use. Aliquots were snap frozen in liquid nitrogen and stored at -80°C.

2.8.2. *Small scale preparation of competent cells*

For small-scale preparation of competent cells, TransformAid (Fermentas) was used according to the manufacturer's instructions. However, for maximum efficiency, DNA was transformed using the method below rather than the method detailed by the manufacturer.

2.8.3. *Transformation of competent cells with plasmid DNA*

An aliquot of competent cells to be transformed was thawed on ice for ~10 minutes. Cells were mixed gently with a pipette tip. A 0.5 ml micro-centrifuge tube was prepared with 1 µl of miniprep plasmid DNA and a 50 µl aliquot of competent cells, briefly mixed by pipetting, and then incubated on ice for 30 minutes. Cells were heat shocked at 42°C for 40 seconds and then incubated for a further 5 minutes on ice. SOC broth was pre-warmed to 37°C and 250 µl added to the cells. Cells were then incubated at 37°C with shaking for 1 hour. Agar plates were pre-warmed to 37°C and the entirety of the transformation mixture plated. Plates were incubated at 37°C overnight.

2.9. DNA manipulations

2.9.1. *Polymerase Chain Reaction (PCR)*

Standard PCR

Phusion polymerase (NEB) with the supplied HF buffer was used for standard PCR and 10 µl of miniprep DNA was used as template. Where plasmid DNA was used as the template, it was diluted with sterile distilled water to either 1:100 (for low copy-number plasmids) or 1:500 (for high copy-number) plasmids. dNTPs (Bioline) were provided as a mixture to a final concentration of 0.25 mM for each dNTP. Oligonucleotide primers were added to a final concentration of 1 µM (each primer). PCR samples were made up to 50 µl final volume with sterile distilled water. An example of a typical cycle is shown in Table 2.3.

Colony PCR

Colony PCR was used to check for the presence of a plasmid or a particular fragment on a plasmid. A tube was prepared with 25 µl Biomix Red (Bioline), 1 µM of each primer and then made up to 50 µl with sterile distilled water. A 200 µl pipette tip was used to pick a colony from an agar plate and the PCR mixture agitated with the tip. A modified PCR cycle was used to amplify fragments (Table 2.3).

2.9.2. *Restriction digestion of DNA*

DNA was digested according to the manufacturer's instructions using either NEB enzymes or Fermentas FastDigest enzymes. Enzyme volume used was between 1 µl and not more than 10% of the final reaction volume. Restriction digests were incubated at 37°C for between 30 minutes and 12 hours. FastDigest enzymes were used for double digests because they use a common buffer. Plasmid DNA that had been digested in preparation for cloning was prevented from re-ligating by treatment with alkaline phosphatase. Either calf-intestinal phosphatase (NEB) or FastAP (Fermentas) was used depending on the restriction enzymes. DNA fragments were then separated by gel electrophoresis or, in some cases, purified using a Qiaquick spin kit without prior electrophoresis (see Section 2.7.3).

Table 2.3 Standard and colony PCR cycles

The extension time (Tx) was calculated as 30 seconds per kb of DNA to be amplified. If product was less than 1 kb long, 30 s was used.

Temperature	Time		Purpose
Standard PCR:			
98°C	30 s		Initial denaturation
98°C	10 s		Denaturation
60°C	20 s	X 30	Primer annealing
72°C	Tx	cycles	Extension
72°C	7 min		Final extension
Colony PCR:			
95°C	10 min		Cell lysis / Initial denaturation
95°C	10 s		Denaturation
60°C	20 s	X 30	Primer annealing
71°C	Tx	cycles	Extension
71°C	5 min		Final extension

2.9.3. DNA ligations

DNA fragments with compatible restriction sites were ligated using T4 DNA ligase (NEB) according to the manufacturer's instructions. Between 1 µl and 15 µl of digested insert DNA and between 3 µl and 6 µl of digested plasmid DNA was used. 2 µl of supplied buffer and 1 µl of T4 ligase was used and reactions were made up to 20 µl with sterile distilled water. Ligation reactions were placed on ice for 10 minutes and then incubated at room temperature for 10 minutes. Between 1 µl and 5 µl of the ligation reactions were then transformed into cells using the method described in section 2.8.3.

2.9.4. DNA sequencing

DNA was sequenced by the Functional Genomics and Proteomics Laboratory at the University of Birmingham using an ABI3730 sequencer. The relevant region of a plasmid was amplified by standard PCR and the PCR product purified using a kit (see Section 2.7.3). 6.8 µl of purified PCR product was mixed with 3.2 µl of 10 µM oligonucleotide primer stock. Sequencing primers are listed in Table 2.4. Cloning of promoter fragments into pRW50, pRW224 and pSR

2.10. Cloning of promoter fragments into pRW50, pRW224 and pSR

2.10.1. Preparation of EcoRI-HindIII promoter fragments by PCR

Table 2.5 lists all EcoRI-HindIII promoter fragments used in this study and Figures 2.14-2.16 show the sequence of each fragment. EcoRI-HindIII fragments were generated by PCR using a plasmid that contained the fragment of interest as template DNA. In some cases, the fragment had the correct restriction sites, whilst in other cases it was necessary to design primers that introduced the desired sites. The PCR product was cut with EcoRI and HindIII enzymes, purified and ligated into prepared pRW50, pRW224 or pSR. Primers used for amplification of fragments are given in Table 2.6

Table 2.4 Sequencing / screening primers used in this work

Code	Sequence	Use
D10520	CCCTGCGGTGCCCCTCAAG	Anneals upstream of EcoRI site in pRW50 and derivatives
D10527	GCAGGTCGTTGAACTGAGCCTG AAATTCAGG	Anneals downstream of HindIII site in pRW50 and derivatives
D5431	ACCTGACGTCTAAGAAACC	Anneals upstream of EcoRI site in pSR and pJW15 Δ100
D4600	GTAGTCGGTGTGTTCAC	Anneals downstream of HindIII site in pSR
D55668	GAGCGGCGACGATAGTCATG	Anneals upstream of BamHI site in pLG339 and derivatives
D70296	ACCTGACGTCTAAGAAACC	Anneals upstream of EcoRI site in pJW15 Δ100 and derivatives
D70297	GTAGTCGGTGTGTTCAC	Anneals downstream of EcoRI site in pJW15 Δ100 and derivatives

Table 2.5 EcoRI-HindIII promoter fragments used in this work

A P80T mutation indicates that the nucleotide at position -80 relative to the transcription start has been mutated to a T.

Name	Description	Source
hcp383-0	Carries the intergenic region between *ybjE* and *hcp* which includes the *hcp* regulatory region. Carries nucleotide sequence from -285 to +98	Filenko *et al.* (2005)
hcp383-1	Derivative of hcp383-0 that carries nucleotide sequence from -125 to +98	This study
hcp383-2	Derivative of hcp383-0 that carries nucleotide sequence from -92 to +98	This study
hcp383-3	Derivative of hcp383-0 that carries nucleotide sequence from -54 to +98	This study
hcp383-1n	Fragment hcp383-1 that carries an A to G mutation and a T to G mutation at positions +9 and +10	This study
hcp383-101	Derivative of the hcp383-1 fragment that carries P110G and P99C mutations	This Study
hcp383-102	Derivative of the hcp383-1 fragment that carries P78C and P67G mutations	This Study
hcp383-1ns	*hcp-hcr* promoter fragment that carries nucleotide sequence from -125 to +46 and contains carries an A to G mutation and a T to G mutation at positions +9 and +10	This study
hcp383-1nvs	Derivative of hcp383-1ns that carries nucleotide sequence from -125 to +8	This study
hcp383-2n	Derivative of hcp383-1n that carries nucleotide sequence from -92 to +98	This study
hcp383-3n	Derivative of hcp383-1n that carries nucleotide sequence from -54 to +98	This study
hcp383-3s	Derivative of hcp383-3 that carries nucleotide sequence from -54 to +46	This study
hcp383-3ns	Derivative of hcp383-1n that carries nucleotide sequence from -54 to +46	This study
ogt100	*E. coli ogt* promoter fragment carrying nucleotide sequences from -270 to +51	Squire *et al.* (2009)
ogt101	Derivative of the ogt100 fragment carrying P80T, P77A, P75G and P73G mutations	M. Xu
ogt102	Derivative of the ogt100 fragment carrying P84G and P63C mutations	M. Xu
ogt104	Derivative of the ogt100 fragment carrying P51G and P40C mutations	D. Browning
ogt1041	Derivative of the ogt100 fragment carrying P80T, P77A, P75G and P73G, P51G and P40C mutations	D. Browning
ogt1042	Derivative of the ogt100 fragment carrying P84G, P63C, P51G and P40C mutations	D. Browning
ogt105	Derivative of the ogt100 fragment carrying P51C, P48C, P44A and P42A mutations	D. Browning
ogt1051	Derivative of the ogt100 fragment carrying P80T, P77A, P75G and P73G, P51C, P48C, P44A and P42A mutations	D. Browning
ogt1052	Derivative of the ogt100 fragment carrying P84G,P63C, P51C, P48C, P44A and P42A mutations	D. Browning

yeaR100	*E. coli yeaR* promoter fragment carrying nucleotide sequences from −294 to +96.	D. Squire
ogt106	Derivative of the ogt100 fragment with positions -23 and -24 deleted	This Study
ogt107	Derivative of the ogt100 fragment carrying P51C, P49C, P48A, P47A, P44A, and P43A mutations	This Study

Figure 2.14 Sequences of *hcp* promoter fragments used in this work

EcoRI and HindIII sites are shaded grey and labelled. The stop codon of *ybjE* and the start codon of *hcp* are shown in bold. The *hcp* transcription start (as identified by Filenko *et al* 2005) is stongly underlined and indicated by a bent arrow labelled +1, The -10 and -35 elements predicted by this study are strongly underlined and labelled. The DNA site for NarL proposed by Filenko *et al.* (2005) is shaded orange and labelled. The DNA site for FNR proposed by Filenko *et al.* (2005) is shaded blue and labelled. The DNA site for NsrR proposed by Filenko *et al.* (2007) is shaded pink and labelled.

Panel A shows the sequence of the hcp383-0 fragment. A more detailed diagram of this fragment is shown in Figure 3.1.

Panel B shows the sequence of the hcp383-1 fragment and indicates mutations introduced to construct derivatives. The hcp383-101 fragment carries mutations that disrupt the DNA site for NarL, the hcp383-102 fragment carries mutations that disrupt the DNA site for FNR and the hcp383-1n fragment carries mutations that disrupt the DNA site for NsrR.

Panel C shows the sequence of the hcp383-3 fragment and indicates mutations introduced to construct the derivative hcp383-2n, which carries mutations that disrupt the DNA site for NsrR.

Panel D shows the sequence of the hcp383-3 fragment and indicates the mutations introduced to construct the derivative hcp383-3n, which carries mutations that disrupt the DNA site for NsrR.

Panel E shows the sequence of the hcp383-1ns fragment

Panel F shows the sequence of the hcp383-1nvs fragment

Panel G shows the sequence of the hcp383-3s fragment and indicates the mutations introduced to construct the derivative hcp383-3ns, which carries mutations that disrupt the DNA site for NsrR.

A: hcp383-0

EcoRI -250 -200

GAATTCTGATTCGCCGCAGCCGCTCTACTGCACTGGGCTTATGCGGTGCCACATCAATGGATTTCACCCTGCCCGTTCTTCAACGTACTGGCGGGCTGGATATGGTCCCGGCGGCAATTGT

-150 -100 -50

TCACGGTTTTATTCTTAGCCTGTTAGTGCCGATCCTCATCGCCTTTTTCTCTGCG**TAA**TACCTCTCTGGCGGGTAGATCCCTGCCGCCAAAATTGCGCTAAATCAATCTCCCTTAAAGTTGC

NarL (-104.5) FNR (-72.5)

+1 +40

ATGAAAAATCCCTTTTATCCCCGCGTTAAGCGTCTTAACCTTAAACATGTATATTAAATATAACTTTAAAAGGTGTGATC**ATG**TTTTGTGTGCAATGTGAACAAACTATCCGTACTCCGGC

-35 -10 NsrR (+6)

+80 HindIII

AGGAAACGGCTGCTCATACGCGCAGGAAGCTT

B: hcp383-1 / hcp383-101 / hcp383-102 / hcp383-1n

G hcp383-101 C C hcp383-102 G

EcoRI -50

GAATTCTTTCTCTGCG**TAA**TACCTCTCTGGCGGGTAGATCCCTGCCGCCAAAATTGCGCTAAATCAATCTCCCTTAAAGTTGC

NarL (-104.5) FNR (-72.5)

+1 GG hcp383-1n +40

ATGAAAAATCCCTTTTATCCCCGCGTTAAGCGTCTTAACCTTAAACATGTATATTAAATATAACTTTAAAAGGTGTGATC**ATG**TTTTGTGTGCAATGTGAACAAACTATCCGTACTCCGGC

-35 -10 NsrR (+6)

+80 HindIII

AGGAAACGGCTGCTCATACGCGCAGGAAGCTT

C: hcp383-2 / hcp383-2n

EcoRI FNR (-72.5) -50

GAATTCCCTGCCGCCAAAATTGCGCTAAATCAATCTCCCTTAAAGTTGC

+1 GG hcp383-2n +40

ATGAAAAATCCCTTTTATCCCCGCGTTAAGCGTCTTAACCTTAAACATGTATATTAAATATAACTTTAAAAGGTGTGATC**ATG**TTTTGTGTGCAATGTGAACAAACTATCCGTACTCCGGC

-35 -10 NsrR (+6)

+80 HindIII

AGGAAACGGCTGCTCATACGCGCAGGAAGCTT

70

D: hcp383-3 / hcp383-3n

EcoRI
`GAATTC`GTTGC

+1   GG hcp383-3n   +40

`ATGAAAAATCCCTTTTTATCCCCGCGTTAAGCGTCTTAACCTTAAACATGTATATTAAATATAACTTTAAAAGGTGTGATCATGTTTTGTGTGCAATGTGAACAAACTATCCGTACTCCGGC`

-35   -10   NsrR (+6)

+80   HindIII

`AGGAAACGGCTGCTCATACGCGCAGGAAGCTT`

E: hcp383-1ns

EcoRI   -100   -50

`GAATTCTTTCTCTGCGTAATACCTCTCTGGCGGTAGATCCCTGCCGCCAAAATTGCGCTAAATCAATCTCCCTTAAAGTTGC`

NarL (-104.5)   FNR (-72.5)

+1   +40   HindIII

`ATGAAAAATCCCTTTTTATCCCCGCGTTAAGCGTCTTAACCTTAAACATGTATATTAAGGATAACTTTAAAAGGTGTGATCATGTTTTGTGTGCAAGCTT`

-35   -10   NsrR (+6)

F: hcp383-1nvs

EcoRI   -100   -50

`GAATTCTTTCTCTGCGTAATACCTCTCTGGCGGTAGATCCCTGCCGCCAAAATTGCGCTAAATCAATCTCCCTTAAAGTTGC`

NarL (-104.5)   FNR (-72.5)

+1   HindIII

`ATGAAAAATCCCTTTTTATCCCCGCGTTAAGCGTCTTAACCTTAAACATGTATATTAAAAGCTT`

-35   -10   NsrR (+6)

G: hcp383-3s / hcp383-3ns

EcoRI   -50

`GAATTC`GTTGC

+1   GG hcp383-3ns   +40   HindIII

`ATGAAAAATCCCTTTTTATCCCCGCGTTAAGCGTCTTAACCTTAAACATGTATATTAAATATAACTTTAAAAGGTGTGATCATGTTTTGTGTGCAAGCTT`

-35   -10   NsrR (+6)

71

Figure 2.15 *ogt* promoter fragments used in this work

EcoRI and HindIII sites are shaded grey and labelled. The start codon of *ogt* is shown in bold. The *ogt* transcription start (as identified by Potter *et al* 1987) is strongly underlined and indicated by a bent arrow labelled +1, The -10 elements identified by Potter *et al.* (1987) is strongly underlined and labelled. The DNA sites for NarL identified by Squire *et al.* (2010) are shaded orange and labelled.
Panel A shows the sequence of the ogt100 fragment.
Panel B shows a partial sequence of the *ogt* promoter fragments used in this work. The *ogt* promoter fragments used do not differ in sequence from the ogt100 fragment outside of the partial sequence shown. Positions shaded in green indicate mutations that improve the DNA site for NarL, relative to the consensus sequence for such sites, TACCYMTnnAKRRGTA. Positions shaded in red indicate mutations that worsen the DNA site for NarL relative to the consensus sequence for such sites. The ogt101 fragment carries mutations that improve only the NarL I site. The ogt102 fragment carries mutations that disrupt only the NarL I site. The ogt104 fragment carries mutations that disrupt only the NarL II site. The ogt1041 fragment carries mutations that improve the NarL I site but disrupt the NarL II site. The ogt1042 fragment carries mutations that disrupt both NarL I and NarL II sites. The ogt105 fragment carries mutations that improve only the NarL II site. The ogt1051 fragment carries mutations that improve both the NarL I and NarL II sites. The ogt1052 fragment carries mutations that disrupt the NarL I site and improve the NarL II site. The ogt106 fragment carries a deletion of 2 bases (positions -23 and -24 of ogt100). The NarL II site of the ogt107 fragment is mutated to the sequence of the DNA site for NarL from the *yeaR* regulatory region.

A: ogt100

EcoRI -250
GAATTCCAATCTGGTCGATTCTCGCCCCCATTTTCGTACCAATGTTTAT

-200 -150
GCTACTTGGCTTTCACCCGGCATTTGCGCAAATCCTCTTTCGTATTGCCGACTCATCCGTATTGCCTTTAGCGCCGGTATCTCCTTTTGTTCCACTGTTTCTTGGATTCCTGCAACGCTAC

-100 -50 +1
AAACCAGACGCGAAACTGGGTACTTACTATTCGTTAGTCTTGCCCTATCCACTTATCTTTTTGGTGGTATGGCTGCTGATGTTGCTGGCGTGGTATCTTGTCGGTCTGCCGATAGGTCCGG
NarL I (-78.5) NarL II (-45.5) -10

+40 HindIII
GTATTTACCCACGTTTGTCTTAAGAGAGAACGGATGAAGCTT

B: Derivatives of ogt100 used in this study

NarL I (-78.5) NarL II (-45.5) +1
TACYYMTnnAKRRGTA -60 TACYYMTnnAKRRGTA -30

ogt100 TGGGTACTTACTATTCGTTAGTCTTGCCCTATCCACTTATCTTTTTGGTGGTATGGCTGCTGATGTTGCTGGCGTGGTATCTTGTCGGTCTGCCGATAGGTCCGG
-10

ogt101 TGGGTACTTATTAATGGGTAGTCTTGCCCTATCCACTTATCTTTTTGGTGGTATGGCTGCTGATGTTGCTGGCGTGGTATCTTGTCGGTCTGCCGATAGGTCCGG

ogt102 TGGGTAGTTACTATTCGCTAGTCTTGCCCTATCCACTTATCTTTTTGGTGGTATGGCTGCTGATGTTGCTGGCGTGGTATCTTGTCGGTCTGCCGATAGGTCCGG

ogt104 TGGGTACTTACTATTCGTTAGTCTTGCCCTATCCACTTAGCTTTTTGGTGCTATGGCTGCTGATGTTGCTGGCGTGGTATCTTGTCGGTCTGCCGATAGGTCCGG

ogt1041 TGGGTACTTATTAATGGGTAGTCTTGCCCTATCCACTTAGCTTTTTGGTGCTATGGCTGCTGATGTTGCTGGCGTGGTATCTTGTCGGTCTGCCGATAGGTCCGG

ogt1042 TGGGTAGTTACTATTCGCTAGTCTTGCCCTATCCACTTAGCTTTTTGGTGCTATGGCTGCTGATGTTGCTGGCGTGGTATCTTGTCGGTCTGCCGATAGGTCCGG

ogt105 TGGGTACTTACTATTCGTTAGTCTTGCCCTATCCACTTACCTCTTTAGAGGTATGGCTGCTGATGTTGCTGGCGTGGTATCTTGTCGGTCTGCCGATAGGTCCGG

ogt1051 TGGGTACTTATTAATGGGTAGTCTTGCCCTATCCACTTACCTCTTTAGAGGTATGGCTGCTGATGTTGCTGGCGTGGTATCTTGTCGGTCTGCCGATAGGTCCGG

ogt1052 TGGGTAGTTACTATTCGCTAGTCTTGCCCTATCCACTTACCTCTTTAGAGGTATGGCTGCTGATGTTGCTGGCGTGGTATCTTGTCGGTCTGCCGATAGGTCCGG

ogt106 ACTGGGTACTTACTATTCGTTAGTCTTGCCCTATCCACTTATCTTTTTGGTGGTATGGCTGCTGATGTCTGGCGTGGTATCTTGTCGGTCTGCCGATAGGTCCGG

ogt107 TGGGTACTTACTATTCGTTAGTCTTGCCCTATCCACTTAACCAATAAATGGTATGGCTGCTGATGTTGCTGGCGTGGTATCTTGTCGGTCTGCCGATAGGTCCGG

EcoRI   -250   -200

GAATTCGCGACGCTGGAACTGGTGAGTTTCTGCTATTTGAGCTTCCTGATTATATCTGGTGCTTTTGTCACGCAGTACATACGTACCAAAAGAAACTGGCTAAAGT

                         -150   -100

TGGCAACTCACTGATTGGTTTGATGTTCGTGGGTTTCGCTGCCCGACTGGCGACGCTGCAATCCTGATGCTTTCAGCCCGCGTTGTCGCGGGCTTCCCATCTATAATCCTCCCTGATTCTT

              -50   +1   +40

CGCTGATATGGTGCTAAAAAGTAACCAATAAATGGTATTTAAAATGCAAATTATCAGGCGTACCCTGAAACGGCTGGAATAAACCGTTTTCAGCGCATTCACCGAAGGAGGGAAAAGG**ATG**

                      NarL (-43.5)   -10

                       +80   HindIII

CTTCAAATCCCACAGAATTATATTCATACGCGCTCAACGCCTTTCTGAAGCTT

Figure 2.16 Sequence of the yeaR100 promoter fragment

EcoRI and HindIII sites are shaded grey and labelled. The start codon of *yeaR* is shown in bold. The *yeaR* transcription start (as reported by Squire *et al* 2009) is strongly underlined and indicated by a bent arrow labelled +1, The -10 elements identified by Squire *et al.* (2009) is strongly underlined and labelled. The DNA site for NarL identified by Squire *et al.* (2010) is shaded orange and labelled.

Table 2.6 Primers used to amplify or generate EcoRI-HindIII promoter fragments

Code	Name	Sequence	Use
D10520		CCCTGCGGTGCCCCTCAAG	Anneals upstream of EcoRI site in pRW50
D10527		GCAGGTCGTTGAACTGAGCCTGAAATTCAGG	Anneals downstream of HindIII site in pRW50
D57415	HCP383 YBJE DEL	CCCGAATTCTTTCTCTGCGTAATACCTCTCTGGCGGT	Used with D10527 to construct hcp383-1, carries EcoRI site
D57416	HCP383 NAR DEL	CCCGAATTCCTGCCGCCAAAATTGCGCTAAATC	Used with D10527 to construct hcp383-2, carries EcoRI site
D57417	HCP383 FNR DEL	CCCGAATTCGTTGCATGAAAAATCCCTTTTATCC	Used with D10527 to construct hcp383-3 and hcp383-3ns, carries EcoRI site
D59726	PHCP333 DS	CCCAAGCTTGCACACAAAACATGATCACACCTT	Used with D10520 to construct hcp383-1ns and hcp383-3s, carries HindIII site

2.10.2. Cloning into pRW50, pRW224 or pSR

Vector DNA was prepared by digestion with EcoRI-HindIII restriction enzymes. Following digestion, a sample of the digestion mix was examined on an agarose gel to ensure complete digestion. The vector was treated with alkaline phosphatase and purified by phenol-chloroform treatment and ethanol precipitation. EcoRI-HindIII products that had been generated by PCR and then cut with EcoRI-HindIII restriction enzymes were purified on an agarose gel and ligated into the prepared vector. Plasmids were then transformed into competent JCB387 cells. Cells were grown overnight on agar plates supplemented with the appropriate antibiotic (tetracycline for pRW50 and pRW224, ampicillin for pSR) and a minimum of 3 colonies then used to inoculate separate overnight cultures. Plasmids were recovered from overnight cultures and the region encompassing the EcoRI-HindIII fragment was amplified by PCR, the product of which was sequenced.

2.11. Site-directed mutagenesis by megaprimer PCR

Due to the size of pRW50, it was not possible to use site-directed mutagenesis techniques that rely on inverse PCR. Therefore, megaprimer PCR (Perrin and Gilliland, 1990) (see Figure 2.17) was used for all site-directed mutagenesis in this project. In the first round of PCR, a mutagenic primer and a flanking primer were used to generate a megaprimer. The megaprimer was purified by electroelution from a 7.5% polyacrylamide gel and then used in a second round of PCR with the opposing flanking primer. The PCR product was analysed on an agarose gel, purified, and then amplified in a final round of PCR with both flanking primers. The fragment was then digested with EcoRI and HindIII restriction enzymes and cloned into pRW50. Mutagenic primers used are given in Table 2.7 and fragments containing point mutations are included in Table 2.5

2.12. Random mutagenesis by error prone PCR

Random mutations were introduced into DNA fragments by PCR in suboptimal
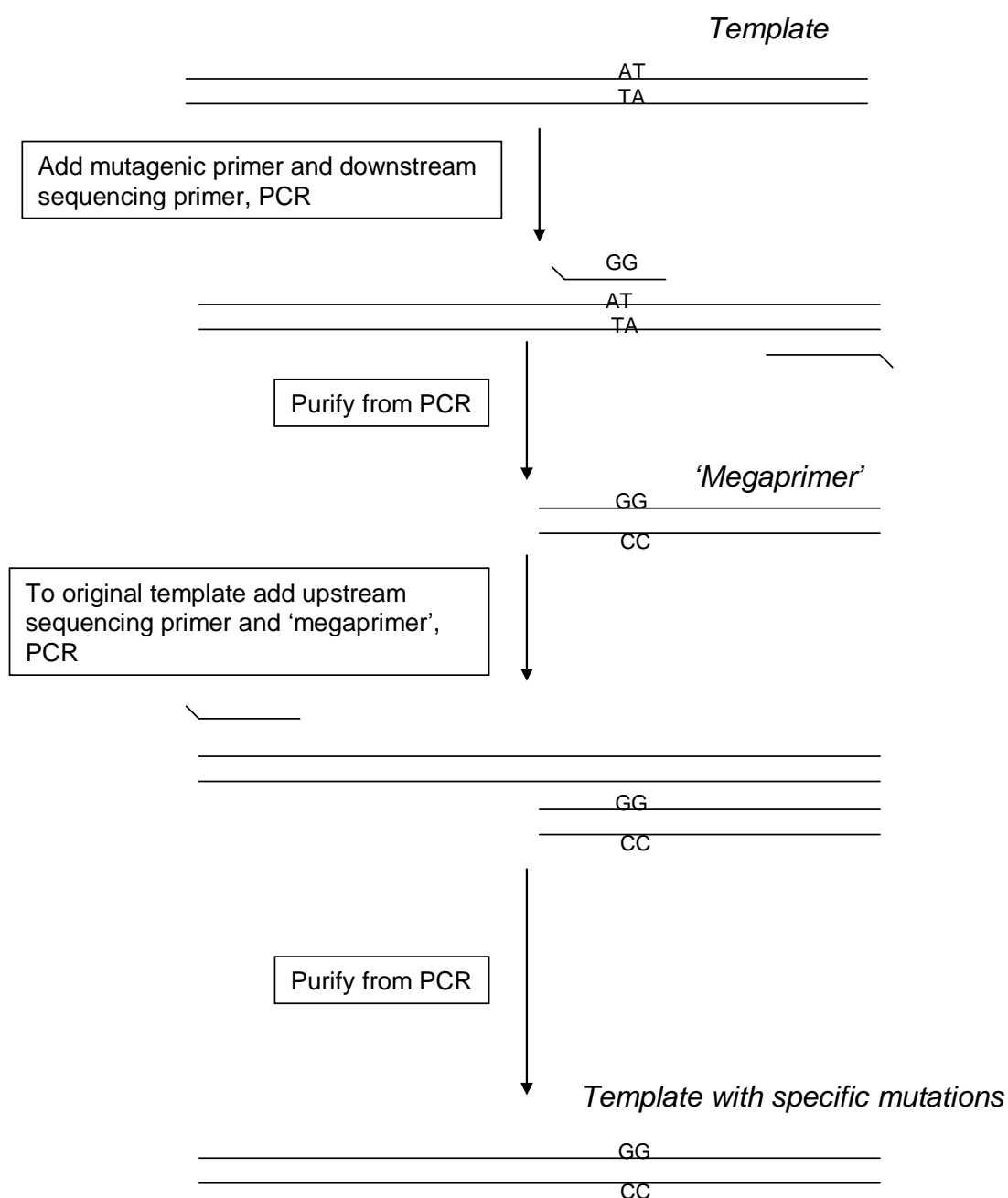
Figure 2.17 Overview of megaprimer PCR

A megaprimer carrying the required point mutation(s) was generated by PCR using plasmid pRW50 carrying the fragment of interest as a template. The megprimer was purified and used in a second round of PCR with the original template and a flanking primer. The product was purified and then amplified in a third round of PCR (not shown).

Table 2.7 Mutagenic primers used in this study

Code	Name	Sequence	Use
D56637	HCP NSRR MUTAG1	CCTTAAACATGTATATT AAGGATAACTTTAAAA GG	Used in PCR with D10527 to produce megaprimer for construction of hcp383-1n
D69558	HCP-L	GCGTAATAGCTCTCTGG CGCTAGATCCCT	Used in PCR with D10527 to produce megaprimer for construction of hcp383-101
D69559	HCP-F	CCAAAATCGCGCTAAA TCGATCTCCCTTAAAGT	Used in PCR with D10527 to produce megaprimer for construction of hcp383-102
D63522	HCP P35V	GAAAAATCCCTTRTATC CCCGCGTTAAG	Used in PCR with D10527 to produce megaprimer to mutate position -35 of the *hcp* promoter
D63521	HCP P37V	GAAAAATCCCVTTTATC CCCGCGTTAAG	Used in PCR with D10527 to produce megaprimer to mutate position -37 of the *hcp* promoter
D61976	P13G OLIGO	GCGTTAAGCGTCTTGAC CTTAAACATGTATATTA A	Used in PCR with D10527 to produce megaprimer to mutate position -13 of of the *hcp* promoter
D63520	HCP P14V	GCGTTAAGCGTCTVAA CCTTAAACATGTATATT AA	Used in PCR with D10527 to produce megaprimer to mutate position -14 of of the *hcp* promoter
D63519	HCP P9V	GCGTTAAGCGTCTTAAC CVTAAACATGTATATTA A	Used in PCR with D10527 to produce megaprimer to mutate position -9 of of the *hcp* promoter
D66082	OGT NARLII MOVE	TGGCTGCTGATGTCTGG CGTGGTATCTT	Used in PCR with D10527 to produce megaprimer for construction of ogt106
D66081	OGT-YEAR TRANSPLANT	CCATATCCACTTAACCA ATAAATGGTATGGCTG CTGATGTT	Used in PCR with D10527 to produce megaprimer for construction of ogt107

conditions. Taq polymerase is not a proof-reading enzyme and reaction conditions can be altered to decrease fidelity. Mutagenic reactions were prepared with 5 µl of each primer (10 pM / µl stock), 10 µl template (plasmid miniprep diluted 1:250), 2 µl dNTP stock (containing 5 mM each dNTP), 5 µl 10 X Taq buffer, 5 µl of 50 nM $MgCl_2$ and 17 µl sterile distilled water and 1 µl BioTaq polymerase (Bioline). Reactions using the same cycling protocol as standard PCR (see Table 2.3). For error prone PCR, the primers flanked the region to be mutagenised (D10520 and D10527) and the products were cut with EcoRI and HindIII and cloned into pRW50 upstream of *lacZ*. Plasmids were then transformed into JCB387, which were plated onto MacConkey-lactose to allow visual identification of colonies containing plasmids with impaired promoters.

2.13. Construction of pDLC plasmids

2.13.1. Construction of pJW15 Δ100 derived plasmids pDLC1, pDLC2, pDLC6, pDLC8 and pDLC9

To express either NarL or NarP from a plasmid, the coding sequence was cloned into pJW15 Δ100 (see Figure 2.3) downstream of the *melR* promoter such that expression of *narL* or *narP* is driven by the *mel*R promoter. Vector was prepared by digesting pJW15 Δ100 with NsiI and HindIII restriction enzymes to remove the *melR* coding region. The vector was treated with alkaline phosphatase and then purified by phenol-chloroform treatment and ethanol precipitation. DNA fragments encoding the NarL or NarP CTDs or full length proteins (see Table 2.8) were amplified from the chromosome of JCB387 using primers given in Table 2.9. To create an NsiI site in the fragment that contained full-length *narL*, the second codon of *narL* was mutated to CAT (See Figure 2.6, Panel B). For fragments encoding only the CTD, the NsiI site was simply appended upstream of the coding region. PCR products were then digested with NsiI and HindIII restriction enzymes and purified by phenol-chloroform treatment and ethanol precipitation. NsiI-HindIII fragments were ligated into prepared pJW15 Δ100 vector and transformed into competent JCB3884 cells.

Table 2.8 Fragments carrying *narL* or *narP*

Name	Description	Source
NarL-CTD	NsiI–HindIII fragment encoding the NarL CTD from amino acid 147	This study
NarL-Full	NsiI-HindIII fragment encoding NarL	This study
NarXL	BamHI-EcoRI fragment containing the *narXL* operon and 226 bp upstream of the translation start	This study
NarP-CTD146	NsiI–HindIII fragment encoding the NarP CTD from amino acid 146	This study
NarP-CTD138	NsiI–HindIII fragment encoding the NarP CTD from amino acid 138	This study
NarP-CTD148	NsiI–HindIII fragment encoding the NarP CTD from amino acid 148	This study
NarPop	BamHI-EcoRI fragment containing the *narp* gene and 327 bp upstream of the translation start	This study

Table 2.9 Primers used to amplify *narL* or *narP*

Code	Name	Sequence	Use
D5431		ACCTGACGTCTAAGAAACC	Primer upstream of the EcoRI site in JW15 Δ100
D66640	NARL CTD R	CCAAGCTTTCAGAAAATGCGCTCCTGATGCACCCA	Anneals to 3′ terminal end of *narL* gene and includes a HindIII site
D66639	NARL CTD F	CCATGCATGCCACTACTGAGCGCGATGTTAACCA	Used with D66640 to construct NarL-CTD fragment
D67900	NARL FULL F	CCAATGCATAGTAATCAGGAACCGGCTACTATCCT	Used with D66640 to construct NarL-Full fragment
D70418	NARL CTD R ECO	CCGAATTCTCAGAAAATGCGCTCCTGATGCACCCA	Anneals to 3′ terminal end of *narL* gene and includes an EcoRI site
D70380	NARX PROMOTER FWD	GGGGATCCACCCATAGTGAGTACAGTGACT	Used with D70418 to construct the NarXL fragment
D66642	NARP CTD R	CCAAGCTTTTATTGTGCCCCGCGTTGTTGCAGG	Anneals to 3′ terminal end of *narP* gene and includes a HindIII site
D70662	NARP CTD F NSII	AATATGCATGGCGCGGAAGAAGATCCCTTCAGCG	Used with D66642 to construct NarL-CTD146 fragment
D71499	NARP CTD 138	AATATGCATTTACGTGAACGTGAAATGTTTGGC	Used with D66642 to construct NarL-CTD138 fragment
D71500	NARP CTD 148	AATATGCATGAAGATCCCTTCAGCGTGCTGAC	Used with D66642 to construct NarL-CTD148 fragment
D70607	NARP CTD R ECO	CCGAATTCTTATTGTGCCCCGCGTTGTTGCAGG	Anneals to 3′ terminal end of *narP* gene and includes an EcoRI site
D71497	NarP operon full F	ATGGATCCTGATGCATTG AGC TTT CAT CCT AT	Used with D70607 to construct the NarPop fragment

2.13.2. Transfer of fragments from pJW15 Δ100 to pLG339 – construction of pDLC3 and pDLC4

The fragments carrying DNA encoding the NarL CTD and full length NarL (see Table 2.8) were transferred to pLG339 so that comparisons could be made with the *narXL* operon expressed from pLG339 (see below). Vector was prepared by digestion of pLG339 with BamHI and EcoRI restriction enzymes (see Figure 2.4). The cut vector was treated with alkaline phosphatase and then purified by phenol-chloroform treatment and ethanol precipitation. The D5431 primer (see Table 2.9), which anneals upstream of the EcoRI site in pJW15 Δ100 was used with either the D70418 or D70607 primer that replaces the HindIII site at the 3′ end of fragments with an EcoRI site. Resulting PCR products were cut with BamHI and EcoRI restriction enzymes, purified by agarose gel electrophoresis, and ligated into prepared pLG339 vector (See Figure 2.7 and 2.8). Plasmids were transformed into JCB387 cells, which were grown overnight. Transformants were selected and plasmids recovered for sequencing.

2.13.3. Cloning of the narXL operon and the narP gene into pLG339 – Construction of pDLC5 and pDLC14

The *narXL* operon and *narP* gene were cloned with their regulatory regions into pLG339. Vector was prepared by digesting pLG339 with BamHI and EcoRI restriction enzymes. The cut vector was treated with alkaline phosphatase and was purified by treatment with phenol-chloroform and ethanol precipitation. The *narXL* operon and the *narP* gene were amplified from the chromosome of JCB387 using primers given in Table 2.9. The PCR products were digested with BamHI and EcoRI restriction enzymes, purified by agarose gel electrophoresis, and ligated into prepared pLG339 (see Figure 2.9 and 2.13). Plasmids were transformed into competent JCB3884 cells, which were grown overnight and the plasmids were recovered for sequencing

2.14. β-Galactosidase assays

2.14.1. Preparation of cells for assays

Plasmids carrying the promoter::*lacZ* fusions of interest were transformed into competent cells that possessed the desired genetic background. Transformants were plated onto MacConkey-lactose plates supplemented with relevant antibiotics and grown at 37°C overnight. Between 1 and 3 colonies were picked from each plate and used to inoculate separate overnight cultures.

2.14.2. Standard assays

β-Galactosidase assays were used to measure the activity of promoter::*lacZ* fusions. Unless stated otherwise, cells were assayed following growth to an OD_{650} of 0.5-0.6 in minimal salts medium with 10% LB, 0.4% glycerol and 40 mM sodium fumarate. If the aerobic activity of the promoter::*lacZ* fusion was to be assayed, a 25 ml conical flask was prepared with 5 ml of minimal medium as described above. Aerobic flasks were pre-warmed to 37°C and inoculated with 50 µl of overnight culture. Aerobic flasks were grown at 37°C with 250 rpm shaking. If the anaerobic activity of a promoter was to be assayed, a 16 cm growth tube was prepared with 10 ml of minimal salt medium as described above. Tubes were pre-warmed to 37°C and inoculated with 400 µl of overnight cultures. Tubes were placed in racks and grown at 37°C in an oven.

Once cultures reached an OD_{650} between 0.5 and 0.6, growth tubes and flasks were placed on ice to slow cell growth. To lyse aerobically grown cells, 30 µl of 1% sodium deoxycholate solution and 30 µl of toluene were added to each flask, and flasks were vigorously mixed. To lyse anaerobically grown cells, 2 ml of culture was transferred to a fresh tube and 30 µl of 1% sodium deoxycholate solution and 30 µl of toluene added to each tube, which was vigorously mixed. Toluene was evaporated by incubating growth tubes and flasks (without caps or bungs

respectively) at 37°C with 100 rpm shaking for 30 minutes. Tubes and flasks were then returned to ice.

Assay tubes were prepared with 2 ml of Z buffer and pre-warmed to 37°C. Growth vessels containing lysate were mixed thoroughly and 100 µl of the relevant lysate was added to each assay tube. To start the reaction, 500 µl of 13 mM ONPG solution was added to an assay tube, which was briefly mixed and placed in a 37°C water bath. The time until a yellow colour developed was measured and the reaction stopped by the addition of 1 ml of 1 M sodium carbonate solution. The OD_{420} of each sample was then measured. In the case of highly active promoters, the time until a yellow colour developed was less 1 minute. Thus, lysates from cells containing highly expressed promoter::*lacZ* fusions were diluted 2-fold with Z-buffer prior to being assayed.

β-Galactosidase activity was calculated as nanomols of ONPG hydrolysed per minute per milligram of bacterial mass using the equation given below :

$$\beta\text{-gal activity} = \frac{1000 \times 2.5 \times 3.6 \times OD_{420nm}}{OD_{650nm} \times 4.5 \times t \times v} = \text{nmol ONPG hydrolysed/min/mg bacterial mass}$$

Where:

2.5 = factor for conversion of OD_{650} into bacterial mass, based on OD_{650} of 1 being equivalent to 0.4 mg/ml bacteria (dry weight).

3.6 = final assay volume (ml)

1000/4.5 = factor for conversion of OD_{420} into nmol o-nitrophenyl (ONP), based on 1 nmol ml^{-1} ONPG having an OD_{420} of 0.0045

t = incubation time (min)

v = volume of lysate added (in ml)

2.14.3. Time course β-galactosidase assays following induction by nitrate

A time course assay was used to assay promoter activity following nitrate induction. Cells were prepared as in Section 2.14.1. 100 ml Duran bottles were prepared with 40 ml of minimal salts medium as described in Section 2.14.2. The bottles were pre-warmed in a 37 °C water bath without shaking. Growth medium was inoculated with 1.8 ml of overnight culture and mixed gently by pipetting, with care taken to not introduce bubbles. Cells were grown to an OD_{650} of ~0.4 and then the 0 time point taken by removal of 3 ml of culture to a separate tube. The OD_{650} of the sample was taken and the remainder of the sample placed on ice. The culture was then supplemented with 0.8 ml of a 1 M sodium nitrate solution or water (control flask) and a timer started. At the desired time points, 3 ml of culture was removed to a separate tube, the OD_{650} of the sample measured, and the sample tube placed on ice.

Once samples for all time points had been extracted the cells were lysed and β-galctosidase activity measured as in Section 2.14.2.

2.15. Electro mobility shift assays (EMSAs)

2.15.1. Preparation of DNA fragments for use in EMSAs

EcoRI-HindIII promoter fragments to be analysed by EMSA were cloned into pSR and transformed into JCB387 cells. Plasmids were recovered as and digested with EcoRI and HindIII restriction enzymes. Fragments were dephosphorylated with alkaline phosphatase and purified by polyacrylamide gel electrophoresis. DNA fragments were end-labelled with $[\gamma\,^{32}P]$ on 5′ terminal hydroxyl groups using T4 polynucleotide kinase (New England Biolabs). 5-17 µl DNA was mixed with 1 µl T4 polynucleotide kinase and 1 µl $[\gamma\,^{32}P]$-ATP in a final volume of 20 µl T4 polynucleotide kinase buffer, and incubated for 30 min at 37°C. Unincorporated $[\gamma\,^{32}P]$-ATP was removed by passing the sample through a Sephadex G-50 column. Columns were prepared by pipetting 400 µl 50% Sephadex G50 suspension into a Bio-Spin column (Biorad) and centrifuging for 2 minutes at 3500 rpm. The flow through was discarded and the

sample loaded onto the packed Sephadex in the column. The column was placed in a fresh tube and centrifuged for 2 minutes at 3500 rpm. The column was then discarded.

2.15.2. EMSA analysis of NarL binding to promoter fragments

Purified NarL protein was supplied by Dr Douglas Browning, University of Birmingham. NarL protein was diluted to the required concentrations in buffer B1. A micro-centrifuge tube was prepared containing 1 µl radiolabelled fragment, 1 µl FNR binding buffer, 1 µl 50% glycerol, 0.5 µl BSA, 0.5 µl herring sperm DNA, 1 µl acetyl phosphate and 4 µl sterile distilled water. Acetyl phosphate was included as a phosphate donor for NarL. 1 µl of NarL protein was added and samples were incubated at 37°C for 20 minutes and then loaded onto a 6% polyacrylamide gel. Gels were run in 0.5X TBE buffer at 160V for ~90 minutes and then soaked in fixing solution for ~30 minutes. Gels were dried and analysed as described in Section 2.9.4.

2.15.3. EMSA analysis of FNR binding to promoter fragments

Purified FNR-Ala154 protein was supplied by Dr Douglas Browning, University of Birmingham. FNR protein was diluted to the required concentrations in HEPES buffer. A micro-centrifuge tube was prepared containing 1 µl radiolabelled fragment, 1 µl FNR binding buffer, 1 µl 50% glycerol, 0.5 µl BSA, 0.5 µl herring sperm DNA and 5 µl sterile distilled water. 1 µl of FNR was added and samples were incubated at 37°C for 20 minutes and then loaded onto a 6% polyacrylamide gel. Gels were run in 0.5X TBE buffer at 160V for ~90 minutes and then soaked in fixing solution for ~30 minutes. Gels were dried and analysed as described in Section 2.9.4.

2.15.4. Preparation of lysate for bandshifts

It was necessary to use whole cell lysates as a source of NsrR because NsrR is difficult to purify. To eliminate other factors known to bind to the *hcp* promoter region from binding, JCB38849S Δ*nsrR* cells were used. Competent cells were transformed either with pBST-Blue

1 (control) or pGIT9, which is pBR322 containing *nsrR,* and transformants used to inoculated overnight cultures. Two conical flasks were prepared with 50 ml of 2X LB and inoculated with 0.5 ml of overnight culture of the cells containing either pBST-Blue 1 or pGIT9. Cultures were grown to an OD_{650} of 0.5-0.7 and placed on ice for 10 minutes. Cells were sedimented in a centrifuge at 4 °C and 4000 rpm for 10 minutes and re-suspended in 4 ml wash buffer. Cells were re-sedimented in a centrifuge at 4 °C and 4000 rpm for 10 minutes and re-suspended in 2 ml wash buffer. Cells were lysed by sonication and cell debris removed by centrifugation for 30 minutes at 13,000 rpm. Lysates were stored at 4°C before use, although not for longer than 48 hours. The protein concentration of lysate was quantified by Biorad Protein Assay (Biorad).

2.15.5. EMSA analysis of whole cell lysate with promoter fragments

Cell lysate was prepared as described above and diluted to the desired concentrations in wash buffer. For each sample and concentration, a microfuge tube was prepared with 1µl radiolabelled DNA, 1 µl FNR binding buffer, 1 µl glycerol, 0.5 µl herring sperm DNA, 3 µl cell lysate, 0.5 µl BSA and 3 µl of sterile distilled water. Samples were incubated at 37 °C for 20 minutes and then loaded onto a 6% polyacrylamide gel. Gels were run in 0.5X TBE buffer at 160V for ~90 minutes and then soaked in fixing solution for ~30 minutes. Gels were dried and analysed as described in Section 2.9.4.

2.16. DNase I protection assays – 'footprinting'

2.16.1. Preparation of radiolabelled DNA fragments for footprinting

Fragments to be analysed by footprinting were cloned into pSR. A maxiprep of pSR carrying the fragment of interest was prepared as described in Section 2.7.7. and 100 µl aliquots were digested by 8 µl HindIII restriction enzyme in a total volume of 120 µl. The restriction products were treated with alkaline phosphatase to remove terminal phosphates and purified by phenol-chloroform treatment and ethanol precipitation. The DNA was then

digested with 8 µl AatI restriction enzyme in a total volume of 80 µl. AatI-HindIII fragments were then purified on a polyacrylamide gel followed by phenol-chloroform treatment and ethanol precipitation. Fragments were then radiolabelled with $[\gamma\text{-}^{32}P]$ on the HindIII end using T4 polynucleotide kinase as described in Section 2.15.1)

2.16.2. Preparation of GA ladder

A GA ladder is a fragment cleaved at G and A nucleotides and allows calibration of a gel to aid determining the position of protein binding. Preparation of the ladder is done according to method of Maxam and Gilbert (1980). DNA was prepared for cleavage by depurination by formic acid. 4 µl of radiolabelled DNA was diluted to a final volume of 12 µl with sterile distilled water and 50 µl of formic acid was mixed with the sample. The sample was incubated for 150 seconds at room temperature and the reaction stopped by the addition of 200 µl 0.3 M sodium acetate (pH 7) and 700 µl of ice-cold ethanol. The DNA was then precipitated as described in Section 2.7.2.The DNA pellet was re-suspended in 100 µl of a 1 M piperidine solution and incubated at 90°C for 30 minutes, during which time piperidine cleaves the bases modified by formic acid treatment. The DNA was then precipitated as described in Section 2.7.2 and the pellet was re-suspended in 20 µl of loading buffer for use in gel calibration.

2.16.3. DNase I footprinting with FNR protein

DNase I footprinting was used to determine the position on the *hcp* promoter fragment that FNR binds to. FNR protein stock (provided by Douglas Browning of the University of Birmingham) was diluted to the concentrations required in HEPES dilution buffer. Each reaction contained 1 µl radiolabelled AatI-HindIII fragment, 2 µl HEPES-Glutamate buffer, 1 µl BSA, 1 µl herring sperm DNA, 1 µl FNR protein and 14 µl sterile distilled water. Samples were incubated at 37°C for 20 minutes to allow FNR to bind the fragment. 3 µl of a 1:1000 dilution of DNase I (10 U/µl, Roche Applied Science) in sterile water was added to each

reaction to digest the DNA. Samples were incubated for ~90 seconds and the reaction was stopped by the addition of 200 µl DNase I stop solution. Samples were then purified by treatment with phenol-chloroform and ethanol precipitation and analysed by polyacrylamide sequencing gel as described in Section 2.6.3

2.16.4. DNase I footprinting with NarL protein

DNase I footprinting was used to determine the position on the *hcp* promoter fragment that NarL binds to. NarL protein stock (provided by Douglas Browning of the University of Birmingham) was incubated with acetyl phosphate to ensure phosphorylation of the protein. The phosphorylation reaction contained 3 µl of NarL stock, 2.7 µl HEPES-glutamate buffer and 0.3 µl of a 500 nM acetyl phosphate solution. The phosphorylation reaction was incubated at 37°C for 20 minutes and then diluted with buffer B1 to the concentrations required for DNase I footprinting. Each reaction contained 1 µl radiolabelled AatI-HindIII fragment, 2 µl HEPES-Glutamate buffer, 1 µl BSA, 1 µl herring sperm DNA, 1 µl NarL protein and 14 µl sterile distilled water. Samples were incubated at 37°C for 20 minutes to allow NarL to bind the fragment. 3 µl of a 1:1000 dilution of DNase I (10 U/µl, Roche Applied Science) in sterile water was added to each reaction to digest the DNA. Samples were incubated for ~90 seconds and the reaction was stopped by the addition of 200 µl DNase I stop solution. Samples were then purified by treatment with phenol-chloroform and ethanol precipitation and analysed by polyacrylamide sequencing gel as described in Section 2.6.3

Chapter 3: Regulation of the *E. coli hcp* promoter

3.1. Introduction

Transcription regulation in response to nitrate or nitrite often involves the integration of multiple signals at promoters, making it difficult to deduce an accurate picture of regulation from data obtained by genomics experiments. Therefore single promoter studies are often necessary to understand and uncover the mechanisms of regulation by multiple factors.

Regulation of the *hcp* gene is of particular interest as the function of the gene product, HCP, is currently unknown, but the *hcp* promoter has been shown to be highly active when fully induced (Filenko *et al.*, 2005). This suggests that the function of HCP is important and greater understanding of the regulation of *hcp* expression could help to ascertain that function. The *hcp* promoter has been studied previously and has been shown to be activated in response to anaerobiosis and by the presence of nitrate and nitrite (Filenko *et al.,* 2005). Filenko and colleagues also demonstrated that NarL and NarP are necessary for nitrate and nitrite induction of *hcp* expression. A later study reported binding of purified FNR and NarL proteins to the *hcp* promoter region *in vitro* and identified NsrR as a repressor of *hcp* expression (Filenko *et al.,* 2007). In this work, the regulation of *hcp* transcription has been studied using *in vivo* and *in vitro* methods.

3.2. Deletion analysis of the *hcp* regulatory region

The starting fragment for this work was hcp383-0 (see Figure 3.1), a 383 bp fragment that contains the entire intergenic region between *hcp* and the upstream gene *ybjE*. The fragment contains the base sequence from position -285 to +98 relative to the σ^{70}-dependent *hcp* transcript start, which was previously identified using 5′ RACE by Filenko *et al.* (2005). The DNA from position -285 to -113 encodes the terminal portion of *ybjE* but does not include the reported σ^{24} dependent *hcp* promoter (Rhodius *et al.*, 2006). Filenko *et al.* (2005, 2007) also proposed transcription factor binding sites in the *hcp* regulatory region based on similarities to published consensus binding sequences. Immediately downstream of the *ybjE* stop

Figure 3.1: The DNA sequence of the hcp383-0 promoter fragment

Numbered positions are given relative to the *hcp* transcript start that was identified using 5′ RACE by Filenko *et al.* (2005) and is shown as an underlined, emboldened 'T' and a '+1' above the sequence. The emboldened sequence from position -115 denotes the stop codon for *ybjE*. Predicted DNA sites for NarL, FNR and NsrR are underlined and marked with coloured arrows underneath the sequence and are aligned with the appropriate consensus binding sequence (in bold). Fasta standard ambiguity codes are used: a Y denotes C or T; an M denotes A or C; a K denotes G or T; an R denotes A or G and an N denotes any base. The centre of the site is given in brackets beneath the arrows. A coordinate finishing in '.5′ denotes that the centre of the site falls between two bases. The -10 and -35 elements identified in this study are underlined and labelled, as is the published start codon. The starting points of other fragments used in this work are marked by a triangle above the sequence.

-280 -260

GAATTC TGATTCGCCGCAGCCGCTCTACTGCACTGGGCTTATGCGGTGCC

EcoRI

-240 -220 -200

ACATCAATGGATTTCACCCTGCCCGTTCTTCAACGTACTGGCGGGCTGGA

-180 -160

TATGGTCCCGGCGGCAATTGTTCACGGTTTTATTCTTAGCCTGTTAGTGC

hcp383-1 **hcp383-2**

▼ ▼

-140 -120

CGATCCTCATCGCCTTTTTCTCTGCG**TAA**TACCTCTCTGGCGGTAGATCC

TACYYMTnnAKRRGTA

NarL (-104.5)

-80 **hcp383-3** -60

▼

CTGCCGCCAAATTGCGCTAAATCAATCTCCCTTAAAGTTGCATGAAAAA

TTGATnnnnATCAA

FNR (-72.5)

-40 -20 +1

TCCCTTTTATCCCCGCGTTAAGCGTCTTAACCTTAAACATG**T**ATATTAAA

-35 -10 **AANATGCATTTnAAA**

NsrR (+6)

+20 +40

TATAACTTTAAAAGGTGTGATCATGTTTTGTGTGCAATGTGAACAAACTA

TGCATNTT Start codon

+60 +80

TCCGTACTCCGGCAGGAAACGGCTGCTCATACGCGCAGG AAGCTT

HindIII

93

codon is a predicted DNA site for NarL/P, a 7-2-7 inverted repeat centred at position -104.5. There is a predicted DNA site for FNR centred at position -72.5 and a predicted DNA site for NsrR centred at position +6. The start codon for *hcp* is located at position +32.

To study the roles of the predicted transcription factor binding sites upstream of the *hcp* transcript start, a nested series of truncated promoter fragments, derived from the hcp383-0 fragment, was used (see Figure 3.1). The hcp383-1 fragment carries a deletion of bases that encode the terminal portion of *ybjE*; the hcp383-2 fragment carries a longer deletion that includes the DNA site for NarL centred at position -104.5, and the hcp383-3 fragment carries a longer deletion that includes the DNA site for FNR centred at position -72.5. Each of the fragments was cloned into pRW50 and expression of the resulting promoter::*lacZ* fusions was measured in JCB387 cells grown either aerobically or anaerobically, with or without nitrate or nitrite supplements.

Results illustrated in Figure 3.2 show that expression of the hcp383-0::*lacZ* fusion is induced by anaerobiosis (when FNR is active), further induction is observed with nitrate (when NarL is thought to be most active) but, in agreement with Filenko *et al* (2005), maximal activity is found when the cells are grown in the presence of nitrite. Similar patterns of expression were observed with the hcp383-1::*lacZ* fusion, demonstrating that sequence upstream of position -125 plays no role in the regulation of *hcp* promoter activity under the conditions tested.

Results also show that the removal of the DNA site for NarL/P in the hcp383-2 fragment prevents induction by nitrate whilst not affecting induction due to nitrite. This suggests that NarL is directly responsible for at least some nitrate dependent activation of the *hcp* promoter. In contrast, NarL binding to the DNA site for NarL at position -104.5 is not essential for induction of the *hcp* promoter by nitrite
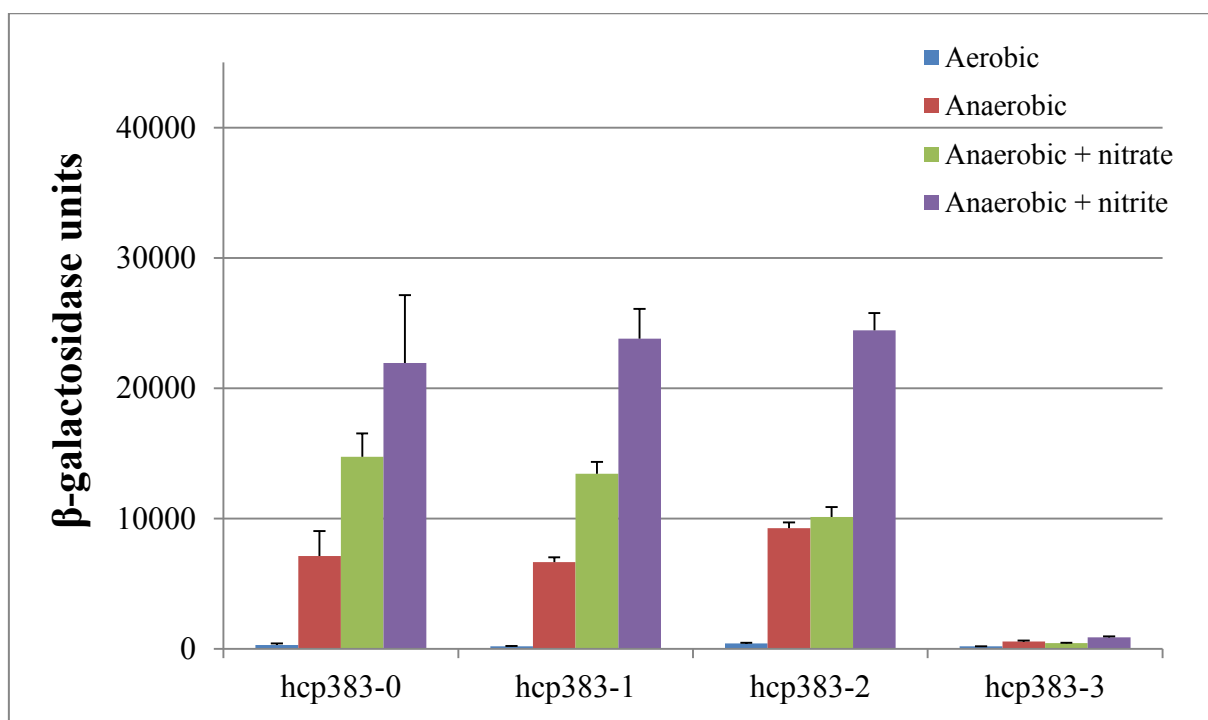
Figure 3.2: Effects of deletions on the activity of the *hcp* promoter

Expression from the full-length hcp383-0 fragment and truncated fragments, cloned as promoter::*lacZ* fusions in pRW50, was measured in *E. coli* strain JCB387 during exponential growth (OD_{650} = 0.5-0.6) in minimal salt medium. Cells were grown aerobically, anaerobically, anaerobically with medium supplemented with 20 mM sodium nitrate or anaerobically with medium supplemented with 2.5 mM sodium nitrite. Data shown are averages from at least 3 biological repeats, and error bars show one standard deviation from the mean.

Finally, removal of the DNA site for FNR in the hcp383-3 fragment prevents all induction of the *hcp* promoter and measured levels of β-Galactosidase were similar in all conditions tested. This corroborates the hypothesis that FNR binding to the proposed target site at position -72.5 induces transcription initiation at the *hcp* promoter.

3.3. Predicted DNA sites for NarL and FNR are important *in vivo*

To ensure that the effects of truncation on *hcp* promoter activity (illustrated in Figure 3.2) are due to the removal of the intended DNA site and not due to removal of other features, point mutations were introduced into the predicted DNA sites.

To confirm the importance of the DNA site for NarL *in vivo,* a derivative of the hcp383-1 fragment, hcp383-101, was constructed with mutations in the DNA site for NarL (see Figure 3.3, panel A). To confirm the importance of the DNA site for FNR *in vivo,* a second derivative of the hcp383-1 fragment, hcp383-102, was constructed with mutations in the DNA site for FNR (see Figure 3.3, panel B). The resulting fragments were cloned into pRW50 and expression of the resulting *lacZ* fusions was measured in JCB387 cells.

Results illustrated in Figure 3.3, panel C, show that the hcp383-1::*lacZ* fusion is induced by nitrate and is optimally induced by nitrite. The hcp383-101::*lacZ* fusion, which has mutations in the DNA site for NarL, shows an identical pattern of expression to that of the hcp383-2::*lacZ* fusion, in which DNA including the proposed NarL site is deleted. Hence, mutation of the DNA site for NarL prevents induction by nitrate, but does not affect induction by nitrite. The hcp383-102::*lacZ* fusion, which has mutations in the DNA site for FNR, shows an identical pattern of expression to that of the hcp383-3::*lacZ* fusion, in which DNA including the proposed site for FNR is deleted. Hence, mutation of the DNA site for FNR prevents all induction. These results indicate that the proposed DNA sites for NarL and FNR are valid and support the conclusions of Section 3.2.

A NarL (-104.5)

TACYYMTnnAKRRGTA
hcp383-1:   GTAATACCTCTCTGGCGGTAGATC
hcp383-101: GTAATAGCTCTCTGGCGCTAGATC

B FNR (-72.5)

TTGATnnnnATCAA
hcp383-1:   CCAAAATTGCGCTAAATCAATCTCCCAA
hcp383-102: CCAAAATCGCGCTAAATCGATCTCCCAA

C
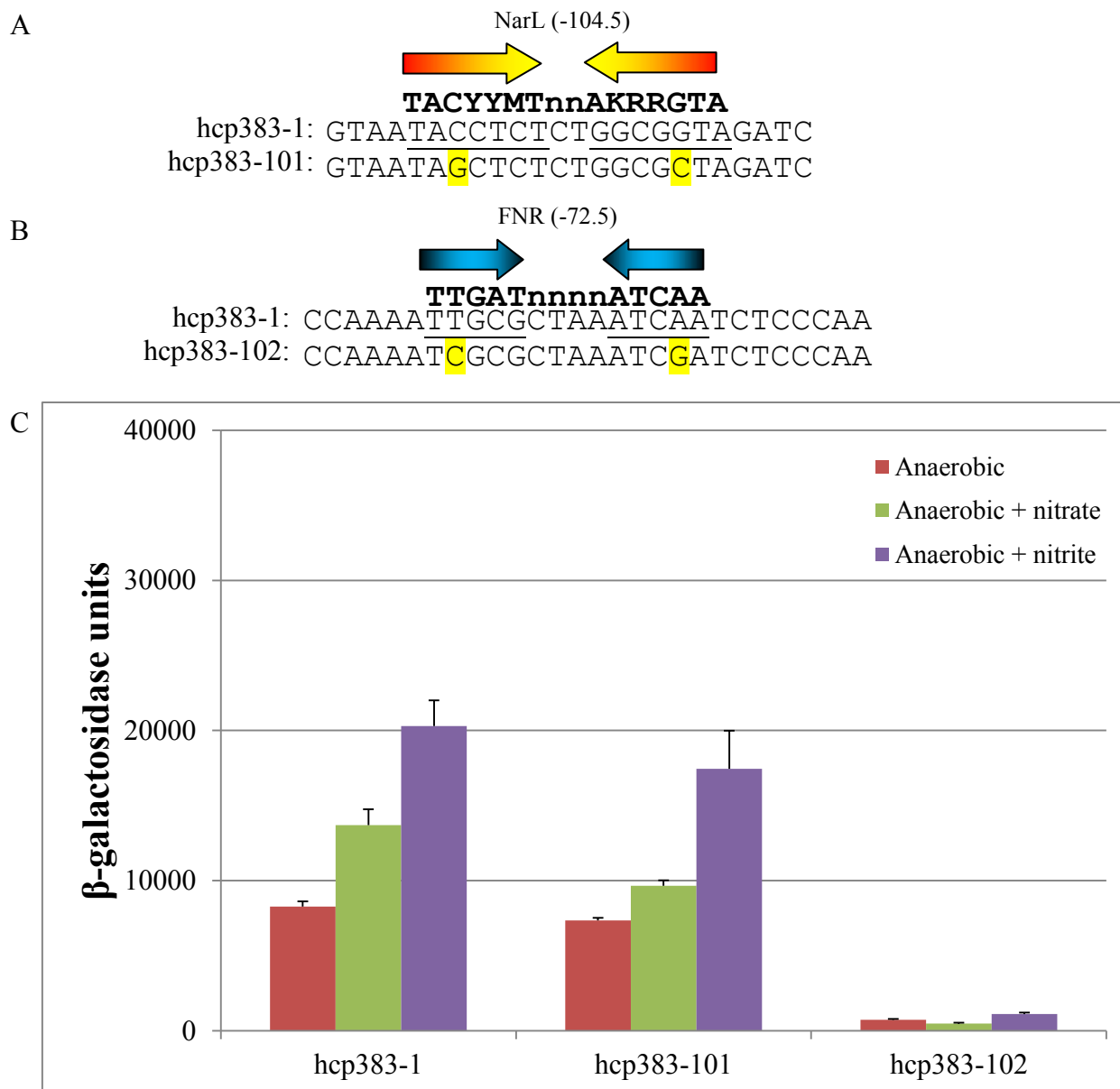
Figure 3.3: Effect of mutations in DNA sites for NarL and FNR on expression from the *hcp* promoter

Panel A shows a partial sequence of the *hcp* promoter fragment hcp383-1 that contains the DNA site for NarL, and a derivative carrying mutations in the site, hcp383-101. Mutations are highlighted in yellow. The DNA site for NarL is indicated by arrows above the sequence and the consensus sequence for a DNA site for NarL is given in bold above the sequence.

Panel B shows a partial sequence of the *hcp* promoter fragment hcp383-1 that contains the DNA site for FNR, and a derivative carrying mutations in the site, hcp383-102. Mutations are highlighted in yellow. The DNA site for FNR is indicated by arrows above the sequence and the consensus sequence for a DNA site for FNR is given in bold above the sequence.

Panel C shows expression of hcp383-1::*lacZ,* hcp383-101::*lacZ* and hcp383-102::*lacZ* fusions measured in JCB387 cells grown to mid exponential phase (OD_{650} 0.5-0.7) in minimal salt medium. Cells were grown anaerobically, anaerobically with medium supplemented with 20 mM sodium nitrate or anaerobically with medium supplemented with 2.5 mM sodium nitrite. Data shown are averages of at 3 biological repeats and error bars show one standard deviation from the mean.

3.4. FNR is an important activator of the *hcp* promoter

To investigate further the role of FNR in regulation of *hcp* promoter activity, the expression of the hcpc383-0::*lacZ* fusion was measured in JCB387 Δ*fnr* cells. Results illustrated in Figure 3.4 show that in Δ*fnr* cells, minimal induction of the *hcp* promoter is observed and expression of the fusion remains at the level of cells grown aerobically. This confirms that FNR is the major activator of *hcp* expression.

3.5. FNR binds *in vitro* to the DNA site at position -72.5 in the *hcp* promoter

Filenko *et. al.* (2005) predicted a single site DNA site for FNR at position -72.5 at the *hcp* promoter and promoter truncations/mutations support this hypothesis. Purified Ala154-FNR protein and ^{32}P labelled hcp383-0 fragment was used in an EMSA experiment to confirm that FNR binds to the *hcp* regulatory region. Results in Figure 3.5, panel A, show a single shift at the lowest concentration of FNR protein used (0.5 μM) and a supershift when 2 μM or higher concentrations of FNR are used. This agrees with that observed by Filenko *et. al.* (2007) and suggests that FNR has a high affinity for its site at the *hcp* promoter. DNase I footprinting was used to map the region to which FNR binds on the hcp383-0 fragment. Results in Figure 3.5, panel B, show a single zone of protection that corresponds to the DNA site for FNR predicted by Filenko *et al.* (2005).

Taken with the *in vivo* evidence, the *in vitro* evidence strongly supports the hypothesis that FNR is necessary for the activation of the *hcp* promoter, and that it does so by acting from the DNA site centred at position -72.5 on the *hcp* promoter.

3.6. NsrR represses *hcp* promoter activity

Filenko *et al.* (2007) showed that NsrR represses *hcp* promoter activity. To investigate this repression, the activity of the *hcp* promoter carried by each of the promoter fragments described in Figure 3.1 was measured in a Δ*nsrR* derivative of JCB387.
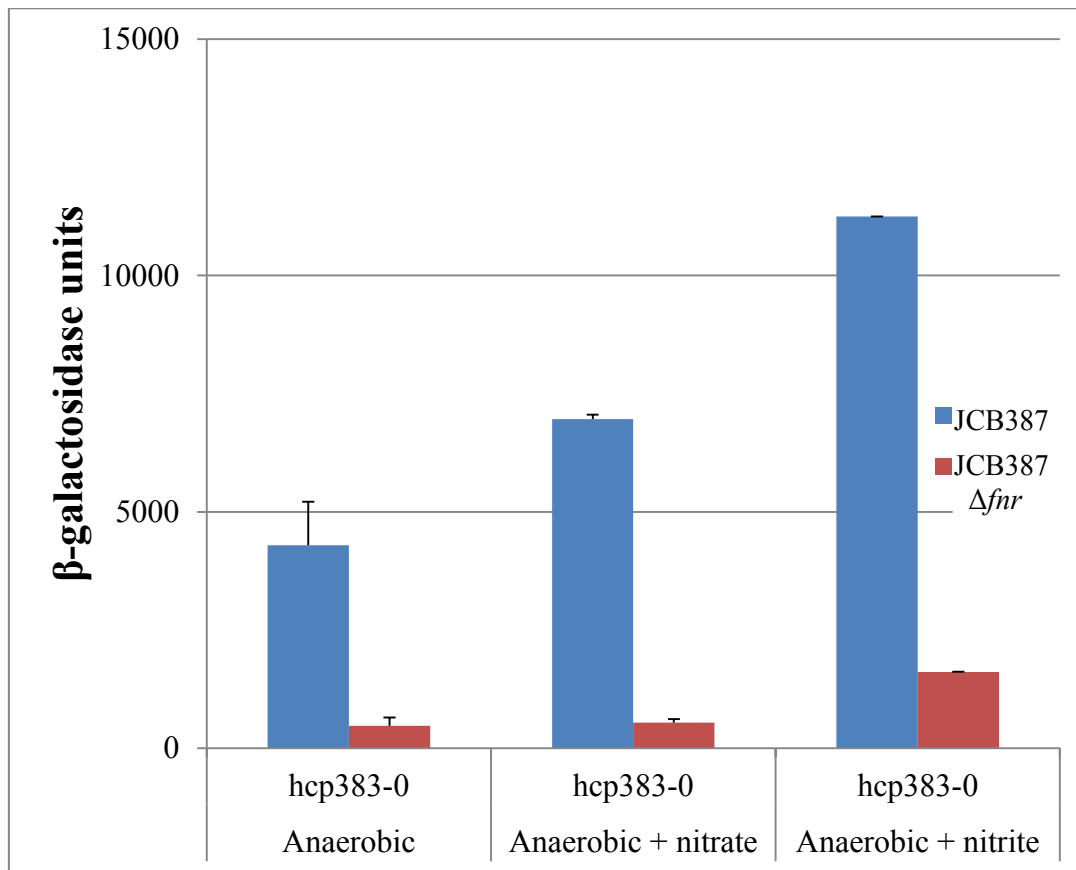
Figure 3.4: FNR is required for activation of the *hcp* promoter

Expression of the hcp383-0::*lacZ* fusion, cloned in pRW50 and measured in JCB387 cells or JCB387 Δ*fnr* cells. Cells were grown to mid exponential phase (OD$_{650}$ 0.5-0.7) in minimal salt medium either anaerobically, anaerobically with medium supplemented with 20 mM sodium nitrate or anaerobically with medium supplemented with 2.5 mM sodium nitrite. Data shown are averages from 3 biological replicates and error bars show one standard deviation from the mean.

Figure 3.5: Binding of FNR to the *hcp* regulatory region

Panel A shows the results of an electrophoretic mobility shift assay using the end-labelled hcp383-0 promoter fragment, incubated with increasing concentrations of FNR. The concentrations of FNR used were: lane 1, no protein; lane 2, 0.5 μM; lane 3, 1.0 μM; lane 4, 2.0 μM; lane 5, 4.0 μM

Panel B shows a DNase I footprint of purified FNR binding at the *hcp* promoter. End-labelled hcp383-1 AatII-HindIII fragment was incubated with increasing concentrations of Ala154 FNR and subjected to DNase I footprint analysis. The concentration of FNR was: lane 1, no protein; lane 2, 0.5 μM; lane 3, 1.0 μM; lane 4, 2.0 μM; lane 5, 4.0 μM. The gel was calibrated using a Maxam-Gilbert 'G+A' sequencing reaction (Lane GA) and relevant positions are indicated. The location of the DNA site for FNR is shown by a box.

A

[FNR]
µM

| 1 | 2 | 3 | 4 | 5 |

- 0.5 1.0 2.0 4.0

B

[FNR]
µM

GA 1 2 3 4 5

0.5 1.0 2.0 4.0

FNR

-90
-80
-70
-60
-50

-10

+1
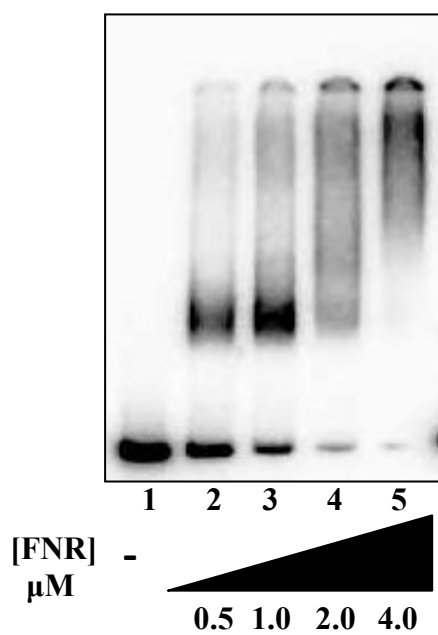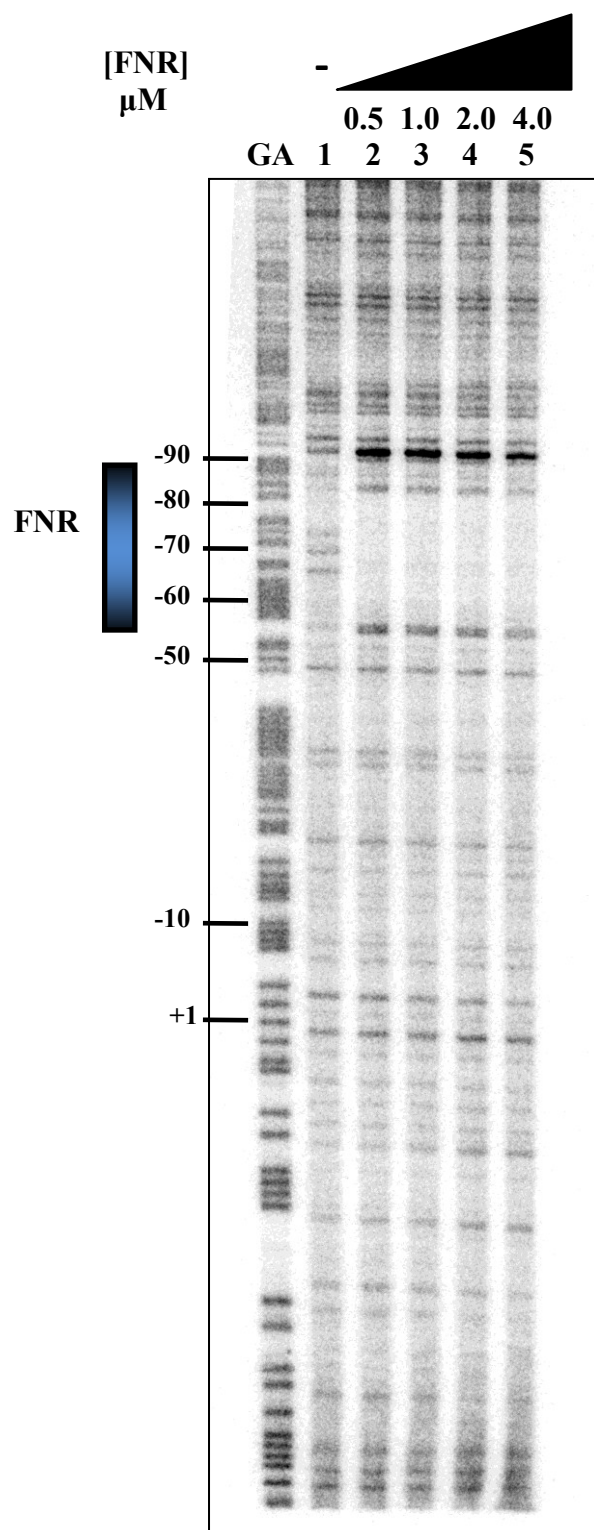
Results illustrated in Figure 3.6 show that *hcp* promoter activity (as measured with the hcp383-0 and hcp383-1 fragments) is higher in Δ*nsrR* cells than in JCB387 cells during anaerobic growth. The promoter appears to be almost fully induced in the Δ*nsrR* background and there is little further induction by nitrite, although activity is reduced by the presence of nitrate. These results confirm that NsrR represses the *hcp* promoter.

Results illustrated in Figure 3.6 also show that removal of the DNA site for NarL in the hcp383-2 fragment has no effect on *hcp* promoter activity in Δ*nsrR* cells and that removal of the DNA site for FNR in the hcp383-3 fragment prevents all induction of *hcp* promoter activity. Taken together, the results in Figure 3.2 and Figure 3.6 indicate that the *hcp* promoter is activated by FNR in response to anaerobiosis by acting from its DNA site centred at position -72.5 and that NsrR is a major regulator of *hcp* promoter activity whilst NarL appears to play a minor role by mediating a small level of induction in response to nitrate in cells that contain NsrR.

3.7. The proposed DNA site for NsrR is important *in vivo*

An *in vivo* effect of NsrR on *hcp* promoter activity has been demonstrated using cells that were deficient in NsrR by Filenko *et al.* (2007) and by results illustrated in Figure 3.6. To confirm the importance of the proposed DNA site for NsrR *in vivo*, a derivative of the hcp383-1 fragment, hcp383-1n, was constructed with mutations in the site. The mutations were designed to change positions identified as highly conserved in DNA site for NsrR (Rodionov *et al.*, 2005) but to not affect positions in the immediate vicinity of the transcription start site for *hcp* (see Figure 3.7, panel A).

The hcp383-1n fragment was cloned into pRW50 and the activities of the hcp383-1::*lacZ* fusion and the hcp383-1n::*lacZ* fusion were measured in JCB387 cells. The results illustrated in Figure 3.7, panel B, show that the hcp383-1n::*lacZ* fusion is more highly expressed than the hcp383-1::*lacZ* fusion in all conditions, indicating that mutations in the DNA site for
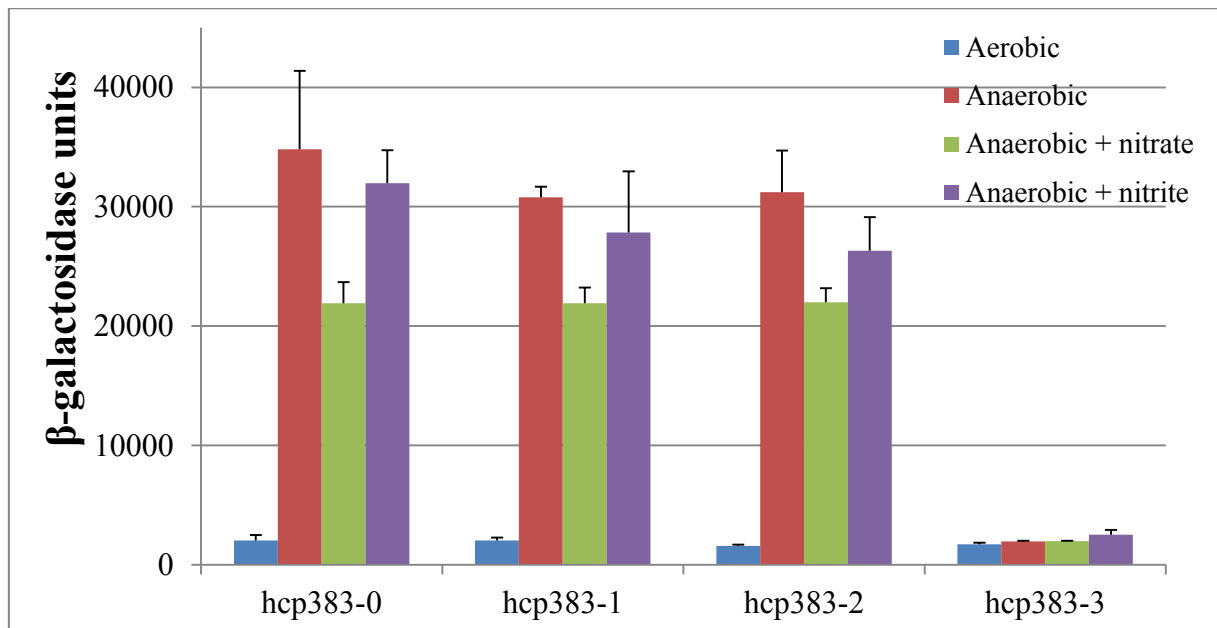
Figure 3.6: Activity of truncated *hcp* promoter fragments in Δ*nsrR* cells

Expression from the full-length hcp383-0 fragment and truncated fragments, cloned as promoter::*lacZ* fusions in pRW50, was measured in a *ΔnsrR* derivative of the *E. coli* strain JCB387 during exponential growth (OD$_{650}$ = 0.5-0.7). Cells were grown in minimal salt medium aerobically, anaerobically, anaerobically with medium supplemented with 20 mM sodium nitrate or anaerobically with medium supplemented with 2.5 mM sodium nitrite. Data shown are averages from at least 3 biological repeats, and error bars show one standard deviation from the mean

Figure 3.7: Effects of mutations in the predicted NsrR site on *hcp* promoter activity

Panel A shows a partial sequence of the *hcp* promoter fragment hcp383-1 that contains the DNA site for NsrR, and a derivative carrying mutations in the site, hcp383-1n. Mutations are highlighted in yellow. The DNA site for NsrR is indicated by arrows above the sequence and the consensus sequence for a DNA site for NsrR is given in bold above the sequence.

Panel B shows expression of the hcp383-1::*lacZ* and hcp383-1n::*lacZ* fusions cloned in pRW50 and measured in JCB387 cells. Cells were grown to mid exponential phase (OD_{650} 0.5-0.7) in minimal salt medium aerobically, anaerobically, anaerobically with medium supplemented with 20 mM sodium nitrate or anaerobically with medium supplemented with 2.5 mM sodium nitrite. Data shown are means of 3 biological repeats and error bars show one standard deviation from the mean.

Panel C shows expression of the hcp383-1::*lacZ* and hcp383-1n::*lacZ* fusions cloned in pRW50 and measured in a Δ*nsrR* derivative of JCB387 cells. Cells were grown to mid-exponential phase (OD_{650} 0.5-0.7) in minimal salt medium aerobically, anaerobically, anaerobically with medium supplemented with 20 mM sodium nitrate or anaerobically with medium supplemented with 2.5 mM sodium nitrite. Data shown are means of 3 biological repeats and error bars show one standard deviation from the mean.

A

NsrR (+6)

AANATGCATTTnAAATGCATNTT
hcp383-1: CCTTAAACATGTATATTAAATATAACTTTAAA
hcp383-1n: CCTTAAACATGTATATTAAGGATAACTTTAAA

B

JCB387 cells

C

JCB387 ΔnsrR cells
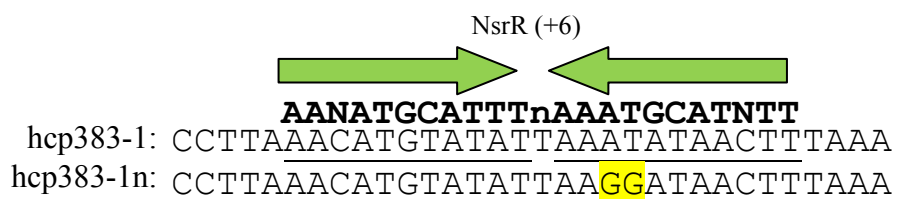
NsrR relieve repression. Activity of the hcp383-1n::*lacZ* fusion is induced by anaerobiosis and by nitrite but does not appear to be induced by nitrate.

The expression of the hcp383-1:*lacZ* fusion and the hcp383-1n::*lacZ* fusion was also measured in JCB387 Δ*nsrR* cells. Results illustrated in Figure 3.7, panel C, show that no significant differences in expression between the two fusions are observed. Anaerobiosis induces expression of both fusions, whilst nitrate has no effect. Nitrite appears to repress activity of both fusions although this effect is less significant.

3.8. NsrR binds *in vitro* to the predicted site at position +6 in the *hcp* promoter

The *E. coli* NsrR protein is a highly oxygen-labile protein and as such is not available in a purified form for use in *in vitro* studies. Therefore, lysates derived from cells that express plasmid encoded NsrR were used as described by Browning *et al.* (2010). To increase the probability of a successful EMSA using whole cell lysate, it is necessary to minimize the binding of other factors to the DNA fragment, and so lysate was prepared from cells that are deficient in the nucleoid-associated protein IHF, NarL/P and NsrR. The hcp383-3 fragment was truncated downstream of the *hcp* transcript start to produce the hcp383-3s fragment and a derivative of the hcp383-3s fragment was constructed, hcp383-3ns, in which the DNA site for NsrR was mutated to prevent binding by NsrR. Both fragments were radiolabelled.

Results illustrated in Figure 3.8 show that when lysate from cells that express NsrR is incubated with the hcp383-3s fragment, a single shift is observed at the lowest concentration of lysate tested (1.5 mg/ml), which indicates strong binding of NsrR to the fragment. As expected, the mutations in the DNA site for NsrR in the hcp383-3ns fragment greatly reduce NsrR binding. No lysate concentration-dependent shifts are observed when lysate from cells that do not express NsrR is incubated with either the hcp383-3s fragment or the hcp383-3ns fragment. These results indicate that NsrR binds to its predicted site centred at position +6 and
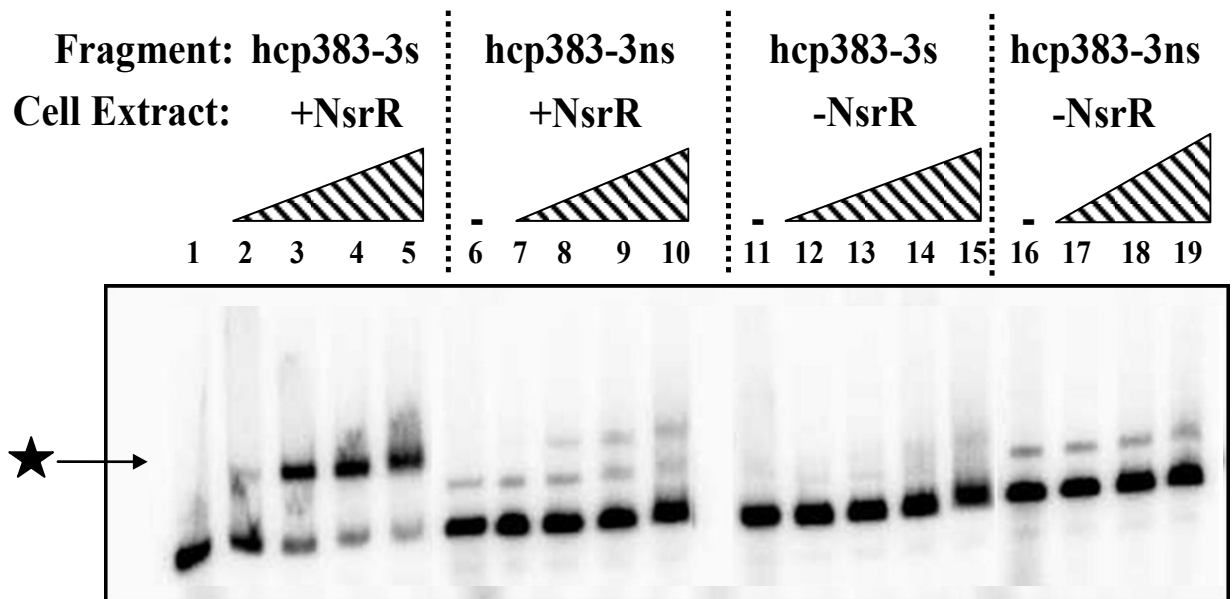
Figure 3.8: Effect of NsrR site mutations on binding of NsrR to the *hcp* promoter region

The figure shows an autoradiogram from an EMSA with end labelled hcp383-3s fragment (lanes 1-5 and 11-15) or the derivative hcp383-3ns fragment with the +9G and +10G mutations that abrogate NsrR binding (lanes 6-10 and 16-19). Labelled fragments were incubated with increasing amounts of soluble cell extract from strain JCB38849S Δ*nsrR* carrying either pGIT9 which expresses NsrR (+NsrR; lanes 2-5 and 7-10), or empty vector pSTBlue-1 (-NsrR; lanes 12-15 and 17-19). The total amount of protein in the extract used in each reaction was: lanes 1, 6, 11 and 16, no protein; lanes 2, 7, 12 and 17, 1.5 µg; lanes 3, 8, 13 and 18, 3 µg; lanes 4, 9, 14 and 19, 6 µg; lanes 5, 10 and 15, 9 µg. The location of the band due to the specific binding of NsrR is indicated by a filled star .

that the engineered mutations at positions +9 and +10 do mostly, but not entirely, prevent NsrR from binding to the fragment.

3.9. NarL is required for maximal activation of the *hcp* promoter

Filenko *et al.* (2005) demonstrated that maximal activity of an *hcp* promoter::*lacZ* fusion requires NarL or NarP. This observation is apparently at odds with the results illustrated in Figure 3.2, which indicate that the DNA site for NarL at the *hcp* promoter is not required for maximal activity. To reinvestigate the effects of NarL and NarP on *hcp* promoter activity, the expression from the hcp383-1::*lacZ* fusion and the hcp383-2::*lacZ* fusion (in which the DNA site for NarL is deleted) was measured in cells that contained both NarL and NarP, NarL only, NarP only, or neither.

Results illustrated in Figure 3.9, panel A, show that expression of promoter::*lacZ* fusions, measured in JCB387 cells, which contain both NarL and NarP, is very similar to that in results illustrated in Figure 3.2. Expression of the hcp383-1::*lacZ* fusion is maximally induced by nitrite but is also induced to a lesser extent by nitrate. Expression of the hcp383-2::*lacZ* fusion is also maximally induced by nitrite, but is not induced by nitrate. This indicates that the DNA site for NarL centred at position -104.5 is not required for maximal activation of the *hcp* promoter by nitrite but is required for the smaller nitrate-dependent induction.

The results illustrated in Figure 3.9, panel B show the expression of the promoter::*lacZ* fusions measured in JCB3875 cells, which contain NarL but not NarP. Both fusions show a similar pattern of induction to that observed in JCB387 cells that express NarP. Thus, expression of the hcp383 1::*lacZ* fusion is induced by nitrate and is maximally induced by nitrite, whilst expression of the hcp383-2::*lacZ* fusion is not induced by nitrate but is maximally induced by nitrite. This demonstrates that NarP is not required for maximal activation of the *hcp* promoter.

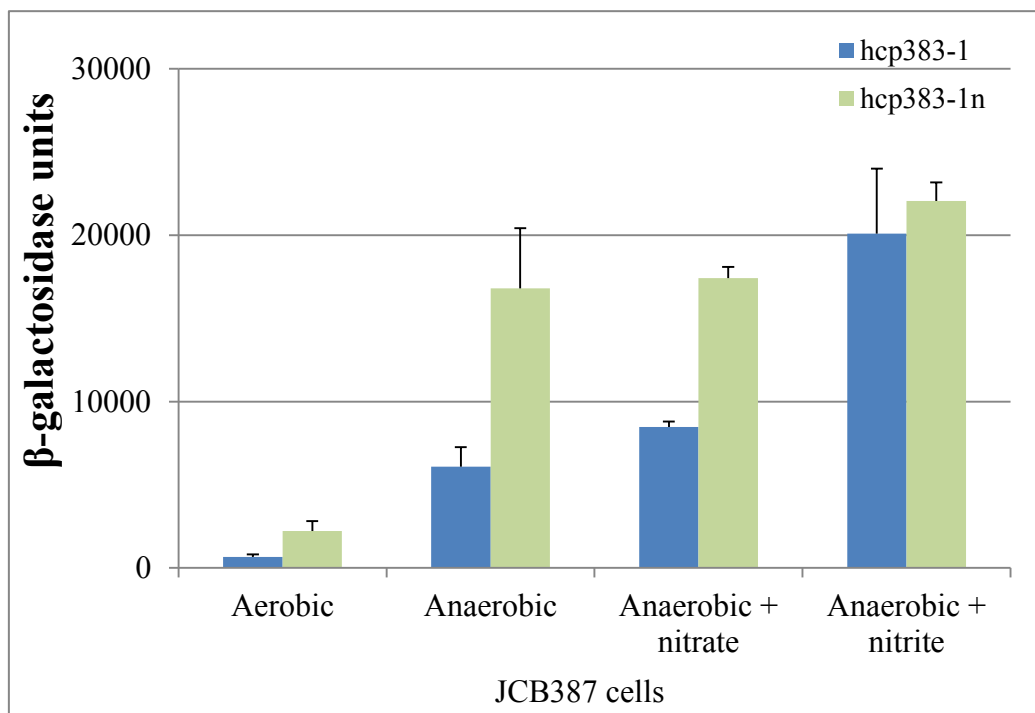**Figure 3.9: Activity of the *hcp* promoter and a derivative lacking the DNA site for NarL**

Panel A shows expression of the hcp383-1::*lacZ* and hcp383-2::*lacZ* fusions cloned in pRW50 and measured in JCB387 cells. Cells were grown to mid exponential phase (OD_{650} 0.5-0.7) in minimal salt medium anaerobically, anaerobically with medium supplemented with 20 mM sodium nitrate or anaerobically with medium supplemented with 2.5 mM sodium nitrite. Data shown are means of 3 biological repeats and error bars show one standard deviation from the mean.
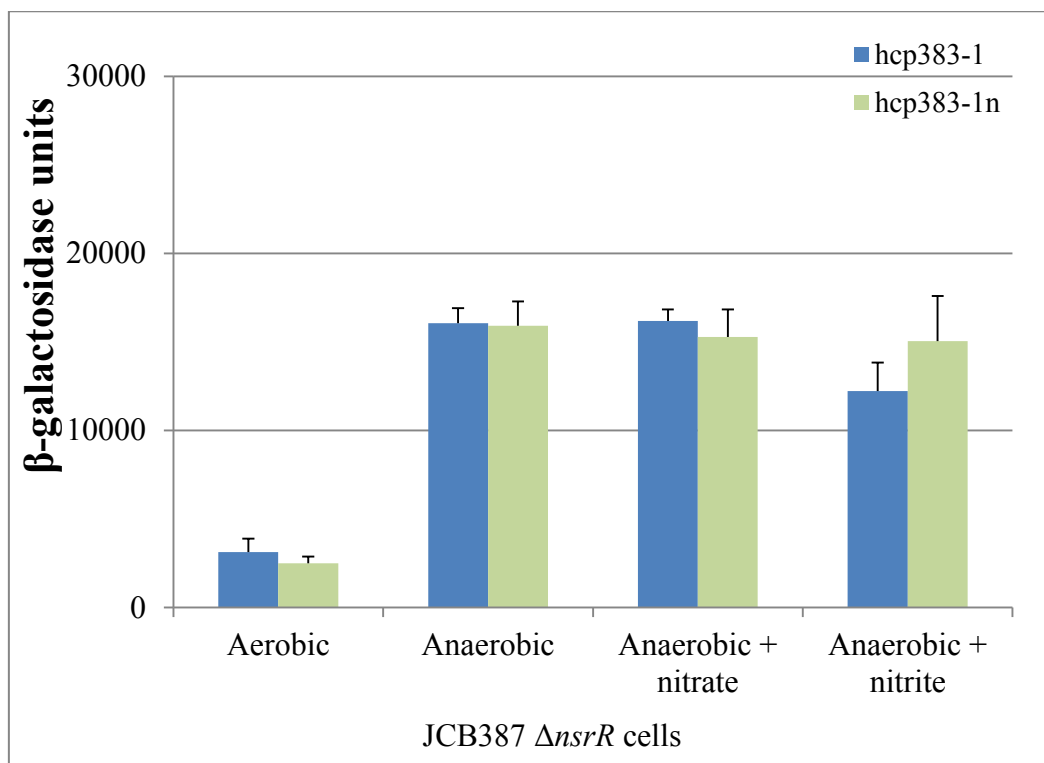
Panel B shows expression of the hcp383-1::*lacZ* and hcp383-2::*lacZ* fusions cloned in pRW50 and measured in JCB3875 cells, which do not express NarP. Cells were grown to mid exponential phase (OD_{650} 0.5-0.7) in minimal salt medium anaerobically, anaerobically with medium supplemented with 20 mM sodium nitrate or anaerobically with medium supplemented with 2.5 mM sodium nitrite. Data shown are means of 3 biological repeats and error bars show one standard deviation from the mean.

A
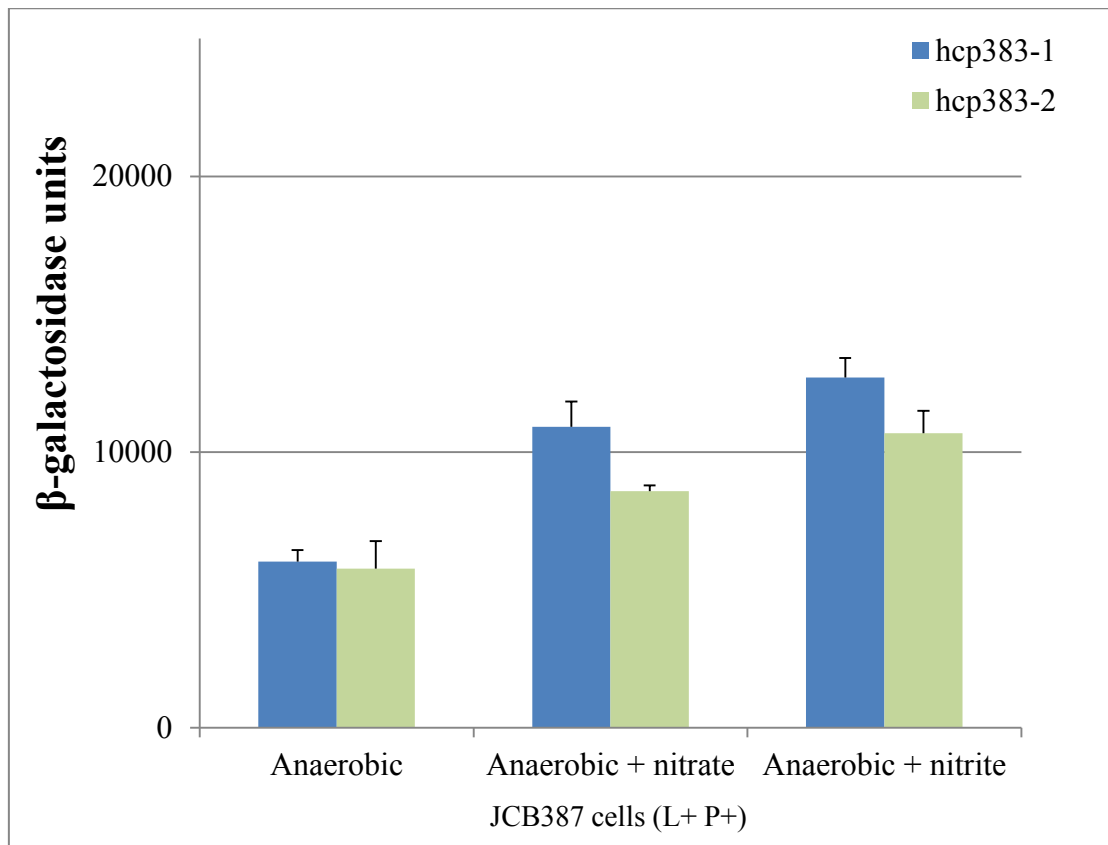
B
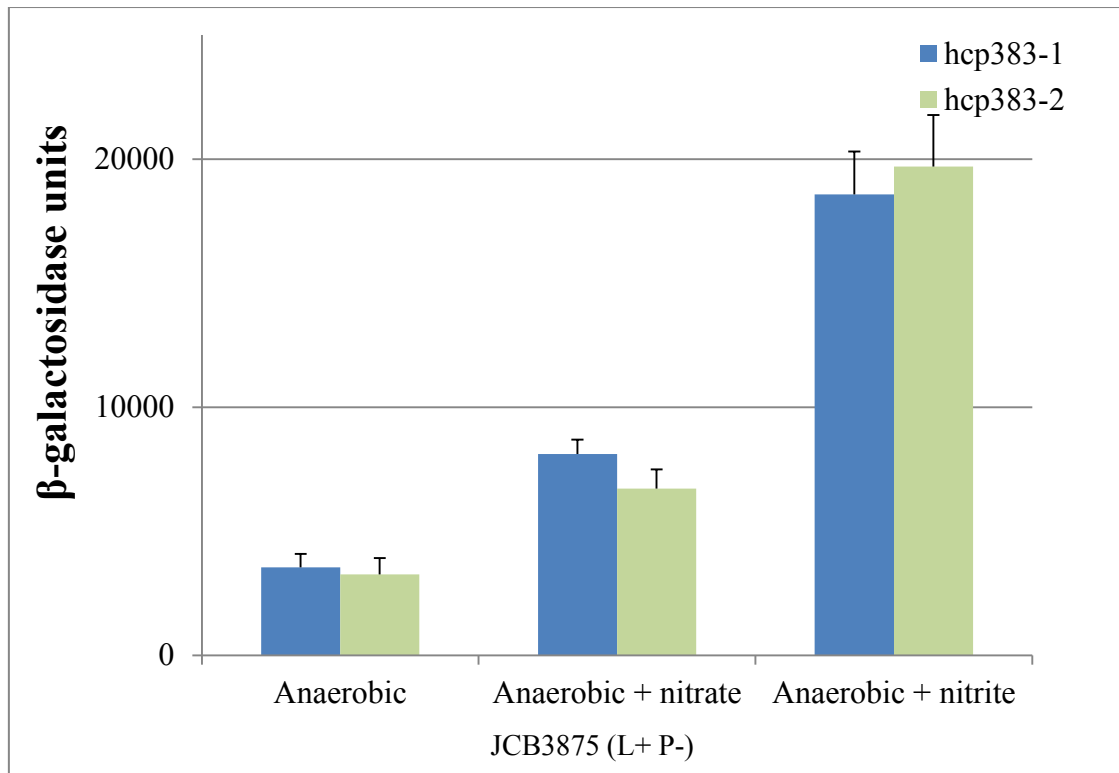
The results illustrated in Figure 3.10, panel A, show expression of the promoter::*lacZ* fusions measured in JCB3883 cells, which contain NarP but not NarL. In JCB3883 cells, levels of nitrite dependent induction are lower than in cells that contain NarL, which indicates that NarL is required for maximal activity of the *hcp* promoter. There is no significant difference between the activities of the two fusions and this indicates that the DNA site for NarL at position -104.5 plays no role in the absence of NarL.

The results illustrated in Figure 3.10, panel B, show expression of the promoter::*lacZ* fusions measured in JCB3884 cells, which lack both NarL and NarP. Both fusions show a similar pattern of expression to that observed in JCB3883 cells that express NarP but not NarL. Expression of both fusions is induced to similar levels by both nitrate and nitrite and the maximal induction seen in cells that contain NarL is not observed.

These results, taken together, show that NarL is required for maximal activation of the *hcp* promoter and this activation is independent of the DNA site for NarL at position -104.5. Hence it is possible that NarL activates *hcp* expression indirectly. Recall that NsrR strongly represses *hcp* transcription and the *hcp* promoter is induced when reactive nitrogen species accumulate in the cell. Therefore, the indirect role of NarL in the induction of the *hcp* promoter may well be because NarL plays a role in the generation of reactive nitrogen species.

To test this hypothesis, expression from the hcp383-1n::*lacZ* and hcp383-2n::*lacZ* fusions, which carry mutations in the DNA site for NsrR, was measured in JCB3884 cells that lack both NarL and NarP. Results illustrated in Figure 3.11 show that NarL is not required for maximal activation of the *hcp* promoter if the DNA site for NsrR at position +6 is mutated. As expected, removal of the DNA site for NarL has no significant effect on activity of the *hcp* promoter in cells that do not express NarL or NarP.

Figure 3.10: Activity of the *hcp* promoter and a derivative lacking the DNA site for NarL in cells that do not contain NarL

Panel A shows expression of the hcp383-1::*lacZ* and hcp383-2::*lacZ* fusions cloned in pRW50 and measured in JCB3883 cells, which do not express NarL. Cells were grown to mid exponential phase (OD_{650} 0.5-0.7) in minimal salt medium anaerobically, anaerobically with medium supplemented with 20 mM sodium nitrate or anaerobically with medium supplemented with 2.5 mM sodium nitrite. Data shown are means of 3 biological repeats and error bars show one standard deviation from the mean.

Panel B shows expression of the hcp383-1::*lacZ* and hcp383-2::*lacZ* fusions cloned in pRW50 and measured in JCB3884 cells, which do not express NarL or NarP. Cells were grown to mid exponential phase (OD_{650} 0.5-0.7) in minimal salt medium anaerobically, anaerobically with medium supplemented with 20 mM sodium nitrate or anaerobically with medium supplemented with 2.5 mM sodium nitrite. Data shown are means of 3 biological repeats and error bars show one standard deviation from the mean.

A

B

Figure 3.11: Activity of derivatives of the *hcp* promoter with mutations in the NsrR site in cells that do not contain NarL or NarP

Expression of the hcp383-1n::*lacZ* and hcp383-2n::*lacZ* fusions, which contain mutations in the DNA site for NsrR, cloned in pRW50 and measured in JCB3884 cells, which do not express NarL or NarP. Cells were grown to mid exponential phase (OD_{650} 0.5-0.7) in minimal salt medium anaerobically, anaerobically with medium supplemented with 20 mM sodium nitrate or anaerobically with medium supplemented with 2.5 mM sodium nitrite. Data shown are means of 3 biological repeats and error bars show one standard deviation from the mean.

3.10. NarL binds *in vitro* to the predicted site at position -104.5 in the *hcp* promoter

Filenko *et. al.* (2005) predicted a single DNA site for NarL at position -104.5 and although promoter truncation and mutation experiments support this conclusion, the possibility remains that NarL might bind to other sites in the *hcp* promoter region. To investigate this possibility, *in vitro* protein binding of purified NarL to the *hcp* promoter region was studied. Initially, radiolabelled hcp383-0 fragment and purified NarL protein were analysed by EMSA and results in Figure 3.12 show 3 shifts with increasing concentrations of NarL protein which confirms that observed by Filenko *et. al.* (2007). In order to explore the possibility that NarL and FNR might act collaboratively in binding to the *hcp* promoter, radiolabelled hcp383-0 fragment, a concentration of purified FNR protein that is known to bind the *hcp* promoter and increasing concentrations of purified NarL protein were used in an EMSA. Results show that the NarL-concentration dependent shifts occur at the same concentrations of NarL regardless of whether FNR is present, indicating that FNR bound to the *hcp* promoter does not enable NarL to bind to the fragment at lower concentrations.

The EMSA results do not preclude the existence of multiple DNA sites for NarL and so binding was examined with a DNase I protection assay. Results in Figure 3.13 show a single zone of protection that corresponds to the predicted site at position -104.5. It is therefore likely that the multiple shifts observed in the EMSA correspond to: initial occupation of a half-site by NarL (first shift); occupation of both half-sites by NarL (second shift); non-specific supershift (third shift).

3.11. Random mutagenesis of the *hcp* promoter region

To identify DNA elements in the *hcp* promoter that are necessary for its activity, random mutagenesis by error prone PCR was used. A reduced fragment was constructed, hcp383-1ns, to improve the efficiency of mutagenesis (Figure 3.14, panel A). The hcp383-1ns fragment, which contains the region from position -125 to +46 together with mutations in the DNA site

115
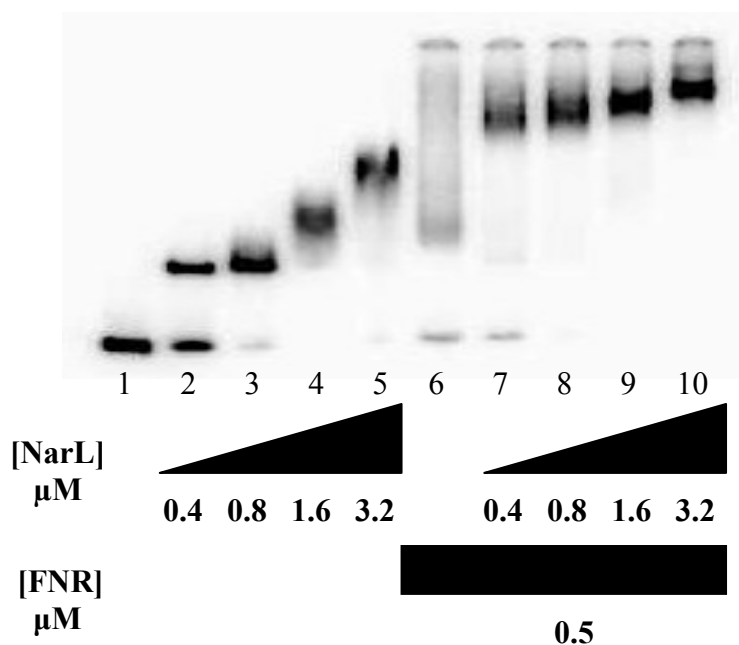
Figure 3.12: NarL and FNR bind to the *hcp* promoter fragment

The figure shows the results of an electrophoretic mobility shift assay using the end-labelled hcp383-0 promoter fragment, incubated with increasing concentrations of NarL, with or without a fixed concentration of Ala154 FNR. The concentrations of NarL used were: lanes 1 and 6, no protein; lanes 2 and 7, 0.4 µM; lanes 3 and 8, 0.8 µM; lanes 4 and 9, 1.6 µM; lanes 5 and 10, 3.2 µM. The concentration of FNR used was 0.5 µM (lanes 6-10)
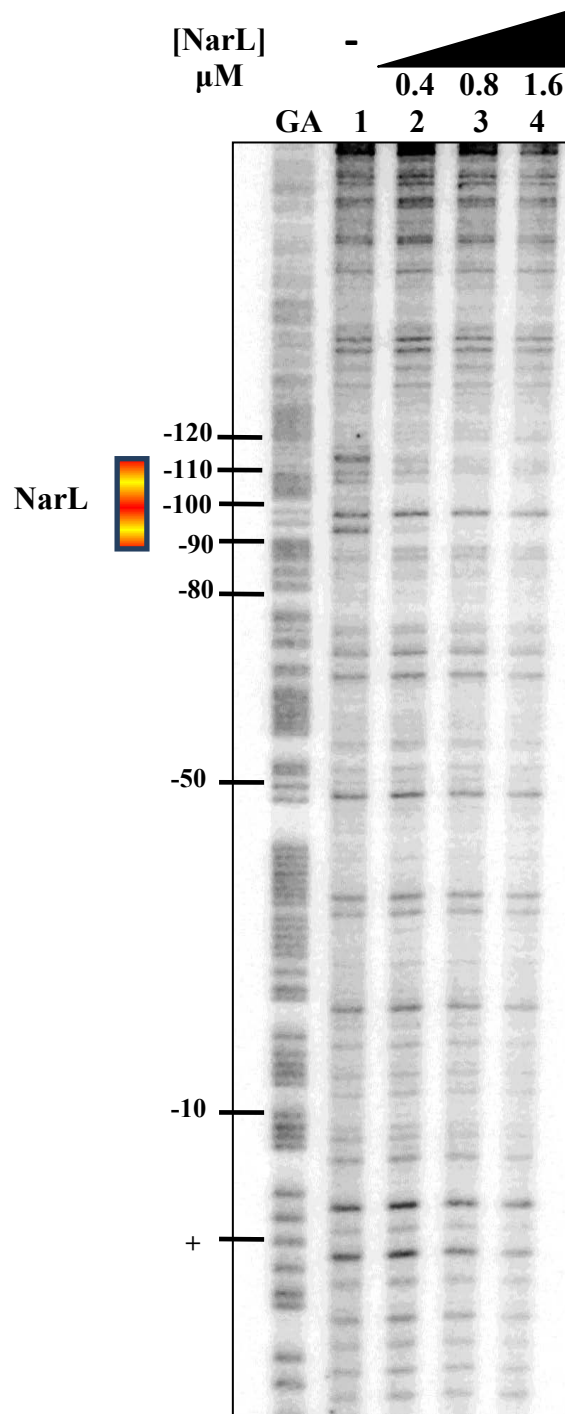
Figure 3.13: DNase I assay of NarL bound to the *hcp* promoter region

The figure shows a footprint of purified NarL binding at the *hcp* promoter. End-labelled hcp383-0 AatII-HindIII fragment was incubated with increasing concentrations of NarL and subjected to DNase I footprint analysis. The concentration of NarL was: lane 1, no protein; lane 2, 0.4 µM; lane 3, 0.8 µM and lane 4, 1.6 µM. The gel was calibrated using a Maxam-Gilbert 'G+A' sequencing reaction and relevant positions are indicated. The location of the DNA site for NarL is shown by a box.

for NsrR, was cloned into pRW50. Expression of the resulting *lacZ* fusion and the hcp383-1n::*lacZ* fusion was measured in JCB387 cells. Results illustrated in Figure 3.14, panel B, show that the promoter carried by the hcp383-1ns fragment is active but displays reduced activity when compared to that carried by the hcp383-1n fragment.

The hcp383-1ns fragment was mutated by error prone PCR and resulting fragments were cloned into pRW50. Plasmids were transformed into JCB387 cells and recovered on MacConkey-lactose plates to allow identification of Lac⁻ colonies. In an experiment performed together with a visiting project student, Gregory Farrant, over 500 transformants were screened and the plasmids of Lac⁻ colonies were isolated and the *hcp* promoter sequenced to determine mutations. The promoter activity was measured in JCB387 cells and results are shown in Figure 3.15. Seven mutations were identified as giving rise to Lac⁻ phenotypes and the mutations can be considered as falling into 2 groups: those that did or did not contain at least one mutation in the DNA site for FNR. That most mutations producing a Lac⁻ phenotype contained at least one mutation in the DNA site for FNR demonstrates the importance of that site for activation of the *hcp* promoter. Two mutant fragments contained mutations in both the DNA sites for FNR and NarL, but promoter activity was not greatly decreased in comparison to that that of fragments containing only a mutation in the DNA site for FNR. No Lac⁻ transformants were identified that contained only a mutation in the DNA site for NarL.

One transformant contained a single mutation at position -36 relative to the *hcp* transcription start site, a position that does not fall into any known motifs. The region surrounding position -36 includes an imperfect inverted repeat with the sequence GAAAAA-TCCCT-TTTATC and so may function as an activator binding site, whilst another possibility is that the mutation fell in the currently uncharacterised -35 element.

A

TACYYMTnnAKRRGTA

-120

GĀATTC TTTCTCTGCG**TAA**TACCTCTCTGGCGGTAGATCC

EcoRI   *ybjE*

NarL (104.5)

-80 **TTGATnnnnATCAA**   -60

CTGCCGCCAAAATTGCGCTAAATCAATCTCCCTTAAAGTTGCATGAAAAA

FNR (-72.5)

-40   -20   +1

AANATGCATTTnAAA

TCCCTTTTATCCCCGCGTTAAGCGTCTTAACCTTAAACATG**T**ATATTAA**G**

NsrR (+6)

TGCATNTT   +30   +50   +70

GATAACTTTAAAAGGTGTGATC**ATG**TTTTGTGTGCAAGCTT

HinDIII

B
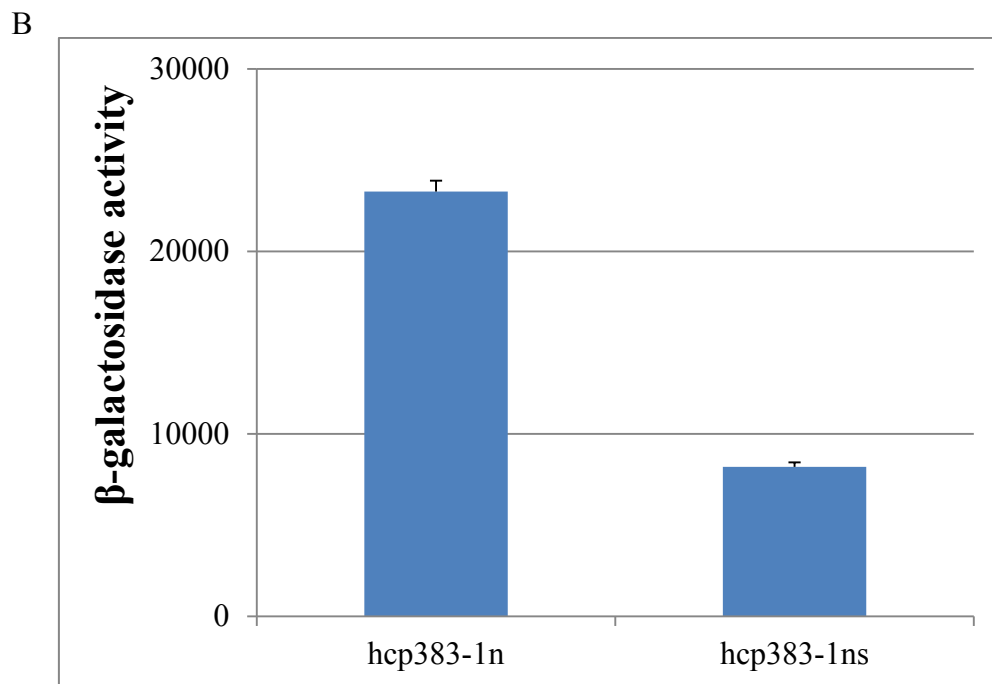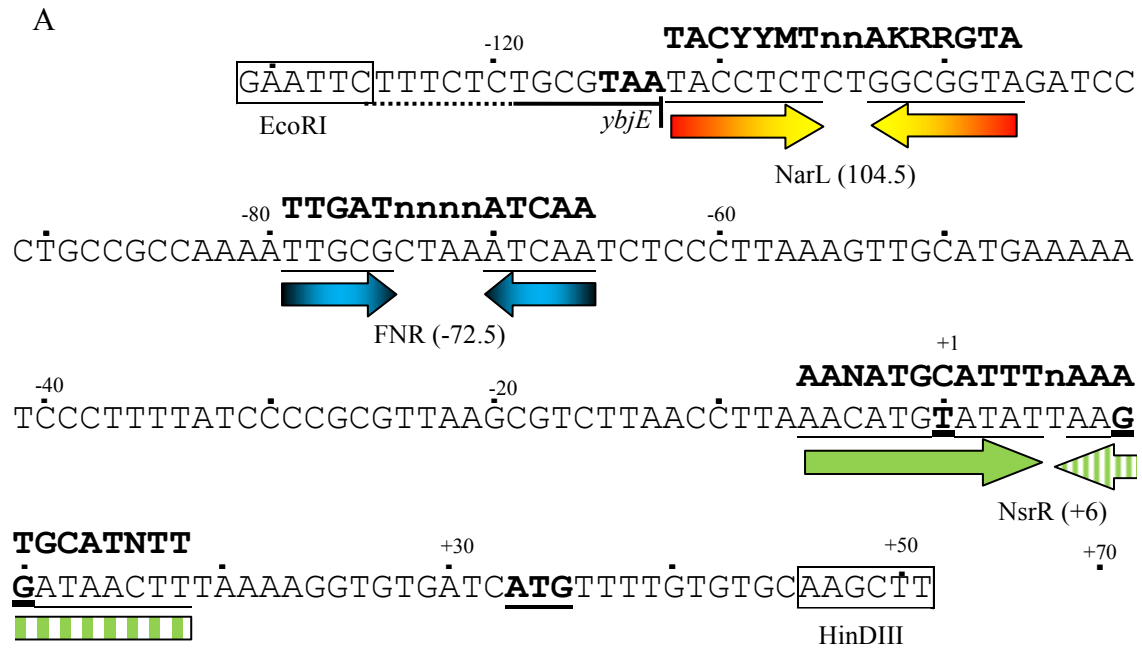
Figure 3.14: Sequence and activity of a truncated fragment containing the *hcp* promoter and mutations in the NsrR site

Panel A shows the sequence of the hcp383-1ns promoter fragment. Emboldened TAA is the stop codon for *ybjE*. DNA sites for NarL, FNR and NsrR are underlined and marked with coloured arrows underneath the sequence. The consensus binding sequences for predicted DNA sites are shown in bold text beneath the site and the positions of the centres of the sites are given in brackets beneath the arrows. The -10 and -35 elements identified in this study are underlined and labelled, as is the published start codon.

Panel B shows expression of the hcp383-1n::*lacZ* and hcp383-1ns::*lacZ* fusions, cloned in pRW50 and measured in JCB387 cells. Cells were grown in minimal salt medium anaerobically to mid-exponential phase (OD_{650} 0.5-0.7) and data shown are the mean of 3 biological repeats. Error bars show one standard deviation from the mean.

NarL (-104.5) FNR (-72.5)

TACYYMTnnAKRRGTA   -80 **TTGATnnnnATCAA**   -60   -40   Relative activity (%)

hcp383-1ns ATACCTCTCTGGCGGTAGATCCCTGCCGCCAAAATTGCGCTAAATCAATCTCCCTTAAAGTTGCATGAAAAATCCCTTTTATCC 100

hcp383-25 ATACCTCTCTGGCGGTAGATCCCTGCCGCCAAAATTGCGCTAAA**C**CAATCTCCCTTAAAGTTGCATGAAAAATCCCTTTTATCC 5.9

hcp383-33 A**C**ACCTCTCTGGCGGTAGATCCCTGCCGCCAAAATTGCGCTAAA**C**CAATCTCCCTTAAAGTTGCATGAAAAATCCCTTTTATCC 5.4

hcp383-26 ATACCTCTCTGGCGGTAGATCCCTGCCGCCAAAAT**C**GCGCTAAATCAATC**C**CCCTTAAAGTTGCATGAAAAATCCCTTTTATCC 2.3

hcp383-27 A**C**ACCTCTCTGGCGGTAGATCCCTGCCGCCAAAA**G**TGCGCTAAATC**G**ATCTCCCTTAAAGTTGCATGAAAAATCCCTTTTATCC 1.7

hcp383-28 ATACCTCTCTGGCGGTAGATCCCTGCCGCCAAAATTGCGCTAAATC**G**ATCTCCCTTAAAGTTGCATGAAAAATCCCTTTTATCC 2.4

hcp383-29 ATACCTCTCTGGCGGTAGATCCCTGCCGCCAAAATTGCGCTAAATC**G**ATCTCCCTTAAAGTTGCATG**C**AAAATCCCTTTTATCC 1.4

hcp383-32J ATACCTCTCTGGCGGTAGATCCCTGCCGCCAAAATTGCGCTAAATCAATCTCCCTTAAAGTTGCATGAAAAATCCCT**C**TTATCC 12.4

Figure 3.15: Mutations in the *hcp* promoter region that affect promoter activity

Figure shows a partial sequence of the *hcp* promoter fragment hcp383-1ns and derivatives containing randomly introduced mutations. Promoter fragments were cloned into pRW50 and activity measured in JCB387 cells. . Cells were grown in minimal salt medium anaerobically to mid exponential phase (OD$_{650}$ 0.5-0.7). Relative activity was calculated by dividing the activity of the promoter fragment in question by that of hcp383-1ns and multiplying by 100.

120

3.12. The inverted repeat near position -36 does not function as an activator binding site

To identify conserved regions of the *hcp* promoter, the sequence of the hcp383-1 fragment was aligned to the relevant region from other bacterial genomes in a study with a visiting student, Gregory Farrant. Results in Figure 3.16 show that the DNA site for NarL is poorly conserved whilst there are regions of conservation that include the DNA site for FNR, the region around position -36 and the region surrounding the transcription start site.

To investigate whether a transcription factor is binding to the inverted repeat near position -36, the inverted repeat was included on a multicopy plasmid in order to titrate out the factor. The fragment hcp383-3ns was selected as it contains few other known binding sites. JCB387 cells were used to measure the expression of the hcp383-1ns::*lacZ* fusion and the hcp383-32J::*lacZ* fusion, which has a mutation at position -36. Cells also contained either the plasmid pSR, which has a high copy number, or pSR carrying the hcp383-3ns fragment.

Results illustrated in Figure 3.17 show that there is slightly greater expression of the hcp383-1ns::*lacZ* fusion when hcp383-3ns/pSR is present in cells grown anaerobically or anaerobically with nitrite. This may indicate that the hcp383-3ns fragment contains the DNA site for a repressor, rather than an activator as hypothesised. If cells are grown anaerobically with nitrate then there is no difference between expression of the hcp383-1ns::*lacZ* fusion in cells containing hcp383-3ns/pSR and expression of the hcp383-1ns::*lacZ* fusion in cells containing pSR. The hcp383-32J::*lacZ* fusion is expressed at lower levels in all conditions and the presence of hcp383-3ns/pSR does not appear to greatly affect expression. These results do not indicate the presence of an activator binding site near position -36.

NarL (-104.5)

TACYYMTNNAKRRGTA

FNR (-72.5)

TTGATNWWNATCAA

-90　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　-50

```
ECO: TAATACCTCTCTGGCGGTAGATCCCT-----GCCGCCAAAAATTGCGCTAAATCAATCTCCCTTAAAGTTGCATGAAA
SEN: TGATACCTCTTTGGCGGTAGGCATCT-----GCCGCCAAAAATTGCGCTAAATCAATCTCCTTGCACATTATGGCAGA
STY: TGATACCTCTTTGGCGGTAGGCATCT-----GCCGCCAAAAATTGCGCTAAATCAATCTCCCTGCACATTATGGCAGA
YEN: TCACTCC--AATAGTGAATAAATTCTCATTCGCAGCAAAAGTTGCGCCAAATCAAAACAAGTTAAAGTTGCATTAAA
YPE: TCCTTCC-CTATATCGAATAAATTCTCATTCGCAGTAATAATTGCGTCAAATCAATATAAGTGAATATTGCATTAAA
YPS: TCCTTCC-CGATATCGAATAAATTCTCATTCGCAGTAATAATTGCGTCAAATCAATATAAGTGAATATTGCATTAAA
```

NsrR (+6)

AANATGCATTTNAAATGCATNTT

-30　　　　　　　　　　　-10

```
ECO: AATCCCTTTT--ATCCCCGCGTTAAGCGTCTTAACCTTAAACATGtATATTAAATATAACTTTAAAAGG---TGTGA
SEN: AATCCCTTTT--ATCCCTTTGTTCACAGGCATAATCTTAAACATGTATATTAAATATAACTTTAAAAGG---TGTGA
STY: AATCCCTTTT--ATCCCTTTGTTCACAGGCATAACCTTAAACATGTATATTAAATATAACTTTAAAAGG---TGTGA
YEN: AAAGCCTTTTTAAAACCGCTTGTGCACATCTATGCTTTTAAACACGCATTACAAATGCAACTTTAAAAGGAAAAATAA
YPE: AATACCTTTTTAAAACCGATGGCGTATCTCTATGCTCCAAAACACGCATTACAAATACAACTTTAAAAAGGAAAATAA
YPS: AATACCTTTTTAAAACCGATGGCGTATCTCTATGCTA-AAAACACGCATTACAAATGCAACTTTAAAAAGGAAAATAA
```

Figure 3.16: Phylogenetic analysis of *hcp* promoter regulatory region
The figure shows an alignment of the *E. coli* K-12 *hcp-hcr* operon promoter region sequence with corresponding sequences from related bacteria, downloaded from the xBASE database (xbase.bham.ac.uk)[29]. The different bacteria (and corresponding xbase Taxon ID) are: ECO, *E. coli* K-12 (83333); SEN, *Salmonella enterica* subspecies enterica serovar Choleraesuis str. SC-B67 (321314); STY, *Salmonella typhimurium* LT2 (99287); YEN, *Yersinia enterocolitica* (393305); YPE, *Yersinia pestis* (229193); YPS, sequence from *Yersinia pseudotuberculosis* (349747). DNA targets for NarL (and NarP), FNR and NsrR are indicated by shaded horizontal inverted arrows. For each target, the consensus binding sequence is shown in bold, aligned with the actual sequence, and the centre of the site is noted. In the consensus sequences, Y denotes either C or T, M denotes A or C, K denotes G or T, R denotes A or G, W denotes A or T, and N denotes any base. The transcription start point in the *E. coli* K-12 sequence is shown in lower case. Bases that are identical in each strain's genome are highlighted with black shading.
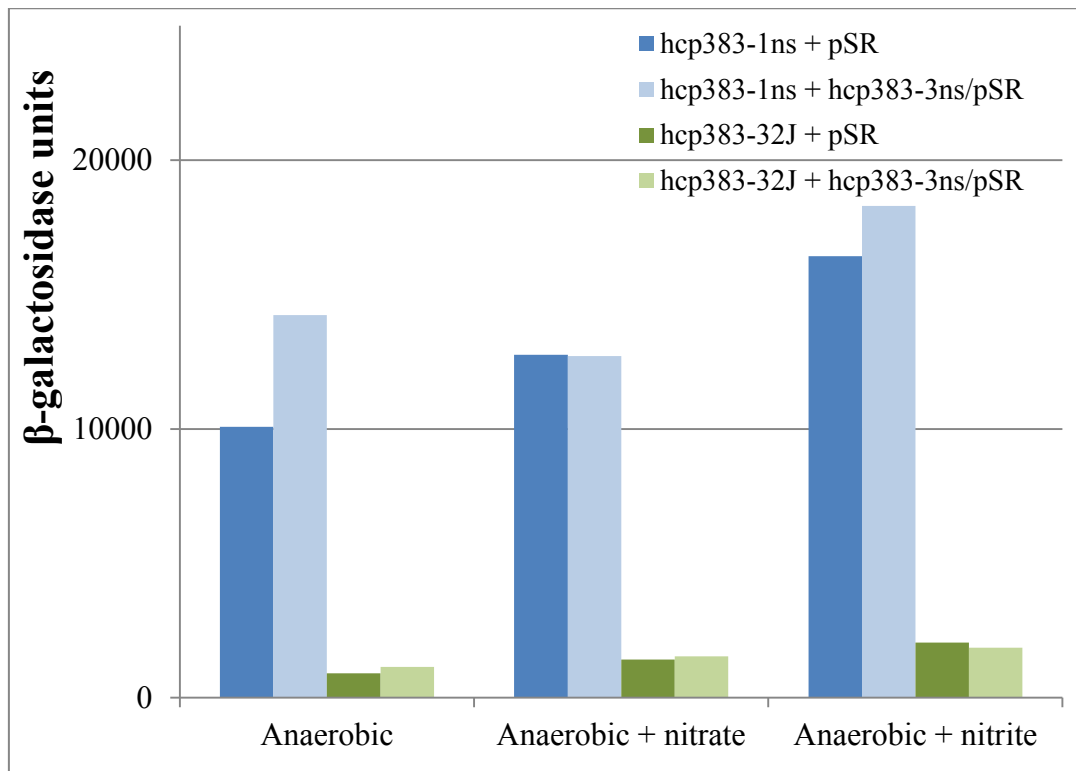
Figure 3.17: Titration of factors regulating expression of the *hcp* promoter

Expression of the hcp383-1ns::*lacZ* and hcp383-32J::*lacZ* fusions, cloned in pRW50 and measured in JCB387 cells. Cells also contained either the plasmid pSR or pSR carrying the truncated *hcp* promoter fragment hcp383-3ns. Cells were grown in minimal salt medium to mid exponential phase (OD_{650} 0.5-0.7) anaerobically, anaerobically with medium supplemented with 20 mM sodium nitrate or anaerobically with medium supplemented with 2.5 mM sodium nitrite.

3.13. Identification of the *hcp* promoter-35 and -10 elements

To uncover the role of the conserved region around position -36, the -35 element was characterised. The -35 element was predicted on the basis of distance from the published transcription start site and similarity to the consensus σ^{70} -35 element (TTGACA). Mutations designed make the sequence more or less resemble the consensus sequence were introduced into the hcp383-1ns fragment and resulting fragments were then cloned into pRW50.

The results in Table 3.1 show that mutations that alter the sequence of the predicted -35 element away from the consensus -35 sequence cause greatly reduced promoter activity. These results agree with those of a similar study by Kobayashi *et al.* (1990). The P35G mutation that alters the sequence of the predicted -35 element to better resemble the consensus sequence produces much higher promoter activity. These results indicate that the region between position -37 and -32 does function as a -35 element. Site directed mutagenesis was also used to construct a derivative of the hcp383-1ns fragment with a consensus -35 element but was unsuccessful as all resulting fragments contained at least one secondary mutation.

The -10 element was predicted on the basis of distance from the reported transcription start site and similarity to the consensus sequence for the -10 element, TATAAT. Site directed mutagenesis was used to introduce mutations as before and results in Table 3.1 show that mutations that alter the sequence of the predicted -10 element away from the consensus sequence result in greatly reduced promoter activity. These results support the hypothesis that the region between position -14 and -9 functions as a -10 element.

3.14. Transcriptional fusions of the *hcp* promoter to *lacZ* are highly active

The hcp383-1ns::*lacZ* fusion, which was constructed by truncation of the downstream region of the hcp383-1n fragment, is expressed 2-fold less than is the hcp383-1n::*lacZ* fusion. This observation raises the possibility that the sequence downstream of the transcription start

-35 Element (TTGACA)	-10 Element (TATAAT)	Mutation	Activity	
			Aerobic	Anaerobic
-37　　-32	-14　　-9			
ṪTTTAṪ	ṪAACCṪ	none	2.9	100.0
Mutations in the -35 element				
TT**G**TAT	TAACCT	-35G	20.0	246.0
T**C**TTAT	TAACCT	-36C	1.1	13.0
CTTTAT	TAACCT	-37C	0.4	6.1
ATTTAT	TAACCT	-37A	0.4	3.3
Mutations in the -10 element				
TTTTAT	TAACC**A**	-9A	0.4	6.2
TTTTAT	TAACC**C**	-9C	0.4	2.8
TTTTAT	T**G**ACCT	-13G	0.5	2.9
TTTTAT	**A**AACCT	-14A	0.2	15.0
TTTTAT	**G**AACCT	-14G	0.4	3.6
TTTTAT	**C**AACCT	-14C	0.4	8.7

Table 3.1: Effect of point mutations in the *hcp* promoter region on activity

β-Galactosidase expression was measured in JCB387 cells carrying pRW50 containing the hcp383-1ns promoter fragment with single point mutations at different locations in the proposed -35 or -10 hexamer elements. The different base changes are shown in the first two columns and the consensus -35 and -10 hexamers for *E. coli* promoters are given at the head of each column. Cells were grown aerobically or anaerobically at 37ºC in minimal medium. Each measurement is the average of three biological replicates and is expressed as a percentage of the measured β-Galactosidase activity during anaerobic growth with the starting hcp383 1ns fragment.

site might contribute to the high activity of the *hcp* promoter. Also, the *lacZ* fusion vector, pRW50, uses the Shine-Dalgarno sequence of the inserted fragment but has a fragment of the *trpBA* operon between the inserted fragment and the *lacZ* gene, making it possible that the *trpBA* operon is interfering with expression of *lacZ*. To address these concerns, the hcp383-1ns fragment was cloned into pRW224, a derivative of pRW50 that lacks the *trpBA* operon, as a transcriptional fusion so that *lacZ* is expressed from the plasmid encoded *galK* Shine-Dalgarno sequence. A further truncated derivative of the hcp383-1ns fragment was constructed to remove the Shine-Dalgarno box of *hcp*. The resulting fragment, hcp383-1nvs was cloned into both pRW50 and pRW224.

JCB387 cells were grown anaerobically and contained fragments in either pRW50 or pRW224 vectors. The expression of the promoter::*lacZ* fusions was measured and results in Figure 3.18 show that, when both fusions are carried by pRW50 as a translational fusion, the expression of the hcp383-1nvs::*lacZ* fusion was 3-fold less than was that of the hcp383-1ns::*lacZ* fusion. When fragments are carried by pRW224 as a transcriptional fusion, the expression of the hcp383-1ns::*lacZ* fusion was roughly 10-fold less than when it was expressed as a translational fusion from pRW50. The expression of the hcp383-1nvs::*lacZ* fusion was 3-fold greater when it was expressed as a transcriptional fusion, which means that the hcp383-1nvs::*lacZ* fusion is as active when using the *galK* Shine-Dalgarno element as is the hcp383-1ns::*lacZ* fusion when using the *hcp* Shine-Dalgarno. This demonstrates that the high activity observed from the *hcp* promoter is not due to using a distinct Shine-Dalgarno element and start codon.

3.15. Discussion

3.15.1. Unusual promoter organisation of the hcp promoter

The *hcp* promoter is highly active when optimally induced but analysis of the DNA sequence suggests a low activity promoter. The core promoter elements of *hcp* bear poor
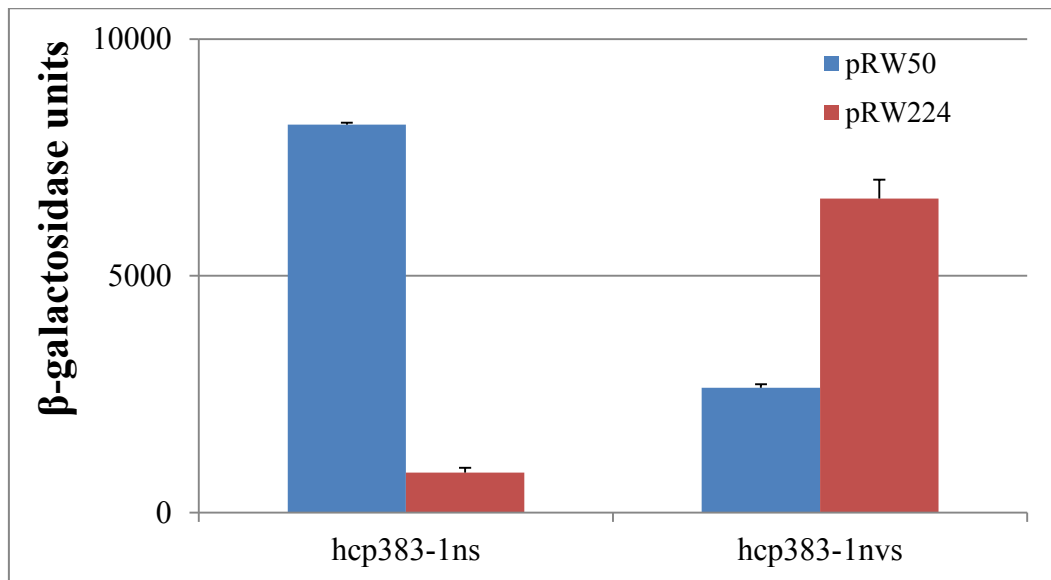
Figure 3.18: Expression of *hcp* promoter::*lacZ* fusions expressed as either a translational or a transcriptional fusion

Expression of the hcp383-1ns::*lacZ* and hcp383-1nv:*lacZ* fusions, cloned in either pRW50 or pRW224 and measured in JCB387 cells. Cells were grown in minimal salt medium to mid exponential phase (OD$_{650}$ 0.5-0.7) anaerobically and data are the mean of 3 biological replicates. Error bars show one standard deviation from the mean.

resemblance to the consensus sequences for such elements, which normally indicates a weak promoter. FNR is an effective activator of promoter activity that can overcome imperfect core promoter elements. However, at the *hcp* promoter, FNR acts from a class I position rather than a class II position, which typically results in weaker activation (Wing *et al.*, 1995). Finally, the DNA site for NarL is in a position where it could contribute to activation but results show that it plays a minor role and is unnecessary in some circumstances.

3.15.2. Direct vs. indirect activation of the hcp promoter

The *hcp* promoter is activated by FNR in response to anaerobiosis, which is crucial for maximal activity. FNR activates directly from its site at position -72.5. Nitrate induces the *hcp* promoter and the DNA site for NarL at position -104.5 plays a minor role. When the site is deleted, there is still a low level of nitrate dependent induction, which suggests that NarL activates *hcp* promoter activity both directly and indirectly. Nitrite induces activity of the *hcp* promoter maximally and this induction is completely independent of the DNA site for NarL at position -104.5. However, maximal activity of the *hcp* promoter is dependent on NarL being expressed by the cell, which corroborates the hypothesis that NarL activates *hcp* promoter activity both directly and indirectly.

The NarL dependence of the the *hcp* promoter is relieved if NsrR is unable to repress activity by binding to its site at position +6. This suggests that NarL may activate by preventing repression by NsrR and it is possible that NarL directly activates by interfering with NsrR binding to the promoter.

The mechanism by which NarL activates indirectly has recently been confirmed, namely that NarG is the major producer of NO in *E. coli* (Claire Vine and Sukhjit Purewal, personal communication). Note that expression of *narG* is strongly induced by NarL (Walker and DeMoss, 1994). In cells lacking NarL, *narG* would not be expressed as highly and so RNS would not accumulate in the cell. Therefore, NsrR would be able to repress *hcp* promoter

activity. If NsrR is unable to bind to the *hcp* promoter then the RNS generated by NarG will not be important.

Another hypothesis to explain NarL-dependent activation by nitrite is that there may be other DNA sites for NarL in the *hcp* promoter region that are currently unknown. NarL bound to such sites may directly prevent NsrR repressing activity. This hypothesis is at odds with reports that NarL is less active in the presence of nitrite than is NarP. The expectation is that NarP, rather than NarL, would activate directly in response to nitrite. Sequence analysis reveals that there are some regions around the transcription start site (where the DNA site for NsrR is located) that partially resemble the NarL consensus. However, no protection of these sites by NarL was observed by DNase I protection assay.

3.15.3. *Factors determining high activity of the hcp promoter*

In this study, the factors crucial for high activity of the *hcp* promoter were found to be: the core promoter elements; activation by FNR from its site at position -72.5; and the absence of repression by NsrR. However, when the region downstream of the *hcp* transcript start was deleted, around 50% of promoter activity was lost. To explain this observation, the secondary structure was predicted for the mRNA that would result from both the longer hcp383-1n fragment and the truncated hcp383-1ns fragment. The predicted structures shown in Figure 3.19 show that the longer hcp383-1n fragment potentially produces mRNA that adopts a hairpin structure and may present the Shine-Dalgarno to a ribosome. Such structures are reported to amplify translation. The truncated fragment is not predicted to give rise to such a structure, which could explain the reduced expression of hcp383-1ns::*lacZ* fusions.

It is possible that the *hcp* promoter is moderately active and that activity is driven by non-consensus but functional core promoter elements and activation by FNR from a non-optimal position. Expression of HCP might then be amplified by the *hcp* mRNA secondary structure. To probe this possibility it would be necessary to study further the hcp383-1nvs fragment
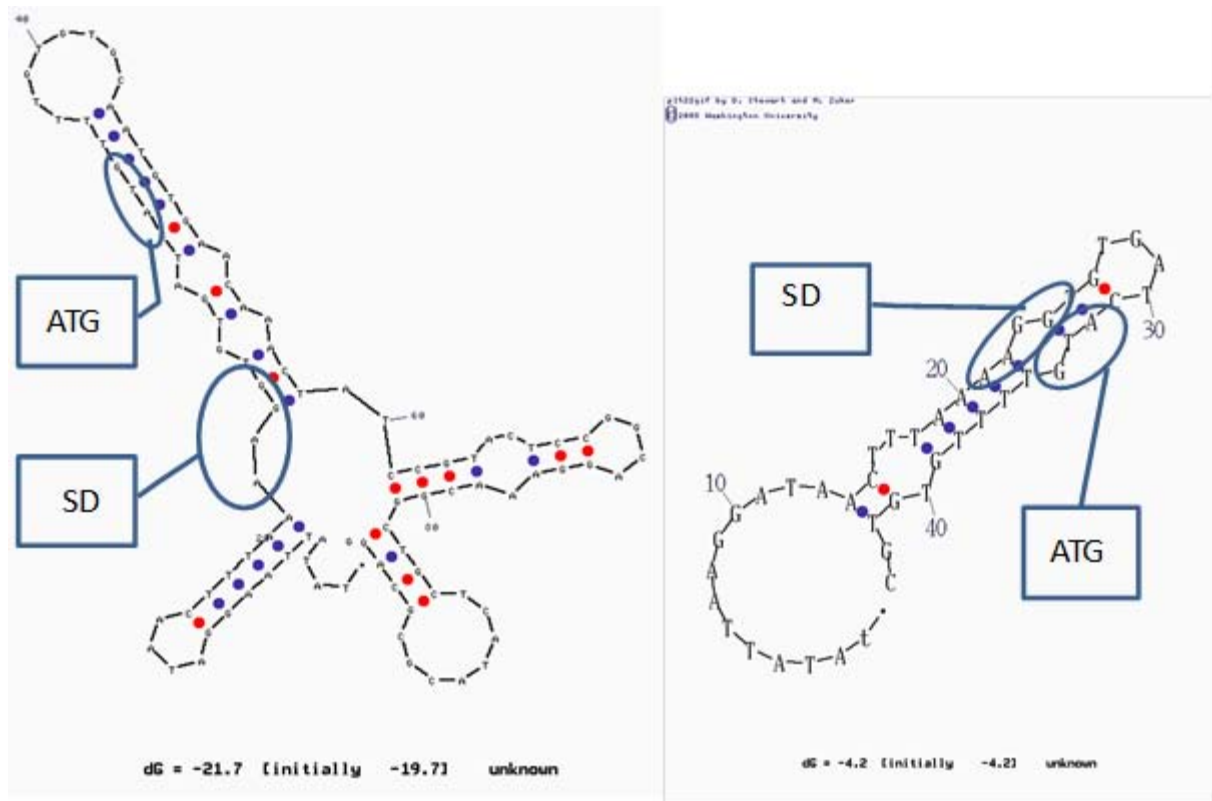
Figure 3.19: Predicted secondary structure of RNA resulting from hcp383-1n and hcp383-1ns

Figure shows predicted secondary structure of *hcp*383-1n and *hcp*383-1ns mRNA. Structures were predicted using mfold (Zuker, 2003). Blue dots denote predicted interaction between bases while red dots denote mismatches.

fused to *lacZ* as a transcriptional fusion or to study *in vitro* transcription. To investigate the predicted secondary structure of the *hcp* mRNA, mutations could be introduced to disrupt the stem of the predicted hairpin. Complementary mutations on the other side of the stem should then recover expression levels.

There remains the possibility that a 'mystery activator' may be responsible for the high activity observed of the *hcp* promoter and that there are DNA sites for it in the *hcp* promoter region that have yet to be identified. One method of identifying such activators is DNA sampling, a technique where the DNA fragment of interest is isolated from cells and analysed by mass spectroscopy to identify bound proteins (Butala *et al.*, 2009). This technique could be applied to determine if there are any factors acting at the *hcp* promoter that have not been previously detected.

3.15.4. Importance of the region near position -36 of the hcp promoter

The region near position -36 of the *hcp* promoter was identified as important for promoter activity by error prone PCR. Further study demonstrated that mutation of position -36 affects *hcp* promoter activity by disrupting the -35 element. However, bioinformatics reveals that the region upstream of and including position -36 is well conserved in other bacteria, which indicates that it may play a role in regulation of *hcp* expression. An explanation for the conservation of this region may be found in the work of Partridge *et al.* (2009), in which a bioinformatics analysis of NsrR sites predicted a second site in the *hcp* promoter from position -56 to -34. The results of this study corroborate the hypothesis of a second, weaker, DNA site for NsrR in the *hcp* promoter.

The EMSA with NsrR-containing cell lysate shows that NsrR binds weakly to an *hcp* promoter fragment with mutations in the site at position +6. Taken together, these results can be interpreted in two ways: either the mutations in the DNA site for NsrR do not entirely dissuade NsrR from binding to that site, or that there is a second, low-affinity, DNA site for

NsrR. Partridge *et al.* (2009) report that NsrR is able to bind both 11 bp half-sites and 11-1-11 inverted repeats. The mutations introduced into the DNA site for NsrR at position +6 of the *hcp* promoter do not affect the upstream half of the site, meaning it is possible that NsrR still binds weakly to the promoter. Finally, *hcp* promoter activity is increased by introducing the hcp383-3ns fragment on a multicopy plasmid that contains the predicted upstream DNA site for NsrR and mutations in the site at position +6. This indicates that a repressor is being titrated out, which supports the existence of a second site for NsrR

3.15.5. *The role of HCP*

An aim of this study was to better our understanding of the regulation of *hcp* expression in the hope of clarifying the role of HCP in the cell. HCP is produced optimally in anaerobic conditions with nitrite present, although nitrate also induces expression. High activity is absolutely dependent on FNR being bound to its site and the inactivation of NsrR. NarL is necessary for high levels of expression although it appears that NarL merely indirectly produces the RNS that inactivate NsrR. Work by Almeida *et al.* (2006) indicates that *hcp* expression is repressed by the transcription factor OxyR, which is inactivated by hydrogen peroxide. The predicted DNA site for OxyR is between positions -105 and -65, which overlaps the DNA sites for NarL and FNR. Unfortunately, Almeida and colleagues do not report quantitative values for expression, which complicates comparison and analysis. However, their and my results show that anaerobiosis induces activity of the *hcp* promoter in the absence of hydrogen peroxide, indicating that FNR is able to displace OxyR bound to its DNA site.

Almeida *et al.* (2006) suggested a role for HCP in defence from reactive oxygen species and reported that HCP reduces hydrogen peroxide. It is possible that the role of HCP is to protect the cell from ROS generated by Hmp, the primary aerobic NO reductase. Like expression of *hcp,* expression of *hmp* is repressed by NsrR so that in the presence of RNS,

hmp expression is up-regulated. Despite detoxifying NO, Hmp presents a risk to the cell because in the presence of oxygen, it produces ROS (Gilberthorpe *et al.*, 2007). Expression of *hmp* is therefore tightly regulated to ensure low expression in the absence of RNS, but high activity in the presence of RNS. It is plausible that bacteria would have evolved a protein to mitigate this risk, and it would be logical that expression of said protein would be up-regulated in response to reactive nitrogen species. However, expression of *hcp* and *hmp* is differentially regulated by FNR. Expression of *hcp* is activated by FNR, while expression of *hmp* is repressed by FNR. If the role of HCP were to protect the cell from ROS generated by Hmp in aerobic conditions, then *hcp* would be expected to be expressed aerobically. Further study could examine the interplay between OxyR and FNR.

Activity of the *hcp* promoter is maximal anaerobically and in the presence of RNS, which strongly suggests that HCP protects the cell from reactive nitrogen species that occur anaerobically, possibly those that result from nitrite reduction. It is perhaps telling that *hcp* promoter activity is low in a *narL* mutant, despite the observation that the DNA site for NarL in the *hcp* regulatory region is unnecessary. It is curious that *hcp* promoter activity is activated by FNR because FNR is known to be inactivated by RNS (Pullan *et al.*, 2007). This suggests that *hcp* expression may be up-regulated in response to a gradual increase of RNS, so that there is still functional FNR in the cell. This could be examined by studying the relative sensitivity of *hcp* expression to anaerobiosis and the presence of RNS. Alternatively, positive regulation of *hcp* expression by FNR may introduce a 'time delay'. Figure 3.20 depicts this model, in which a strong burst of RNS would initially inactivate FNR and NsrR. In such conditions, expression of the iron-sulphur cluster repair protein YtfE is up-regulated (Overton *et al.*, 2008). YftE then repairs the iron-sulphur cluster of FNR, perhaps damaging itself in the process. As FNR becomes re-activated, *hcp* expression would be up-regulated and HCP may then repair damaged YtfE. Eventually NsrR would be repaired and *hcp* expression shut down.

For this model to be true, YtfE would need to repair the iron-sulphur cluster of FNR before that of NsrR, which could be tested experimentally.

Finally, HCP is known to possess 2 different iron-sulphur clusters, which may indicate dual function. It is possible that HCP has a role in defence from both reactive oxygen species and reactive nitrogen species. It may deal with either threat when they occur in the cell naturally or may have a role in defence from the host immune system. ROS and RNS are known to be used by the host immune system to kill bacteria. If HCP is involved in protection from macrophages, its role must not be crucial, as an *hcp* mutant of *S. enterica* does not appear to be adversely affected in assays with immune cells (Kim *et al.*, 2003).

The regulation of *hcp* expression is reasonably complex, with multiple factors regulating, and can vary from low expression to very high expression depending on the stimuli. This makes assigning function particularly difficult, but suggests that it is important for the cell to have the right amount of HCP at the right time. It should be noted that assigning function based on regulation can be flawed, for example, the function of the Ogt protein as a DNA repair enzyme would not be inferred from regulation of its expression by nitrate (covered in Chapter 4). The function of HCP will therefore likely be solved by biochemical studies of the active protein and by studying the phenotype of *hcp* knockout cells. The evidence from this work and others does however support the hypothesis that HCP is involved in defence from reactive species of some description and so study should focus on roles related to defence from such threats.
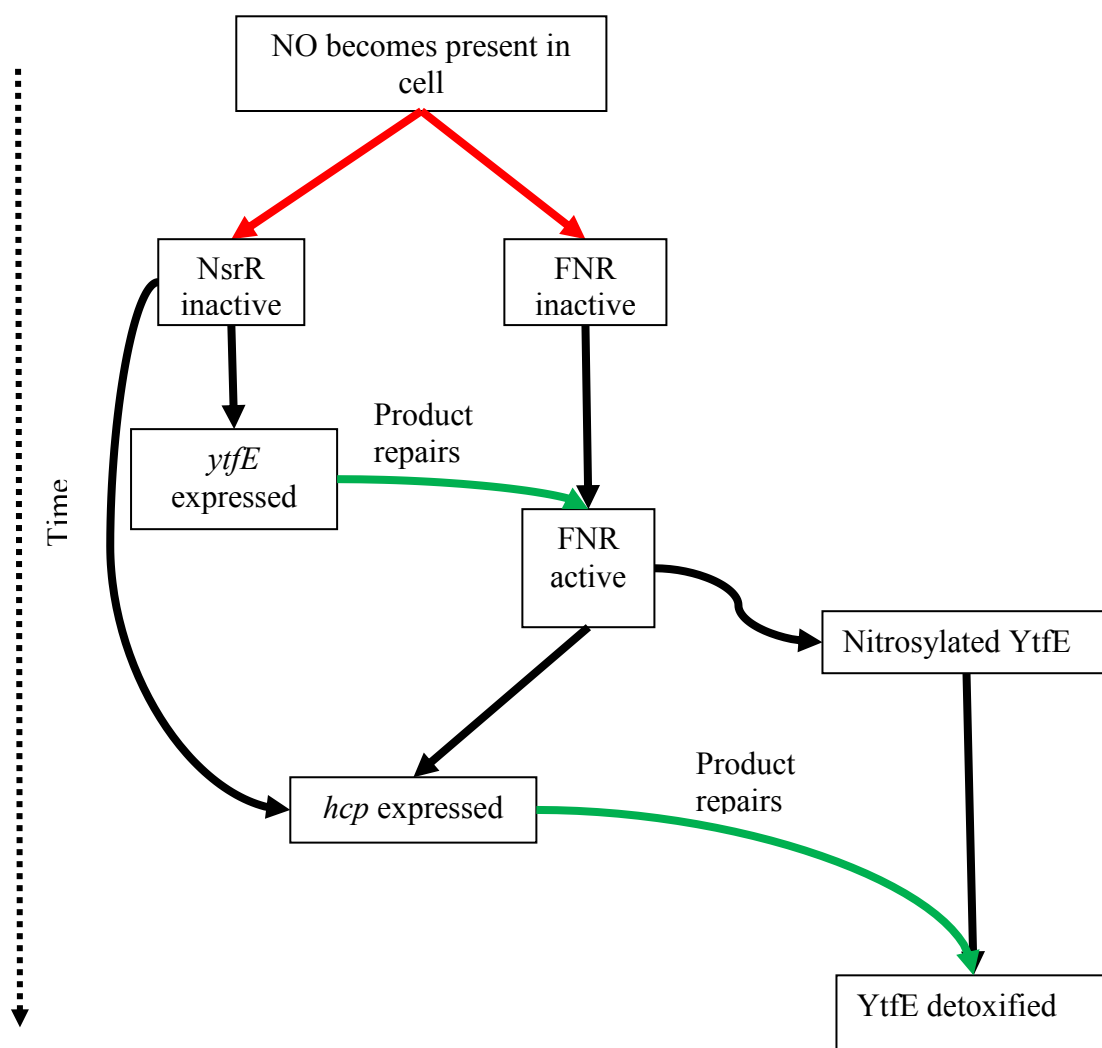
Figure 3.20 Model of regulation of *hcp* expression

The figure is a diagram showing a proposed model for the regulation of *hcp* expression. The figure is organised by the sequence of events from an initial pulse of NO. Red arrows indicate a disruptive effect and green arrows indicate a reparative effect. Black arrows indicate a general progression, i.e. "leads to". Dashed black arrow indicates the order of events temporally.

Chapter 4: Regulation of the *E. coli ogt* promoter

4.1. Introduction

In the absence of oxygen, *E. coli* is able to utilise alternative electron acceptors. The most efficient of these acceptors is nitrate, followed closely by nitrite (Richardson *et al.* 2000) . *E. coli* has evolved systems to reduce nitrate and nitrite and to deal with the by-products of reduction, RNS. The major transcriptional regulators in response to nitrate and nitrite, NarL and NarP, are able to up or down regulate transcription of relevant operons. However, the majority of promoters that are activated by NarL and NarP are also activated by FNR, the master anaerobic sensor / regulator. This is logical as oxygen is a more powerful oxidiser than is nitrate and so proteins to process nitrate would only be required in the absence of oxygen. However, NarL and NarP are members of the LuxR protein family, other members of which are able to activate transcription by contacting polymerase (Choi and Greenberg 2001). Therefore the question has arisen, are NarL and NarP able to activate transcription independently of FNR?

The work of Constantinidou *et al.* (2006) identified several genes that were upregulated by NarL / NarP in Δ*fnr* cells. Further investigation by Lin *et al* (2008) and Squire *et al.* (2009) confirmed that the *yeaR-yoaG* and *ogt* operons are activated by NarL independently of FNR. The *yeaR-yoaG* operon encodes genes of currently unknown function, while *ogt* encodes an O6-alkylguanine DNA methyltransferase, the function of which is to repair DNA damage such as that caused by RNS.

The *yeaR* and *ogt* promoters both contain DNA sites for NarL, which were confirmed by Squire *et al.* (2009) (see Figure 4.1). There is a single DNA site for NarL at the *yeaR* promoter, which is centred at position -43.5 relative to the *yeaR* transcription start site, which was profiled using 5′ RACE (Lin *et al.* 2008). At the *ogt* promoter there are two DNA

Figure 4.1: Diagrams and partial sequences of the yeaR100 and ogt100 fragments

Panel A shows sketch diagrams of the yeaR100 and ogt100 DNA fragments, which contain the *E coli yeaR* and *ogt* promoters respectively. The capital E and H at the ends of the fragments denote EcoRI and HindIII restriction sites. The mapped transcription start site is denoted with a bent arrow and "+1". The DNA sites for NarL are indicted with inverted arrows, which are labelled, and the central position of the site is given beneath the arrows relative to the transcription start site. A position ending in .5 denotes that the centre falls between two positions.

Panel B is reproduced from Squire *et al.* (2009) and shows a partial sequence of the yeaR100 fragment from positions −64 to +86 with respect to the transcription start site (+1). The −10 hexamer element is boldface and underlined, the transcription start point is designated by a lower-case letter and a bent arrow, and the translation start ATG codon is underlined. The DNA sites for NarL and NsrR are identified by shaded arrows and a rectangle respectively, and the centre of each site is numbered with respect to the transcription start site.

Panel C is reproduced from Squire *et al* (2009) and shows a partial sequence of the ogt100 fragment, from position −106 to +57 with respect to the transcription start site (+1). Then −10 hexamer element is emboldened and underlined. The transcription start site is designated by a lower-case letter and a bent arrow, and the translation start ATG codon is underlined. The DNA sites for NarL are identified by shaded arrows and the centre of each site is numbered with respect to the transcription start site.

138

A

yeaR100 fragment

NarL

E ──────────────────────────── H

+1

-43.5

ogt100 fragment

NarL   NarL

E ──────────────────────────── H

+1

-78.5   -45.5

B

yeaR100 fragment

NarL   **NsrR**

-60   -40   -20

TGGTGCTAAAAAGTAACCAATAAATGGTATTTAAAATGCAAATTATCAGG

(-43.5)   (-32)

CG**TACCCT**GAAACG**g**CTGGAATAAACCGTTTTCAGCGCATTCACCGAAGG

(-10)   +1   +20

+40   +60   +80

AGGGAAAAGGATGCTTCAAATCCCACAGAATTATATTCATACGCGCTCAA

C

ogt100 fragment

NarL I

-100   -60

AAACCAGACGCGAAACTGGGTACTTACTATTCGTTAGTCTTGCCCTATCCACT

(-78.5)

NarL II   -20

TATCTTTTTGGTGGTATGGCTGCTGATGTTGCTGGCGTGG**TATCTT**GTCGGTC

(-45.5)   (-10)

+20   +40

tGCCGATAGGTCCGGGTATTTACCCACGTTTGTCTTAAGAGAGAACGGATG

+1

139

sites for NarL, which are centred at positions -78.5 and -45.5 relative to the *ogt* transcription start which had been mapped using primer extension by Potter *et al.* (1987). The organisation of DNA sites for NarL suggests that NarL activates the *yeaR* promoter by a class II mechanism and the *ogt* promoter by a class III mechanism. Squire *et al.* (2009) confirmed that NarL was required for activation of the *yeaR* promoter and that it bound to the predicted site. They also investigated the *ogt* promoter and some of their results are reproduced in Figure 4.2.

Figure 4.2, panel A, shows that expression of an *ogt* promoter::*lacZ* fusion was induced by nitrate in a NarL-dependent fashion. Figure 4.2, panel B shows that results of an EMSA with NarL and radiolabelled DNA fragments containing the *ogt* promoter. NarL was seen to bind the *ogt* promoter region at higher concentrations only. By comparison, NarL bound to the *yeaR* promoter at much lower concentrations, which suggests that the single DNA site for NarL is a better target for NarL than are either of the sites in the *ogt* promoter region (Squire *et al.* 2009). Figure 4.2, panel C, shows the results of a DNase I protection assay with NarL and a radiolabelled DNA fragment containing the *ogt* promoter. NarL was seen to protect 2 regions of the fragment that correspond to the DNA sites for NarL predicted by sequence analysis. Finally, Figure 4.2, panel D, shows the expression of an *ogt* promoter::*lacZ* fusion in JCB387 cells and a Δ*fnr* derivative. The results confirmed that the presence or absence of FNR do not affect activity of the *ogt* promoter.

Squire *et al.* (2009) also reported that nitrate induction of the *ogt* promoter was prevented if either of the DNA sites for NarL were mutated to prevent NarL binding. This indicated that, at the *ogt* promoter, NarL activates by a cooperative class III mechanism, as both DNA sites are required.

The aim of this work was to continue the work of Squire *et al.* (2009) and to study further FNR-independent activation by NarL. A range of biochemical techniques were used to

Figure 4.2: NarL regulates the *E coli ogt* promoter

Reproduced from Squire *et al* (2009).
Panel A shows the measured β-Galactosidase activities in JCB387 and JCB3883 (Δ*narL*) cells carrying pRW50, containing the ogt100 promoter fragment. Cells were grown in minimal medium and, where indicated, sodium nitrate or nitrite was added to a final concentration of 20 and 2.5 mM respectively. β-Galactosidase activities are expressed as nmol of ONPG hydrolysed/min per mg of dry cell mass.

Panel B shows EMSA experiments with purified NarL protein. End-labelled ogt100 promoter fragment was incubated with increasing concentrations of NarL and protein–DNA complexes were separated by PAGE. The concentration of NarL was: lane 1, no protein; lane 2, 0.2 μM NarL; lane 3, 0.8 μM NarL; lane 4, 1.6 μM NarL; lane 5, 3.2 μM NarL.

Panel C shows an in vitro DNase I footprint experiment with purified NarL. End-labelled ogt100 AatII– HindIII fragment was incubated with increasing concentrations of NarL and subjected to DNase I footprint analysis. The concentration of NarL was: lane 1, no protein; lane 2, 0.4 μM; lane 3, 0.8 μM; lane 4, 1.6 μM; lane 5, 3.2 μM. The gel was calibrated using a Maxam–Gilbert 'G+A' sequencing reaction and relevant positions are indicated. The location of NarL-binding sites is shown by boxes and a hypersensitive site due to NarL binding is marked with a star.

Panel D shows the measured β-Galactosidase activities in JCB387 and a Δ*fnr* derivative, carrying pRW50 containing the ogt100 promoter fragment. Cells were grown in minimal salt medium and, where indicated, 20 mM sodium nitrate was added.

A

B

C *ogt100*

D
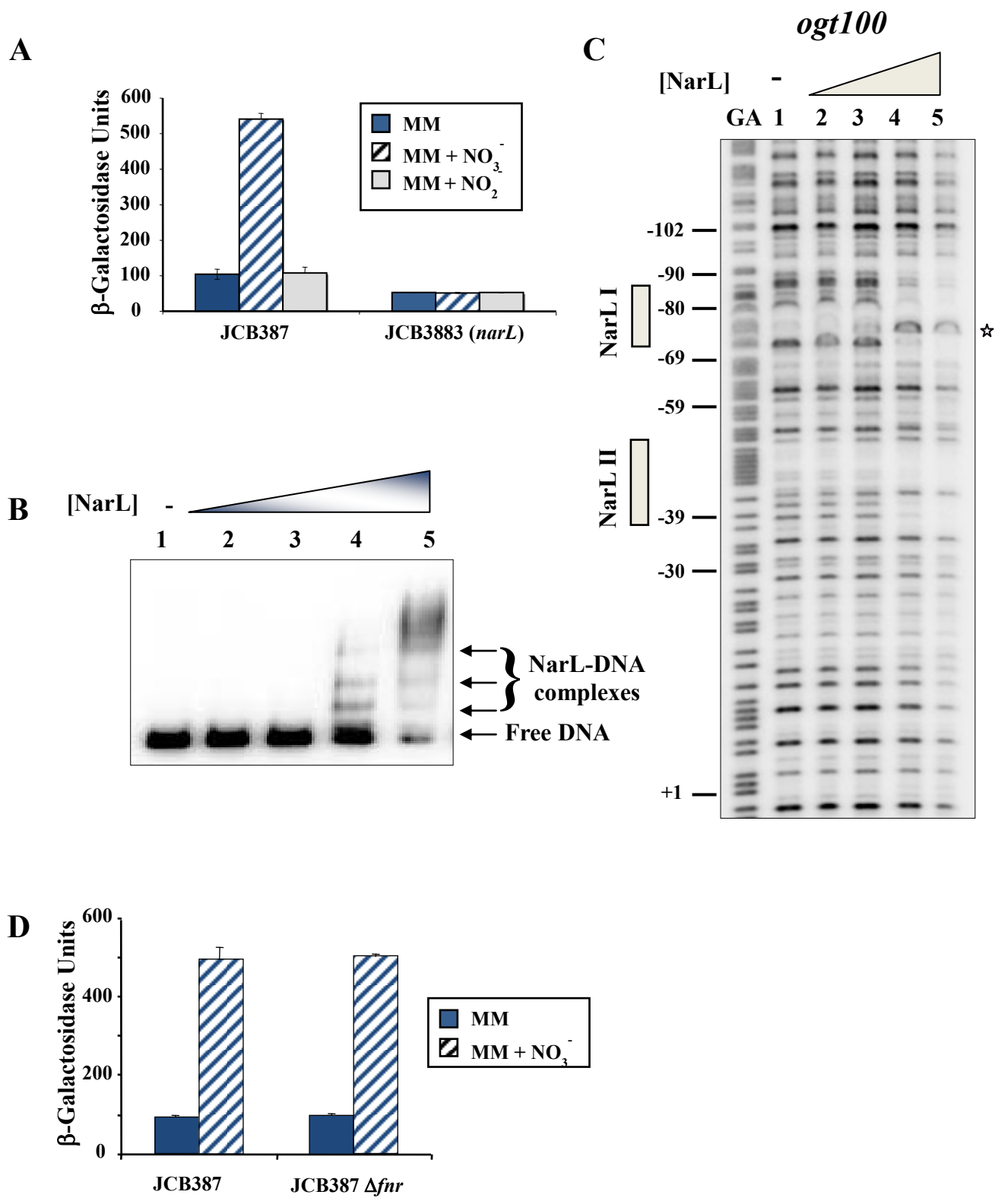
investigate the range of promoter organisations from which NarL / NarP can activate.

4.2. Activation of promoter activity by NarL from a single DNA site

The results of Squire *et al.* (2009) contain a curious result: that NarL induction of the *yeaR* promoter was 5-fold greater than was NarL induction of the *ogt* promoter. Induction of the *yeaR* promoter demonstrated that NarL was able to activate transcription from a single site, which suggests that the different induction of the two promoters might result from differences between the promoter proximal DNA sites for NarL. There are 2 important differences between the promoter proximal DNA sites for NarL at the *yeaR* and *ogt* promoters. First, the DNA site for NarL at the *yeaR* promoter is centred at position -43.5, whilst the promoter proximal DNA site for NarL at the *ogt* promoter is centred at position -45.5, which may be suboptimal. Secondly, the DNA site for NarL at the *yeaR* promoter corresponds better to the consensus DNA site for NarL, which is why NarL binds more readily to the *yeaR* promoter than it does to the *ogt* promoter.

The starting fragment for this work was the ogt100 fragment, a 320 bp EcoRI-HindIII DNA fragment that contains the region from position -270 to +51 relative to the mapped *ogt* transcription start site (see Figure 4.1, Panel C). Both DNA sites for NarL identified by Squire *et al.* (2009) are included in the fragment as is the Shine-Dalgarno element and start codon for *ogt.* To investigate the differing induction of the *yeaR* and *ogt* promoters, derivatives of the ogt100 fragment were constructed (see Figure 4.3, panel A). The ogt106 fragment was constructed by deleting 2 bases between the NarL II site and the transcription start site in order to move the NarL II site so that it is centred at position -43.5. The bases deleted were chosen so as not to disrupt the potential -35 element immediately downstream of the NarL II site. The ogt107 fragment was constructed by mutating the sequence of the NarL II site to that of the DNA site for NarL at the *yeaR* promoter. The resulting fragments were then cloned into pRW50 and plasmids were transformed into JCB387 cells and JCB3883

Figure 4.3: NarL can activate from a single strong DNA site

Panel A shows a partial sequence of the ogt100 and yeaR100 fragments and derivatives of the ogt100 fragment, ogt106 and ogt107. The promoter proximal DNA site for NarL is marked with coloured text with the consensus sequence for such sites given above the sequences in emboldened text. The promoter proximal DNA site for NarL from the ogt100 fragment is marked in green and the DNA site for NarL from the yeaR100 fragment is marked in yellow. The transcription start site is denoted by an emboldened and underlined character with a "+1" above the sequence.

Panel B shows expression from the ogt100, yeaR100, ogt106 and ogt107 fragments cloned as promoter::*lacZ* fusions in pRW50, which was measured in *E. coli* strain JCB387 during exponential growth (OD_{650} = 0.5-0.6) in minimal salt medium. Cells were grown anaerobically or anaerobically in medium supplemented with 20 mM sodium nitrate. Data shown are averages from at least 3 biological repeats, and error bars show one standard deviation from the mean.

Panel C shows expression from the ogt100, yeaR100, ogt106 and ogt107 fragments cloned as promoter::*lacZ* fusions in pRW50, which was measured in *E. coli* strain JCB3884 (Δ*narXL* Δ*narP)* during exponential growth (OD_{650} = 0.5-0.6) in minimal salt medium. Cells were grown anaerobically or anaerobically in medium supplemented with 20 mM sodium nitrate. Data shown are averages from at least 3 biological repeats, and error bars show one standard deviation from the mean.

A

TACYYMTnnAKRRGTA

ogt100 CCCTATCCACTTATCTTTTTGGTGGTA Chart Area GCTGATGTTGCTGGCGTGGTATCTTGTCGGTCTG
+1

yeaR100 TGGTGCTAAAAAGTAACCAATAAATGGTATTTAAAATGCAAATTATCAGGCGTACCCTGAAACGGC

ogt106 CCCTATCCACTTATCTTTTTGGTGGTATGGCTGCTGATGTCTGGCGTGGTATCTTGTCGGTCTG

ogt107 CCCTATCCACTTAACCAATAAATGGTATGGCTGCTGATGTTGCTGGCGTGGTATCTTGTCGGTCTG

B

β-galactosidase units

4000
3000
2000
1000
0

ogt100   yeaR100   ogt106   ogt107

JCB387 cells

■ Anaerobic
■ Anaerobic + nitrate

C

β-galactosidase units

4000
3000
2000
1000
0

ogt100   yeaR100   ogt106   ogt107

JCB3883 cells

■ Anaerobic
■ Anaerobic + nitrate

145
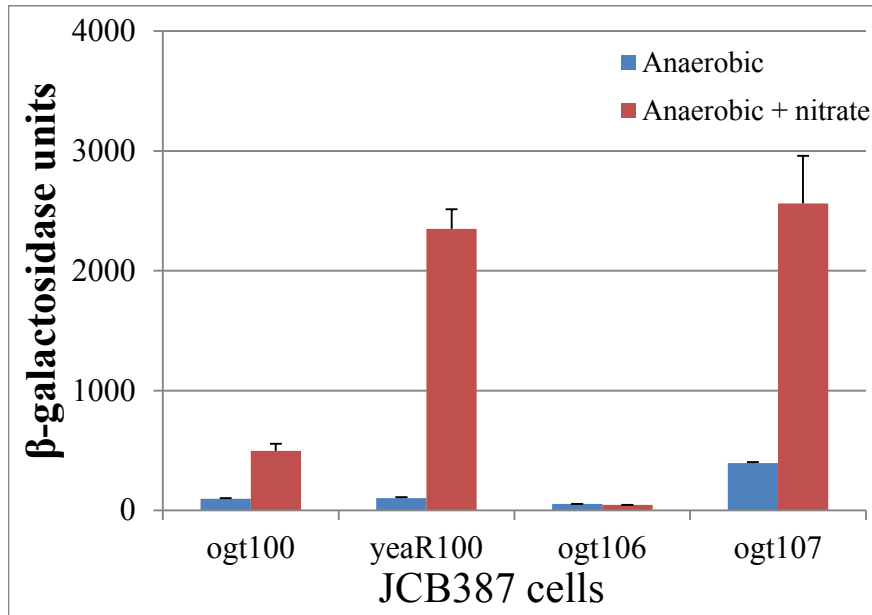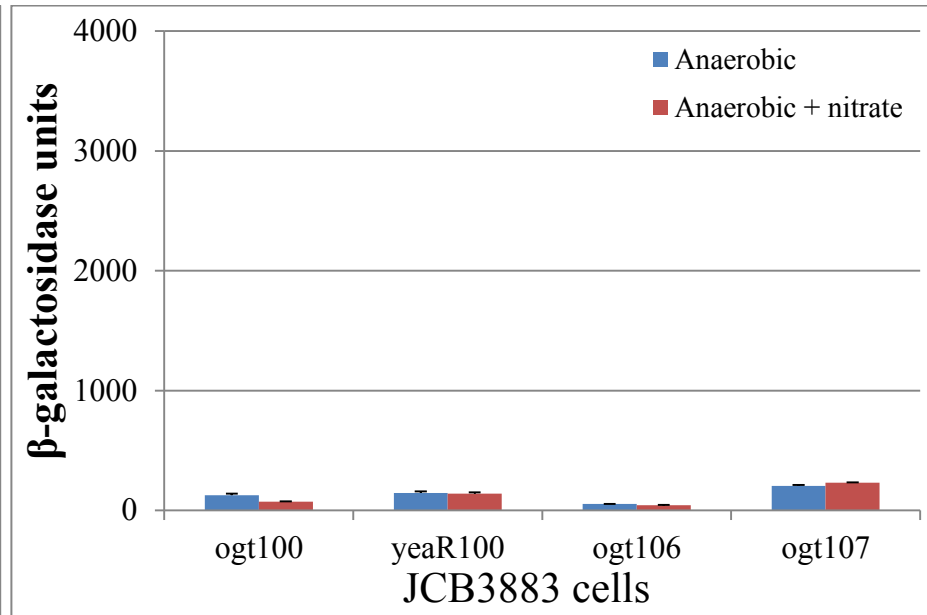
cells, which lack NarL. Cells were grown in minimal salt medium anaerobically or anaerobically with nitrate-supplemented medium.

The results in Figure 4.3, panel B show expression of the promoter::*lacZ* fusions in JCB387 cells. As expected, expression of the ogt100::*lacZ* fusion was induced by nitrate, as was expression of the yeaR100::*lacZ* fusion. In agreement with Squire *et al.* (2009), induction of the *yeaR* promoter was ~5-fold greater than induction of the *ogt* promoter. The ogt106::*lacZ* fusion was expressed poorly under all conditions. This is possibly because the spacing of core promoter elements was disturbed by the deletion of bases in constructing the fragment. Nitrate induced expression of the ogt107::*lacZ* fusion and the yeaR100*::lacZ* fusion to a similar degree, which indicates that the sequence of the DNA site for NarL is the important determinant of efficient induction by NarL. The results in Figure 4.3, panel C show expression of the promoter::*lacZ* fusions in JCB3883 cells, which lack NarL. No nitrate-dependent induction of promoter activity was observed, which indicates that induction observed in the results in Figure 4.3, panel B, is NarL dependent.

Taken with the results of Squire *et al.* (2009), these data confirm that NarL can activate promoter activity independently of FNR and that it can do this from a single DNA site, providing the site is a strong target. The results also demonstrate that NarL can activate efficiently regardless of whether the DNA site is centred at position -43.5 or position -45.5.

4.3. Effect of consensus DNA sites for NarL on *ogt* promoter activity

The results of the previous section indicate that activity of the *ogt* promoter is increased if the promoter proximal DNA site for NarL is modified to better resemble the consensus sequence for such elements. To investigate the effects of modifying one or both DNA sites for NarL to match the consensus sequence, a series of *ogt* promoter fragments was constructed. Derivatives of the ogt100 fragment were prepared, in which one or both DNA sites for NarL were mutated to match the consensus NarL binding sequence (see Figure 4.4, panel A).
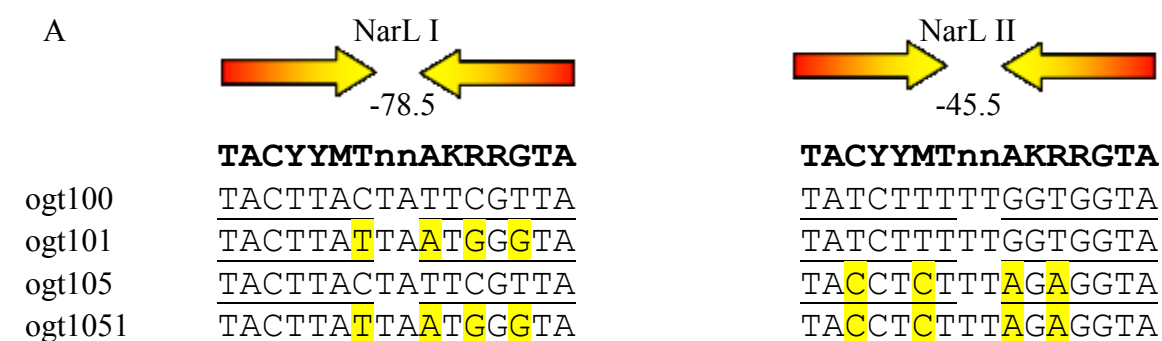
Figure 4.4: Effect of consensus DNA sites for NarL on *ogt* promoter activity

Panel A shows the sequence of the DNA sites for NarL in the ogt100, ogt101, ogt 105, and ogt1051 fragments. The consensus sequence for DNA sites for NarL is given above the site in emboldened text. The central position of the site relative to the *ogt* transcription start is given above the site. Mutations introduced to modify the site to better match the consensus sequence are highlighted in yellow.

Panel B shows sketches of promoters contained on fragments and measured β-Galactosidase units. A wild type DNA site for NarL is indicated in the sketch by a slim yellow bar and an improved site is denoted with a thick yellow bar. The transcription start site is denoted with a bent arrow. β-Galactosidase units were measured in JCB387 cells grown anaerobically in minimal salt medium, with or without 20 mM sodium nitrate, as indicated. Cells contained fragments carried by pRW50. Data shown are the mean of 3 biological repeats and numbers in brackets indicate one standard deviation from the mean.

Resulting fragments were cloned into pRW50 and plasmids were transformed into JCB387 cells. Promoter activity was measured following growth in minimal salt medium anaerobically or anaerobically with nitrate.

Results in Figure 4.4, panel B, show that expression of the ogt100::*lacZ* fusion, which contains the wild type *ogt* promoter, was induced ~4-fold by nitrate. Expression of the ogt101::*lacZ* fusion, in which the NarL I site is modified to better resemble the consensus, was induced ~13-fold by nitrate. Anaerobic expression of the ogt105::*lacZ* fusion, in which the NarL II site is modified to better resemble the consensus, was ~16-fold greater than was anaerobic expression of the ogt100::*lacZ* fusion. Nitrate induced expression of the ogt105::*lacZ* fusion by a further 2-fold. Anaerobic expression of the ogt1051::*lacZ* fusion, in which both DNA sites for NarL were modified to better resemble the consensus sequence, was around 12-fold greater than was anaerobic expression of the ogt100 fragment. Nitrate further induced expression of the ogt1051::*lacZ* fragment by ~4-fold.

These results demonstrate that *ogt* promoter activity is increased if DNA sites for NarL are improved by modifying them to resemble better the consensus binding sequence for NarL. The results also reveal that anaerobic expression of the ogt105::*lacZ* fusion and the ogt1051::*lacZ* fusion is greater than the wild type. It is possible that in the absence of nitrate there is still a small amount of activated NarL, which occupies the improved NarL II site. This suggests that NarL readily binds to and efficiently activates from the NarL II site.

4.4. NarL can activate by class I, II, and III mechanisms

The work of Squire *et al.* (2009) and Lin *et al.* (2008) suggests that NarL can function as a class II activator of the *yeaR* promoter and as a class III activator of the *ogt* promoter. My results suggest that NarL activates by a class III mechanism at the *ogt* promoter, rather than a class II mechanism, because the DNA sites for NarL bear poor resemblance to the consensus sequence. It is reasonable to expect that NarL may also be able to act as a class I activator if

the upstream site is improved, as Squire *et al.* (2008) demonstrated that NarL uses the upstream NarL I site at the *ogt* promoter during class III activation. To investigate the range of promoter organisations that NarL can activate from, a series of DNA fragments was constructed by improving or worsening the DNA sites for NarL in the ogt100 promoter fragment (see Figure 4.5, panel A). Fragments were then cloned into pRW50 and plasmids were transformed into JCB387 cells. Cells were grown in minimal salt medium anaerobically or anaerobically with nitrate and β-Galactosidase activity was measured.

Initially, the effect of disrupting DNA sites for NarL was confirmed. Results presented in Figure 4.5, panel B, show that expression of the ogt100::*lacZ* fusion was induced by nitrate. The ogt102::*lacZ* fusion, which contains a disrupted NarL I site, was poorly induced by nitrate, as was the ogt104::*lacZ* fusion, which contains a disrupted NarL II site. This agrees with the data from Squire *et al.* (2009). As expected, nitrate did not induce expression of the ogt1042::*lacZ* fusion, in which both the NarL I and NarL II sites are disrupted. These results confirm that NarL can only activate by a class III mechanism at the *ogt* promoter if both DNA sites for NarL are present.

To investigate the effect of improving one DNA site for NarL whilst the other is disrupted, the expression of the ogt1041::*lacZ* fusion and the ogt0152::*lacZ* fusion was measured. Anaerobic expression of the ogt1041::*lacZ* fusion, which contains an improved NarL I site but a disrupted NarL II site, was lower than was that of the wild type ogt100::*lacZ* fusion. However, nitrate induced expression of the ogt1041::*lacZ* fusion by ~10-fold. This indicates that NarL can activate by a class I mechanism if the DNA site is modified to resemble better the consensus binding sequence for NarL. Anaerobic expression of the ogt1052::*lacZ* fusion, which contains a disrupted NarL I site and an improved NarL II site, wass 11-fold greater than was that of the wild type ogt100::*lacZ* fusion. Nitrate further

A

NarL I
−78.5
TACYYMTnnAKRRGTA

ogt100	TACTTACTATTCGTTA
ogt102	TAGTTACTATTCGCTA
ogt104	TACTTACTATTCGTTA
ogt1042	TAGTTACTATTCGCTA
ogt1041	TACTTATTAATGGGTA
ogt1052	TAGTTACTATTCGCTA

NarL II
−45.5
TACYYMTnnAKRRGTA

TATCTTTTTGGTGGTA
TATCTTTTTGGTGGTA
TAGCTTTTTGGTGCTA
TAGCTTTTTGGTGCTA
TAGCTTTTTGGTGCTA
TACCTCTTTAGAGGTA
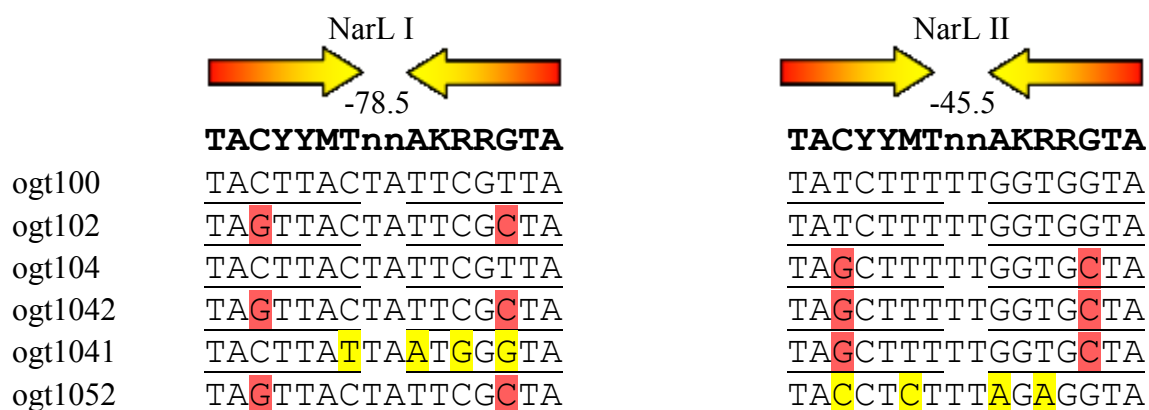
B

Fragment		Anaerobic	Anaerobic + nitrate
ogt100		104 (12)	445 (55)
ogt102		169 (40)	187 (64)
ogt104		115 (4)	180 (5)
ogt1042		67 (2)	67 (1)
ogt1041		63 (4)	641 (73)
ogt1052		1161 (29)	3710 (14)

Figure 4.5: NarL can activate by class I, II or III mechanisms

Panel A shows the base sequence of the two DNA sites for NarL in the ogt100, ogt102, ogt 104, ogt1042, ogt1041 and ogt1052 fragments. The consensus sequence for DNA sites for NarL is given above the site in emboldened text. The central position of the site relative to the ogt transcription start is given above the site. Mutations introduced to modify the site to better match the consensus sequence are highlighted in yellow and mutations introduced to disrupt the DNA site for NarL are highlighted in red.

Panel B shows sketches of promoters contained on fragments and measured β-Galactosidase units. A wild type DNA site for NarL is denoted in the sketch with a slim yellow bar, an improved site is denoted with a thick yellow bar and a disrupted site is denoted with a red X. The transcription start site is denoted with a bent arrow. β-Galactosidase units were measured in JCB387 cells grown anaerobically in minimal salt medium, with or without 20 mM sodium nitrate, as indicated. Cells contained fragments carried by pRW50. Data shown are the mean of 3 biological repeats and numbers in brackets indicate one standard deviation from the mean.

induced expression of the ogt1052::*lacZ* fusion by ~3-fold. This adds to the evidence that NarL can activate *ogt* promoter activity by a class II mechanism if the NarL II site is modified to better resemble the consensus binding sequence.

Taken together, these results show that NarL can activate the *ogt* promoter by class I, II and III mechanisms. However, near consensus DNA sites for NarL are required for NarL to activate by class I and class II mechanisms.

4.5. NarL/P dependent activation of *ogt* promoter derivatives

The results in the previous section indicate that NarL can activate by class I, II and III mechanisms. NarP is known to bind to DNA sites for NarL if the site is organised as an inverted repeat and so it is therefore possible that NarP may also activate *ogt* promoter derivatives that contain improved NarL sites. To investigate this possibility, the activity of *ogt* promoter derivatives that are induced by nitrate was measured in cells that contain both NarL and NarP, just NarL, just NarP or neither.

Plasmids containing *ogt* promoter fragments were transformed into: JCB387, which contain both NarL and NarP; JCB3875, which contain NarL but not NarP; JCB3883, which contain NarP but not NarL; and JCB3884, which do not contain either NarL or NarP. Cells were grown in minimal salt medium anaerobically with or without nitrate and expression of promoter::*lacZ* fusions was measured.

Results illustrated in Figure 4.6, panel A, show expression of promoter::*lacZ* fusions measured in JCB387, which contain both NarL and NarP. Expression of the ogt100::*lacZ* fusion, which contains the wild type *ogt* promoter, was induced by nitrate. Expression of the ogt101::*lacZ* fusion, which contains an improved NarL I site, was induced efficiently by nitrate to a higher level than was the wild type *ogt* promoter. Expression of the ogt1041::*lacZ* fusion, which contains an improved NarL I site and a worsened NarL II site,

Figure 4.6: Activity of *ogt* promoter derivatives in cells that lack NarL, NarP, or both.

Panel A shows expression from the ogt100, ogt101, ogt1041, ogt105, ogt1051, and ogt1052 fragments cloned as promoter::lacZ fusions in pRW50. The status of the NarL I and NarL II sites in the fragments are given below the fragment name. WT indicates the site is wild type, an upwards arrow indicates the site is modified to better resemble the NarL consensus binding site and a downwards arrow indicates that that site has been disrupted. For example, "I: WT II: ↑" indicates that the NarL I site is wild type whilst the NarL II site has been modified to better resemble the consensus. Expression of promoter::*lacZ* fusions was measured in *E. coli* strain JCB387 during exponential growth (OD_{650} = 0.5-0.6) in minimal salt medium. Cells were grown anaerobically or anaerobically in medium supplemented with 20 mM sodium nitrate. Data shown are averages from at least 3 biological repeats, and error bars show one standard deviation from the mean.

Panel B shows expression from the ogt100, ogt101, ogt1041, ogt105, ogt1051, and ogt1052 fragments cloned as promoter::lacZ fusions in pRW50. Expression was measured in JCB3875 cells, which lack NarP, during exponential growth (OD_{650} = 0.5-0.6) in minimal salt medium. Cells were grown anaerobically or anaerobically in medium supplemented with 20 mM sodium nitrate. Data shown are averages from at least 3 biological repeats, and error bars show one standard deviation from the mean.

Panel C shows expression from the ogt100, ogt101, ogt1041, ogt105, ogt1051, and ogt1052 fragments cloned as promoter::lacZ fusions in pRW50. Expression was measured in JCB3883 cells, which lack NarL, during exponential growth (OD_{650} = 0.5-0.6) in minimal salt medium. Cells were grown anaerobically or anaerobically in medium supplemented with 20 mM sodium nitrate. Data shown are averages from at least 3 biological repeats, and error bars show one standard deviation from the mean.

Panel D shows expression from the ogt100, ogt101, ogt1041, ogt105, ogt1051, and ogt1052 fragments cloned as promoter::lacZ fusions in pRW50. Expression was measured in JCB3884 cells, which lack NarL and NarP, during exponential growth (OD_{650} = 0.5-0.6) in minimal salt medium. Cells were grown anaerobically or anaerobically in medium supplemented with 20 mM sodium nitrate. Data shown are averages from at least 3 biological repeats, and error bars show one standard deviation from the mean.

A

B

C

D

153
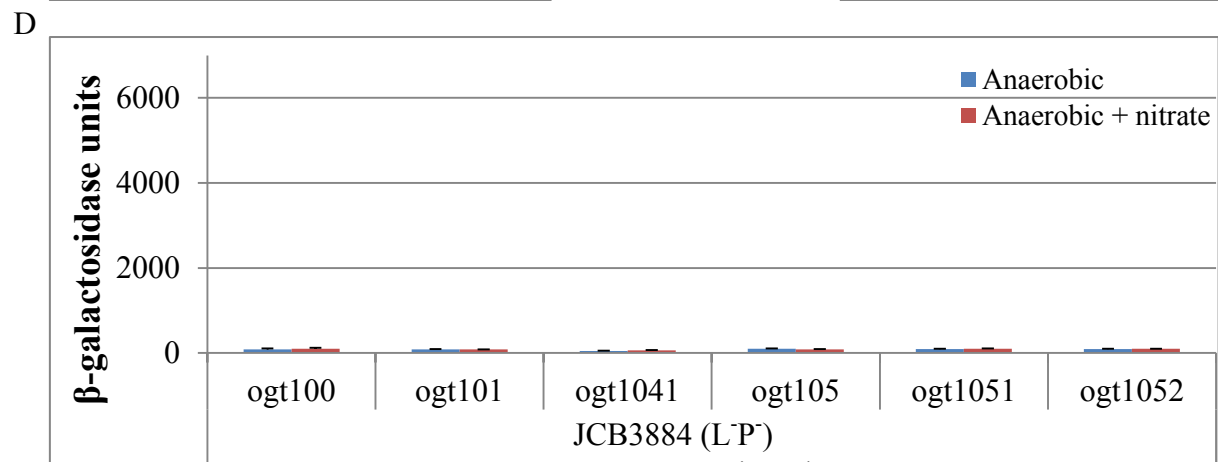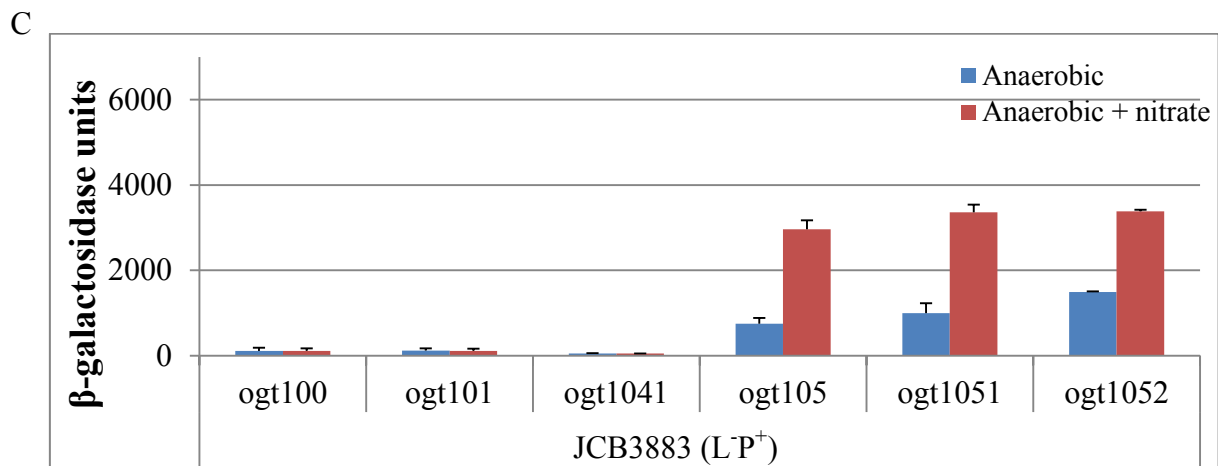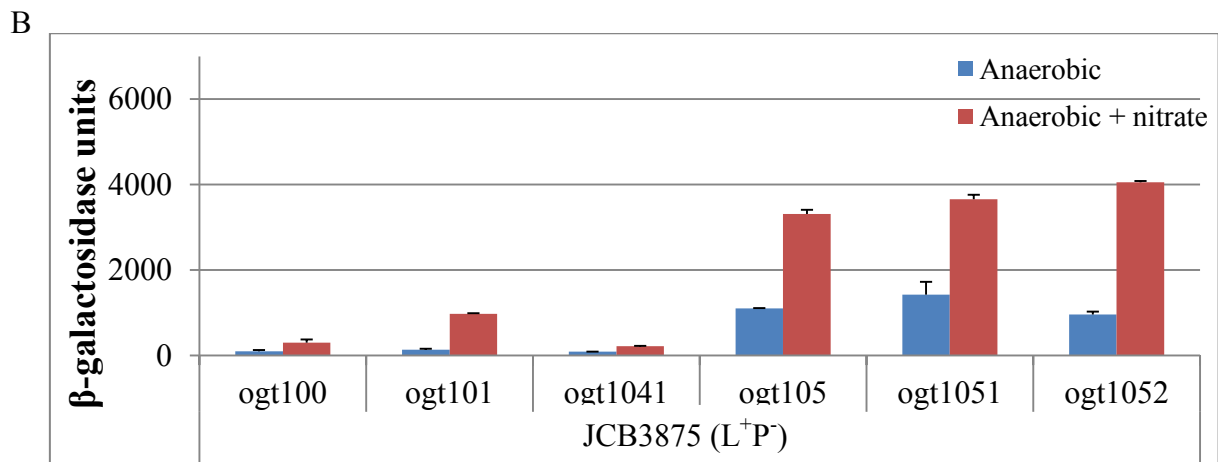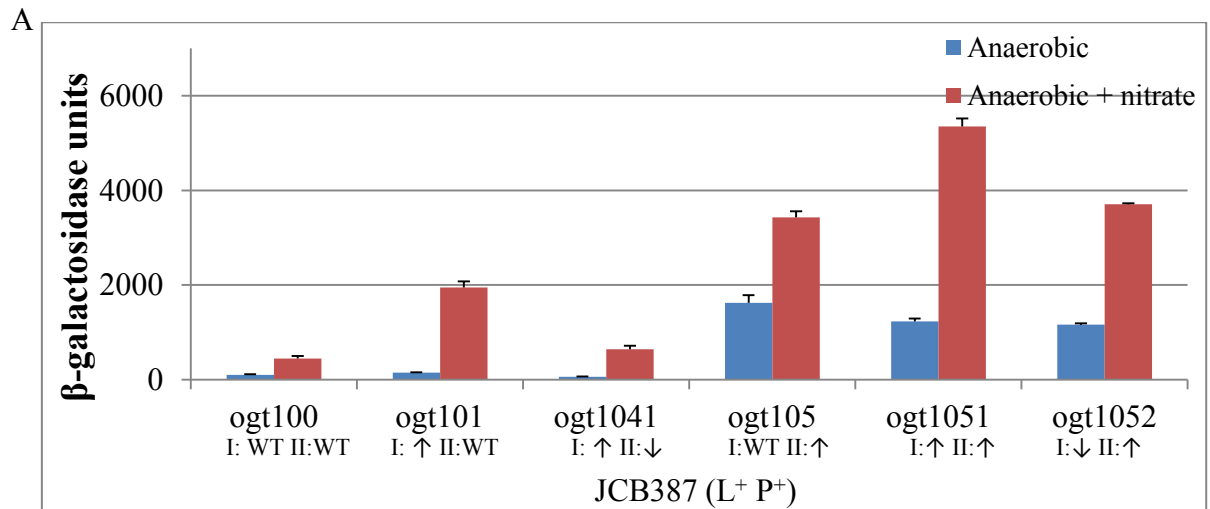
was induced by nitrate to a similar level as the wild type *ogt* promoter. The ogt105::*lacZ* fusion, which contains an improved NarL II site, wass expressed highly anaerobically and nitrate induced expression further. The ogt 1051::*lacZ* fusion, in which both the NarL I and NarL II sites are improved, was expressed highly anaerobically and nitrate induced further. The ogt1052::*lacZ* fusion, in which the NarL I site is worsened and the NarL II site improved, showed a similar pattern of expression to that of the ogt105::*lacZ* fusion. This confirms that promoter::*lacZ* fusions with improved DNA sites for NarL are expressed more highly than is the wild type ogt100::*lacZ* fusion.

Results illustrated in Figure 4.6, panel B, show expression of promoter::*lacZ* fusions measured in JCB3875 cells, which contain NarL but not NarP. Absolute expression levels of the ogt101::*lacZ* fusion and the ogt1051::*lacZ* fusion were marginally lower in cells that do not contain NarP, which may indicate a subtle repressive effect of NarP. Overall, expression of the fusions was very similar to that observed in cells that contained both NarL and NarP. This indicates that none of the fusions rely solely on NarP for activation of expression.

Results illustrated in Figure 4.6, panel C, show expression of promoter::*lacZ* fusions in JCB3883 cells, which contain NarP but not NarL. Expression of the ogt100::*lacZ* fusion was not induced by nitrate, which agrees with that reported by Squire *et al* (2009). Neither expression of the ogt101::*lacZ* fusion nor that of the ogt1041::*lacZ* fusion was induced by nitrate, which indicates that NarP is unable to activate from the promoter-distal NarL I site. That expression of the ogt1041::*lacZ* fusion was not activated by NarP indicates that NarP is unable to activate by a class I mechanism. Expression of the promoter::*lacZ* fusions that contain improved NarL II sites, ogt105, ogt1051 and ogt1052, was induced by nitrate, which indicates that NarP is able to activate by a class II mechanism from the promoter proximal NarL II site.

Results illustrated in Figure 4.6, panel D, show expression of promoter::*lacZ* fusions in JCB3884 cells, which do not contain NarL or NarP. As expected, nitrate did not induce expression of any of the fragments. This suggests that the induction observed in JCB387 cells is entirely dependent on either NarL or NarP and that no other factors are mediating induction by nitrate.

Taken together, these results suggest that NarL is able to activate by class I, II and III mechanisms, whilst NarP is only able to act via a class II mechanism. However, a good target site is required for NarL to activate by class I or II mechanisms.

4.6. *ogt* promoter derivatives that are induced by nitrite

Squire *et al.* (2009) demonstrated that an ogt100::*lacZ* fusion is induced by nitrate but not by nitrite. Recall that when cells are grown in nitrate, NarL is thought to be active whilst when cells are grown in nitrite, NarP is thought to be most active. The results in Figure 4.6 indicate that NarP cannot induce expression of the ogt100::*lacZ* fusion and so it follows that nitrite would not induce expression. However, NarP can activate expression of the ogt105::*lacZ*, ogt1051::*lacZ* and 1052::*lacZ* fusions and so nitrite induction of these fusions was investigated. Plasmids were transformed into JCB387 cells and cells were grown in minimal salt medium anaerobically, anaerobically with nitrate and anaerobically with nitrite.

Results in Figure 4.7 confirm that expression of the ogt100::*lacZ* fusion was induced by nitrate but was not induced by nitrite. Anaerobic expression of the ogt105::*lacZ* fusion, in which the NarL II site is improved, was higher than that of the ogt100::*lacZ* fusion and both nitrate and nitrite induced activity further.

Expression of the ogt1051::*lacZ* fusion, in which both NarL I and NarL II sites are improved, was induced by nitrate and nitrite. Expression of the ogt1052::*lacZ* fusion, in which the NarL I site is worsened and the NarL II site improved, was induced by

155
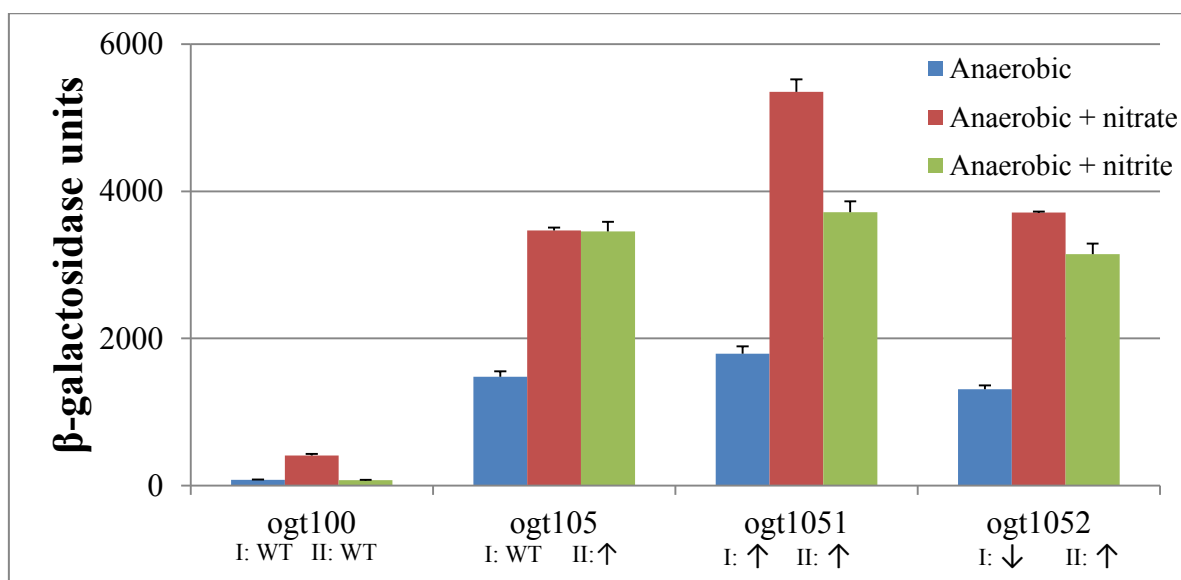
Figure 4.7: *ogt* **promoter derivatives that are induced by nitrite**

The figure shows expression from the ogt100, ogt105, ogt1051, and ogt1052 fragments cloned as promoter::*lacZ* fusions in pRW50. The status of the NarL I and NarL II sites in the fragments are given below the fragment name. WT indicates the site is wild type, an upwards arrow indicates the site is modified to better resemble the NarL consensus binding site and a downwards arrow indicates that that site has been disrupted. For example, "I: WT II: ↑" indicates that the NarL I site is wild type whilst the NarL II site has been modified to better resemble the consensus. Expression of promoter::*lacZ* fusions was measured in *E. coli* strain JCB387 during exponential growth (OD$_{650}$ = 0.5-0.6) in minimal salt medium. Cells were grown anaerobically, anaerobically in medium supplemented with 20 mM sodium nitrate, or anaerobically in medium supplemented with 2.5 mM sodium nitrite. Data shown are averages from at least 3 biological repeats, and error bars show one standard deviation from the mean.

nitrate and was maximally induced by nitrite.

These results, in combination with previous results, suggest that NarP can activate by a class II mechanism in a nitrate and nitrite-dependent manner.

4.7. Induction kinetics of the *hcp* and *ogt* promoters

The results strongly suggest that NarL induces *ogt* expression directly in a nitrate dependent manner. By contrast, NarL is thought to induce expression of *hcp* indirectly, by its involvement in the production of RNS, which cause de-repression by NsrR (see Chapter 3). It is therefore predicted that, following the addition of nitrate, induction of *ogt* expression would occur more rapidly than would induction of *hcp* expression.

To test this prediction, JCB387 cells were transformed with pRW50 plasmids containing either the ogt100 fragment or the hcp383-0 fragment to generate *lacZ* fusions. Cells were grown in minimal salt medium anaerobically and pulsed with nitrate. The results in Figure 4.8 show that expression of the ogt100::*lacZ* fusion increased ~12 minutes after the culture was pulsed with nitrate. Expression of the hcp383-0::*lacZ* fusion did not increase until ~21 minutes after the culture was pulsed with nitrate. These results support the hypothesis that NarL directly activates the *ogt* promoter but indirectly activates the *hcp* promoter. It is likely that the difference of 9 minutes between induction of the *ogt* promoter and induction of the *hcp* promoter represents the time taken for RNS to be generated from nitrate.

4.8. Discussion

4.8.1. Regulation of ogt *expression*

The *E. coli* Ogt protein has been shown to repair methylated DNA, as has another protein, Ada (Potter *et al.*, 1987). Ada is a fascinating example of self-regulation since it is both a methyltransferase and a transcription factor: the structure of Ada is modified when it repairs a methylated nucleotide, which makes it competent to activate expression of its own gene, *ada*
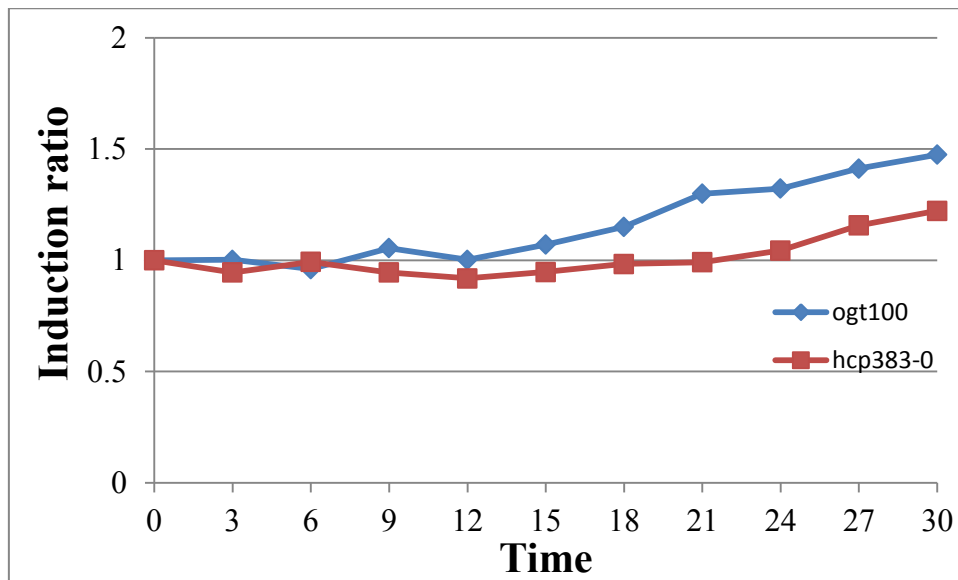
Figure 4.8: Induction kinetics of the *hcp* and *ogt* promoters

Cultures of strain JCB387 carrying pRW50 containing the hcp383-0 or ogt100 promoter fragments were grown anaerobically at 37 ºC in minimal medium. 20 mM sodium nitrate was added when cells reached OD_{650} of ~0.4. β-Galactosidase activities were measured immediately prior to the addition of sodium nitrate (timepoint 0) and at 3 minute intervals following addition. The figure shows the ratio of the measured activity relative to timepoint 0 plotted against time after the addition of sodium nitrate.

(Teo *et al.,* 1986). This elegant mechanism directly connects expression of *ada* to DNA damage. By contrast, the *ogt* gene was initially reported to be constitutively expressed at a low level, although recent work has shown that expression is activated by NarL in a nitrate-dependent manner (Squire *et al.*, 2009). Squire *et al.* (2009) also demonstrated that FIS regulates *ogt* expression by repressing promoter activity when cells are growing in rich media.

It is interesting that the expression of a gene encoding a DNA repair enzyme is activated by NarL. NarL is activated in response to the presence of nitrate and is partially activated in response to the presence of nitrite, neither of which is reported to cause DNA damage directly. However, recall that nitrate and nitrite are reduced when *E. coli* is growing anaerobically and by-products of this reduction include RNS, which are reported to cause DNA damage (Weiss, 2006). It is therefore possible that the regulation of *ogt* expression by NarL is a prophylactic mechanism designed to protect cells from future products of nitrate reduction, which in turn suggests that Ada is a general-purpose DNA repair enzyme whilst Ogt mitigates a particular threat. It is curious that Ogt is not upregulated in response to nitrite. Expression of *ogt* may be repressed by FIS may because, in rich media, cell growth and division is faster than it is in non-rich media and the energy expended by the cell on generating Ogt might be more advantageously used for cell growth and division.

4.8.2. NarL regulation of the ogt promoter

NarL has been found to activate *ogt* promoter activity by a class III mechanism from 2 DNA sites that differ significantly from the consensus NarL binding sequence. However, even when fully induced, the *ogt* promoter is not highly active, which means that Ogt would be expected to accumulate slowly in the cell. This work has shown that NarL can activate by a class I mechanism from an upstream DNA site that matches the consensus sequence for NarL binding. Results in Figure 4.4 show that expression of the ogt100::*lacZ* fusion, which contains the wild type *ogt* promoter, and expression of the ogt1041::*lacZ* fusion, which contains a

consensus upstream NarL I site and a disrupted NarL II site, are similar. This indicates that NarL could activate the *ogt* promoter by either a class III or class I mechanism and produce a similar level of expression, leading to the question: why has the *ogt* promoter evolved to be activated by a class III mechanism rather than a class I mechanism?

First, it is possible that NarL and FIS interact in the regulation of *ogt* expression. The NarL I site and the DNA site from which FIS represses overlap (see Figure 4.9) and therefore it is possible that if there were not a promoter proximal DNA site for NarL, then NarL would be unable to overcome FIS repression. Alternatively, if NarL activated *ogt* expression from a site centred at position -45.5 only, then FIS might not be able to repress. Interaction of FIS and NarL at the *ogt* promoter could be investigated using a wild type *ogt* promoter::*lacZ* fusion and a class I / II derivative in FIS$^+$ and FIS$^-$ cells. Competitive *in vitro* studies of FIS and NarL binding to the *ogt* promoter fragment may also help to elucidate the relative functions of the proteins.

The second possibility is that two poor DNA sites for NarL may ensure that *ogt* will only be optimally expressed when higher levels of activated NarL are present in the cell due to higher concentrations of nitrate. NarL is thought to activate target promoters hierarchically as weaker target sites are only bound once stronger target sites are already occupied. If the *ogt* promoter contained a single strong target for NarL, then expression of *ogt* might be induced in response to a lower concentration of nitrate than is optimal for the cell. This could be investigated by studying induction of the wild-type *ogt* promoter and a derivative activated by a class I mechanism in response to a range of nitrate concentrations. The relative fitness of cells could be gauged using a competitive growth experiment in a chemostat with cells where *ogt* expression is driven by the natural promoter and cells where *ogt* expression is driven by a class I activated derivative of the natural promoter (Clegg *et al.*, 2006).
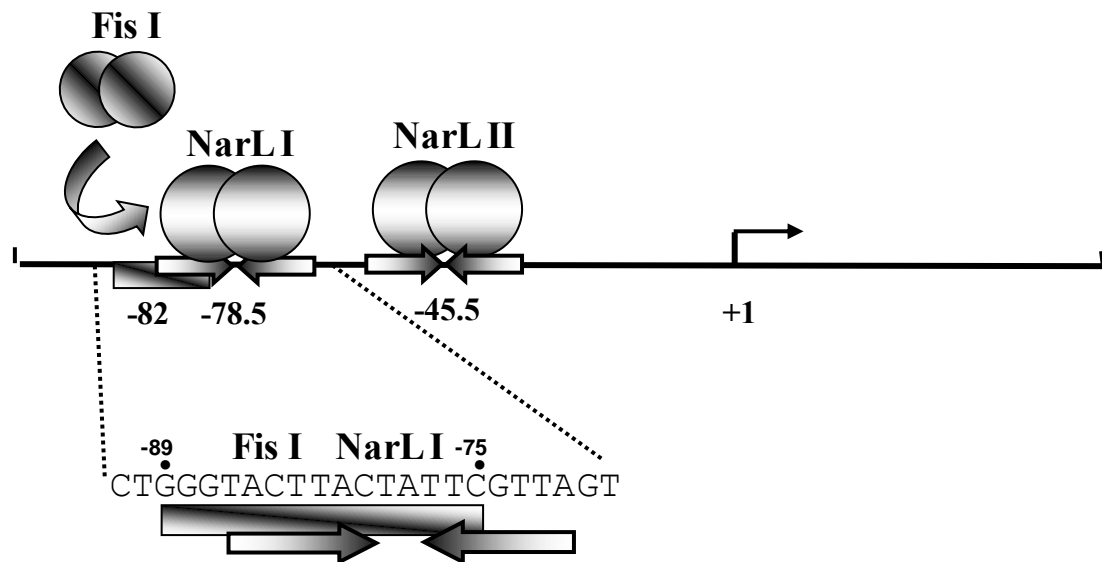
Figure 4.9: DNA site for FIS and NarL overlap at the *ogt* promoter

Reproduced from Squire *et al.* (2009). The figure is a diagram that shows the NarL I and II sites and the Fis I site at the *ogt* promoter. The Fis I site is represented by a diagonally shaded box and Fis protein is shown as diagonally shaded circles. The DNA sites for NarL are inverted horizontal arrows and NarL is represented as vertically shaded circles. The *ogt* transcription start site is denoted with a bent arrow and the label "+1". The positions given below the DNA sites represent the centre point for the site relative to the *ogt* transcription start. The partial sequence below the diagram shows the sequence from position -91 to -69 of the *ogt* promoter and shows the juxtaposition of the Fis I and NarL I sites.

4.8.3. *NarL can activate by class I, II and III mechanisms*

This work has shown that NarL can activate transcription independently of FNR by Class I, II and III mechanisms. This is to be expected, as other members of the LuxR family (to which NarL belongs) are able to activate directly.

NarL activation of *ogt* promoter derivatives by a class I mechanism appears to be less efficient than is activation by a class II mechanism. It is possible that this is because the NarL I site in the *ogt* promoter is at a sub-optimal location. However, experiments to move the NarL I site to a different position relative to the *ogt* transcription start do not markedly improve promoter activity (Patawarachin Ruanto, personal communication). It is also possible that if the NarL I site is made to match the consensus NarL binding sequence, then transcription will start from a different position than it does in the wild type *ogt* promoter. Patawarachin Ruanto (University of Birmingham, UK) has shown that if the confirmed *ogt* -10 element is disrupted in the ogt1041 fragment, which contains a consensus NarL I site and a disrupted NarL II site, then promoter activity is minimal (Figure 4.10). This indicates that the same *ogt* promoter is functioning in the ogt1041 fragment and observed activity is not due to the chance activation of a second promoter.

My work has shown that NarL can activate the *ogt* promoter very efficiently from a class II position. Derivatives of the *ogt* promoter that contain a consensus NarL II site are over 10-fold more active than the wild type promoter. Interestingly, derivatives of the *ogt* promoter that contain a consensus NarL II site are also highly active in the absence of nitrate, when NarL is not thought to be active. It is to be expected that there are small levels of active NarL in the absence of nitrate and that this active NarL is enough to activate promoter activity. However, the *yeaR* promoter has a near consensus DNA site for NarL in a class II position but does not show high levels of promoter activity in the absence of nitrate (Squire *et al.* 2009). A possible explanation for this discrepancy is that the position of the DNA site for
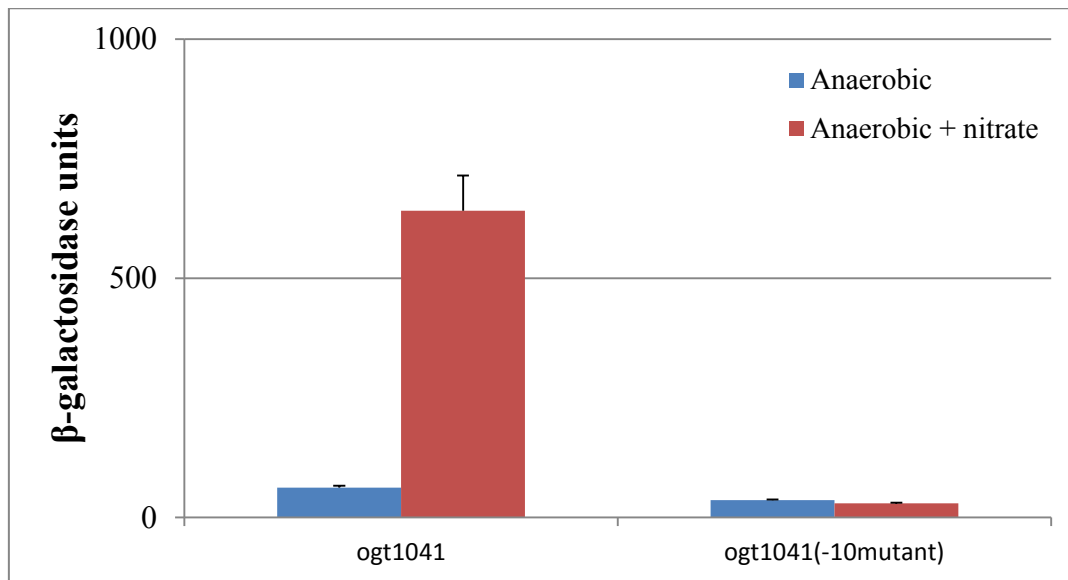
Figure 4.10: Mutation of the -10 element in the ogt1041 fragment abrogates promoter activity

Data from Patawarachin Ruanto (University of Birmingham) that shows expression of either an ogt1041::*lacZ* fusion or and derivative with a guanine substitution at position -12. Expression was measured in JCB387 cells following growth in minimal salt medium anaerobically or anaerobically and with medium supplemented with 20 mM nitrate. Data shown are the average of 3 biological repeats and error bars denote one standard deviation from the mean.

NarL in the *ogt* promoter may be more optimal than is the position of the DNA site for NarL in the *yeaR* promoter. This hypothesis is disproved however by the results in Figure 4.3, panel B, which shows that a derivative of the *ogt* promoter that contains the sequence of the DNA site for NarL from the *yeaR* promoter, but at the *ogt* NarL II site position, is not highly active in the absence of nitrate.

Another possible explanation is that the improved NarL sites in the *ogt* promoter derivatives are a better target for NarL than is the DNA site for NarL in the *yeaR* promoter. Figure 4.11 shows the sequences of the improved NarL I / II sites from *ogt* promoter derivatives used in this study and that of the DNA site for NarL from the *yeaR* promoter aligned with the consensus binding sequence for NarL. It shows that the improved DNA sites for NarL in the *ogt* promoter match the consensus binding sequence for NarL at all 14 positions, whilst the sequence of the DNA site for NarL at the *yeaR* promoter only matches the consensus sequence at 9 of the 14 positions. This implies that the consensus NarL II site at the *ogt* promoter would present a stronger target for NarL than would the DNA site for NarL at the *yeaR* promoter.

Overall, my results show that NarL can activate the *ogt* promoter by a class III mechanism from two weak DNA sites, both of which are crucial. However, if one site is modified to match the consensus sequence for DNA sites for NarL, then the other site becomes unnecessary. Thus, when a site is improved, it is no longer clear by which mechanism NarL activates. Results in Figure 4.4 show that the ogt101::*lacZ* fusion, in which the NarL I site is improved and the NarL II site is wild type, is expressed 3-fold greater than is the ogt1041::*lacZ* fusion, in which the NarL I site is improved and the NarL II site is disrupted. This indicates that the wild type NarL II site in the ogt101 fragment has a role in NarL-dependent activation of expression of the ogt101::*lacZ* fusion, but that the NarL II site is no longer crucial. Further study would be necessary to determine the mechanism by which NarL

		Score
Consensus DNA site for NarL	**TACYYMTnnAKRRGTA**	Score
Improved NarL I site (ogt105 fragment)	TACTTATTAATGGGTA	14 / 14
Improved NarL II site (ogt105 fragment)	TACCTCTTTAGAGGTA	14 / 14
DNA site for NarL (*yeaR* promoter / ogt107 fragment)	TAACCAATAAATGGTA	9 / 14

Figure 4.11: Alignment of DNA sites for NarL to the consensus sequence

Figure shows the consensus sequence for DNA sites for NarL, the sequences of improved NarL I and II sites from *ogt* promoter derivatives and the sequence of the DNA site for NarL from the *yeaR* site. Fasta standard ambiguity codes are used: a Y denotes C or T; an M denotes A or C; a K denotes G or T and an R denotes A or G. Positions that match the consensus sequence are highlighted purple and score shows how many positions match out of a possible 14. The centre 2 base pairs in between the NarL heptamers are not included in this score.

activates transcription from such a promoter derivative. For example, a tethered nuclease such as FeBABE could be used to locate the positions of the CTD of the RNAP α subunit when bound to the *ogt* promoter, which may help clarify the mechanism of NarL activation (Miller and Hahn, 2006).

4.8.4. *NarP is a more discriminating activator than is NarL*

An interesting observation from the results of this work is that NarL can activate from positions, and by mechanisms, that NarP cannot. Hence, NarL must be capable of an activation mechanism that NarP is not. Specifically, NarP appears to be able to activate by a class II mechanism only, whilst NarL can activate by class I, II and III mechanisms. Also, NarP is unable to activate the *yea*R promoter and the *ogt* promoter derivative carried by the ogt107 fragment, whilst NarL is able to activate both (see Figure 4.3, panels B and C).

These observations pose two separate problems. The inability of NarP to activate by class I or class III mechanisms is likely to be a functional problem that arises from the structural differences between NarL and NarP. Although NarL and NarP are homologous, they are not identical. The inability of NarP to activate transcription of the *yeaR* promoter is possibly because the DNA site is too poor a target for NarP. An EMSA study of NarP binding to a radiolabelled DNA fragment containing the *yeaR* promoter did not find evidence of binding (personal communication from Derrick Squire, University of Birmingham, UK). Some of these issues are tackled in the work described in Chapter 5.

Chapter 5: Functional studies on the NarL and NarP proteins

5.1. Introduction

NarL belongs to the LuxR-GerE-TraJ family of transcription factors. Members of this family such as LuxR, have been reported to activate transcription by directly interacting with RNA polymerase (Choi and Greenberg, 1991). Because of its similarity to other family members, NarL would also be expected to be able to interact directly with RNAP to activate transcription. This hypothesis is supported by the work of Lin *et al.* (2007) and Squire *et al.* (2009), which showed that NarL can activate the *yeaR* and *ogt* promoters independently of other factors.

Like many of the LuxR-GerE-TraJ family, NarL has 2 distinct domains, the NTD and the CTD. The CTD contains a LuxR type helix-turn-helix DNA binding motif and has been shown to be competent to bind DNA independently of the NTD (Maris *et al.*, 2002). The NTD is thought to regulate DNA binding by the CTD by obscuring the helix-turn-helix motif. When NarL is phosphorylated by NarX or NarQ, it is thought to undergo a conformational change that repositions the NTD away from the CTD and permits DNA binding.

The research described in this chapter seeks to examine whether NarL interacts directly with RNAP and to understand the relative contributions of the NTD and CTD with regard to promoter activation.

5.2. Residues of αCTD important for activation by NarL at the *ogt* promoter

The results of this work and those of Squire *et al.* (2009) indicate that NarL activates transcription from the *ogt* promoter by binding to sites at positions where it might directly interact with the RNAP polymerase α subunit. To investigate whether this is the case, the effect on NarL-dependent activation of 69 alanine substitutions within the RNA polymerase α subunit, covering residues 255 to 239, was measured.

To ascertain the effect of alanine substitutions of the RNAP α on NarL-dependent activation, JCB387 or JCB3875 (*ΔnarP*)cells were transformed with pRW50 carrying an *ogt* promoter fragment. Cells were then transformed with plasmids expressing either wild-type α or derivatives carrying alanine substitutions in the CTD. Cells were grown anaerobically in minimal salt medium and in the presence of nitrate to ensure NarL was phosphorylated before the β-Galactosidase activity was measured. Mutation of a residue was said to improve or impair activation if it resulted in average activity of the *ogt* promoter that was equal to or greater than 15% different from the activity of the *ogt* promoter in cells containing wild type RNA polymerase α subunit.

5.2.1. *Class III activated,* ogt *promoter*

Results in Figure 5.1 show expression of the ogt100::*lacZ* fusion, which contains the wild type, class III activated *ogt* promoter, in JCB387. The NarL-dependent expression of the ogt100::*lacZ* fusion was affected, positively and negatively, by substitution of different residues in the αCTD with alanine. Alanine substitution of residues 258, 266, 273, 287, 288, 290, 293, 307 impaired class III activation by NarL. Alanine substation of residues 259, 260, 262, 263, 265, 268, 269, 278, 283, 284, 291, 297, 298 and 303 notably improved class III activation by NarL. These results indicate that, when activating transcription by a class III mechanism, NarL does interact with RNAP via the αCTD.

Residues of αCTD identified by the alanine scan as being involved in class III activation by NarL were mapped onto the structural model of αCTD. Figure 5.2 shows that many residues identified are surface exposed, which supports the hypothesis that they are involved in interaction with NarL. The majority of the residues that, when mutated to alanine, negatively affect class III activation by NarL can be seen to cluster into a band around the outside of the α CTD. This could represent multiple distinct contacts with NarL, such as is seen at class II CRP-dependent promoters, or could indicate that NarL makes one, extensive,.
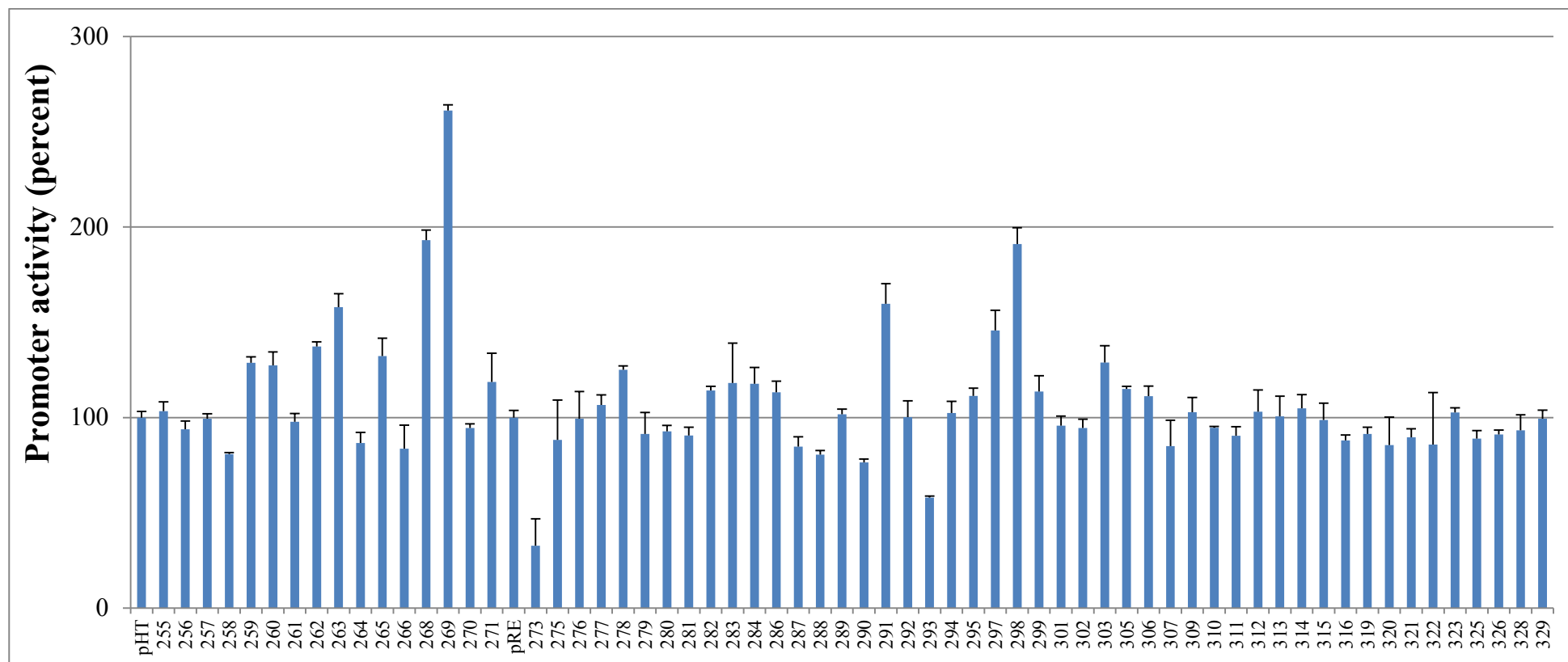
Figure 5.1 Residues of αCTD that are important for NarL-dependent activation of the *ogt* promoter

JCB387 cells containing pRW50 carrying the ogt100 fragment were transformed with derivatives of plasmids pHTf1 (substitutions at 255 to 271 and 302) or pREII (substitutions at remaining positions). Wild type pHTf1 and pRE11 are labelled pHT and pRE respectively. Each plasmid encoded an α derivative with a single alanine substitution between residues 255 and 329 as indicated in the figure. Galactosidase activities are expressed as percentages of the activity obtained with cells transformed with plasmids encoding wild-type α (100% = 648 Miller units). Residues 267, 272, 274, 308, 324, and 327 are alanines in the wild-type protein. Each bar shows the average of 3 biological replicates and errors bars indicate 1 standard deviation from the mean.
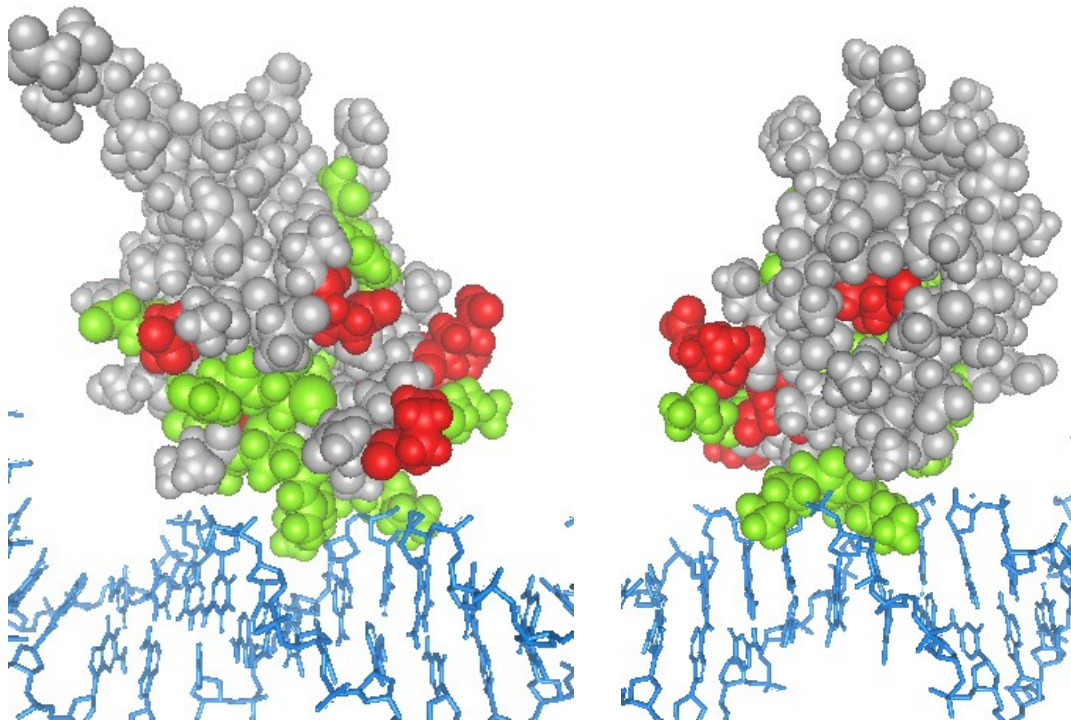
Figure 5.2 Model of αCTD indicating position of residues implicated in class III activation by NarL

Structure of αCTD (Jeon *et al.*, 1995) showing side chains identified as critical for class I activation by NarL. Two views of the structure, related by a 180° rotation on the vertical axis, are shown. Residues where alanine replacement impairs class I activation by NarL are coloured red and those that, when replaced, improve class I activation are coloured green.

contact with the αCTD. Residue 266 appears to be inside the core structure of the αCTD Some of the residues that, when mutated to alanine, improved NarL-dependent class III activation fall in the DNA binding domain of the αCTD. Residues 268, 269 and 298 have all been identified as being involved in recognition of UP sites (Gourse and Gaal, 2000). It is possible that by disrupting the interaction of the αCTD with DNA, NarL is better able to sequester the αCTD. Together these results support the hypothesis that NarL interacts directly with RNAP to recruit it to the *ogt* promoter.

5.2.2.   Class I activated derivative of the ogt *promoter*

If NarL interacts with RNAP directly during class III activation then it is highly likely that it does so during class I activation as well. Recall that class III activation can be viewed as a hybrid of class I and class II activation (see Figure 1.7). Thus, the contacts that NarL makes with the αCTD for class I activation would be expected to be a subset of those contacts it makes for class III activation.

Results in Figure 5.3 show expression of the ogt1041::*lacZ* fusion, which contains a class I activated derivative of the *ogt* promoter, in JCB3875. The NarL-dependent expression of the ogt1041::*lacZ* fusion is affected, positively and negatively, by mutation of residues in the αCTD to alanine. Alanine substitution of residues 259, 261, 271, 273, 280, 288, 290, 293, 296, 302, and 328 impaired class I activation by NarL. Alanine substitution of residues 257, 260, 262, 268, 269, 270, 278, 281, 284, 291, 298, 300, 303, 304, 305, 307, 314, and 316 notably improved class I activation by NarL. These results suggest that NarL interacts with RNAP via the αCTD when activating by a class I mechanism. As expected, there is an overlap between the residues of αCTD involved in class I activation by NarL and those involved in class III activation. Class III activation typically involves more contacts with polymerase that does class I activation (see Section 1.10). Therefore, it is unexpected that more residues of αCTD appeared to be involved in class I activation by NarL than are in class III activation by

Figure 5.3 Residues of αCTD that are important for class I activation by NarL

JCB3875 cells containing pRW50 carrying the ogt1041 fragment were transformed with derivatives of plasmids pHTf1 (substitutions at 255 to 271 and 302) or pREII (substitutions at remaining positions). Wild type pHTf1 and pRE11 are labelled pHT and pRE respectively. Each plasmid encoded an α derivative with a single alanine substitution between residues 255 and 329 as indicated in the figure. β-Galactosidase activities are expressed as percentages of the activity obtained with cells transformed with plasmids encoding wild-type α (100% = 405 Miller units). Residues 267, 272, 274, 308, 324, and 327 are alanines in the wild-type protein. Each bar shows the average of 3 biological replicates and errors bars indicate 1 standard deviation from the mean.

NarL. One possible explanation for this observation is that less overall contacts are needed for class III activation because both αCTDs are bound by NarL. At class I activated promoters NarL is only predicted to bind one αCTD, which may necessitate more contacts.

Residues of αCTD identified by the alanine scan as being involved in class I activation by NarL were mapped onto the structural model of αCTD. Figure 5.4 shows that all residues identified are surface exposed, which supports the hypothesis that they are involved in interaction with NarL. As was observed with class III activation, αCTD residues implicated as necessary for class I activation cluster in a band around the outside of the αCTD, although one, residue 320, is towards the linker region. Residue 320 could be a target for NarL or, when mutated to alanine, might impair flexibility of the linker / αCTD and therefore the ability of NarL to bind the αCTD. As with class III activation, many of the residues that, when mutated to alanine, improve NarL-dependent class III activation fall in the DNA binding determinant of the αCTD. However, many residues also are not surface exposed and are inside the structure of the αCTD. This indicates that mutation of such residues may disrupt the 3D structure of the αCTD, although it is unclear by what mechanism that this improves NarL-dependent class I activation.

5.2.3. *Class II activated derivative of the* ogt *promoter*

Results so far indicate that NarL interacts with the αCTD when activating by class I and, III mechanisms. However, recall that transcription factors that activate by a class II mechanism are placed where they might contact the αNTD, σ or even β (see Section 1.10). Therefore, it is entirely possible that NarL makes no contacts with the αCTD when activating transcription by a class II mechanism.

Results in Figure 5.5 show expression of the ogt1052::*lacZ* fusion, which contains a class II activated derivative of the *ogt* promoter, in JCB3875. The NarL-dependent expression of the ogt1052::*lacZ* fusion is affected, positively and negatively, by mutation of residues in the
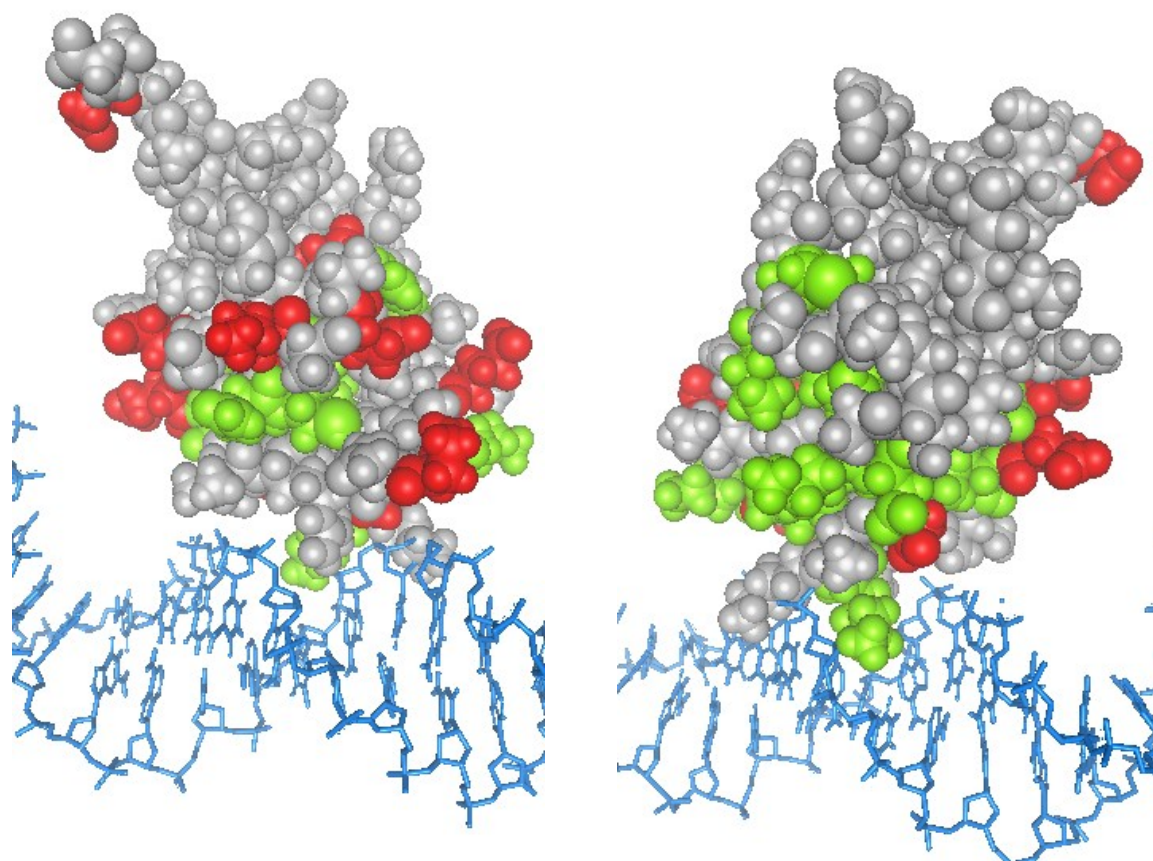
Figure 5.4 Model of αCTD indicating position of residues implicated in class I activation by NarL

Structure of αCTD (Jeon *et al.*, 1995) showing side chains identified as critical for class I activation by NarL. Two views of the structure, related by a 180° rotation on the vertical axis, are shown. Residues where alanine replacement impairs class I activation by NarL are coloured red and those that, when replaced, improve class I activation are coloured green.
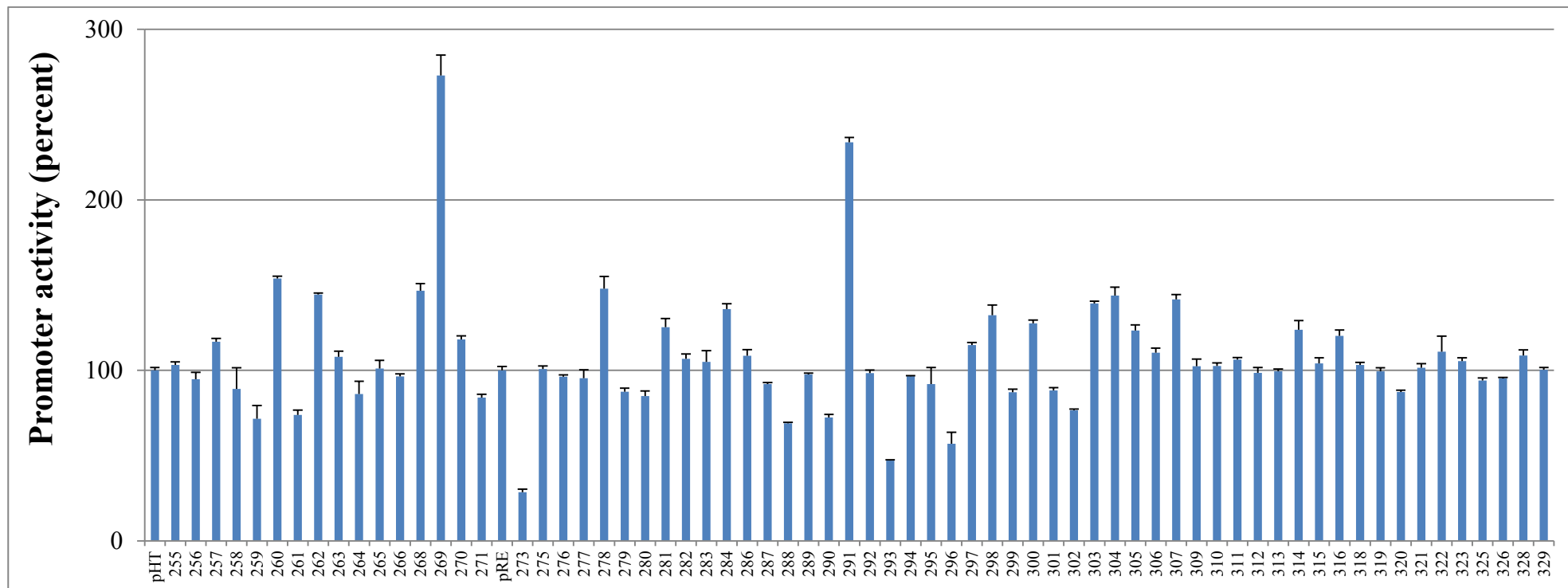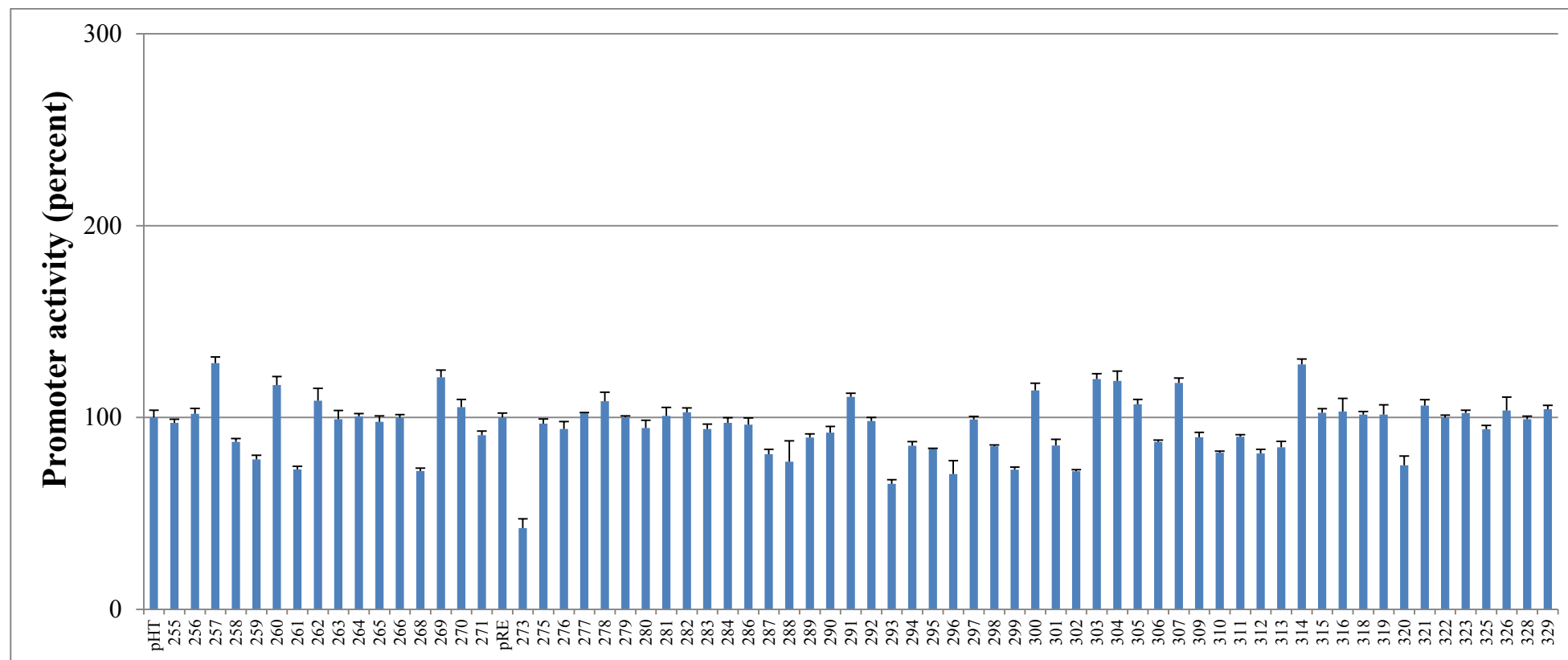
Figure 5.5 Residues of αCTD that are important for class II activation by NarL

JCB3875 cells containing pRW50 carrying the ogt1052 fragment were transformed with derivatives of plasmids pHTf1 (substitutions at 255 to 271 and 302) or pREII (substitutions at remaining positions). Each plasmid encoded an α derivative with a single alanine substitution between residues 255 and 329 as indicated in the figure. Galactosidase activities are expressed as percentages of the activity obtained with cells transformed with plasmids encoding wild-type α (100% = 405 Miller units). Residues 267, 272, 274, 308, 324, and 327 are alanines in the wild-type protein. Each bar shows the average of 3 biological replicates and errors bars indicate 1 standard deviation from the mean.

αCTD to alanine. Mutation of a residue is said to improve or impair activation if it results in average activity of the *ogt* promoter that is equal to or greater than 15% different from the activity of the *ogt* promoter in cells containing wild type RNA polymerase α subunit. Alanine substitution of residues 259, 261, 268, 273, 287, 288, 293, 295, 296, 299, 302, 310, 312, 313, and 320 impaired class II activation by NarL. Alanine substitution of residues 257, 260, 269 303, 304, 307 and 314 improved class II activation by NarL. These results indicate that NarL interacts with RNAP via the αCTD when activating by a class II mechanism.

Residues of αCTD identified by the alanine scan as being involved in class II activation by NarL were mapped onto the structural model of αCTD. Figure 5.6 shows that all residues identified are surface exposed, which supports the hypothesis that they are involved in interaction with NarL. As was observed with class I and class III activation, αCTD residues implicated as necessary for class II activation cluster in a band around the outside of the αCTD. The residues of αCTD implicated in class II activation are broadly similar to those implicated as important in class I activation. However, the residues from 307-313 of αCTD appear to have a small effect on class II activation by NarL, but no effect on class I activation by NarL. Class II activation by NarL appears to involve similar residues of αCTD as are involved in class III activation. However, as with class I activation, more residues of αCTD are needed for class II activation by NarL than are by class III activation by NarL. The results indicate that NarL contacts the αCTD when activating by a class II mechanism. However, the results do not eliminate the possibility that NarL contacts other parts of the transcription machinery when activating by a class II mechanism.

5.3. The CTD of NarL is competent to activate transcription by a class II mechanism

The results in Figures 5.1 – 5.6 indicate that NarL directly interacts with RNAP,
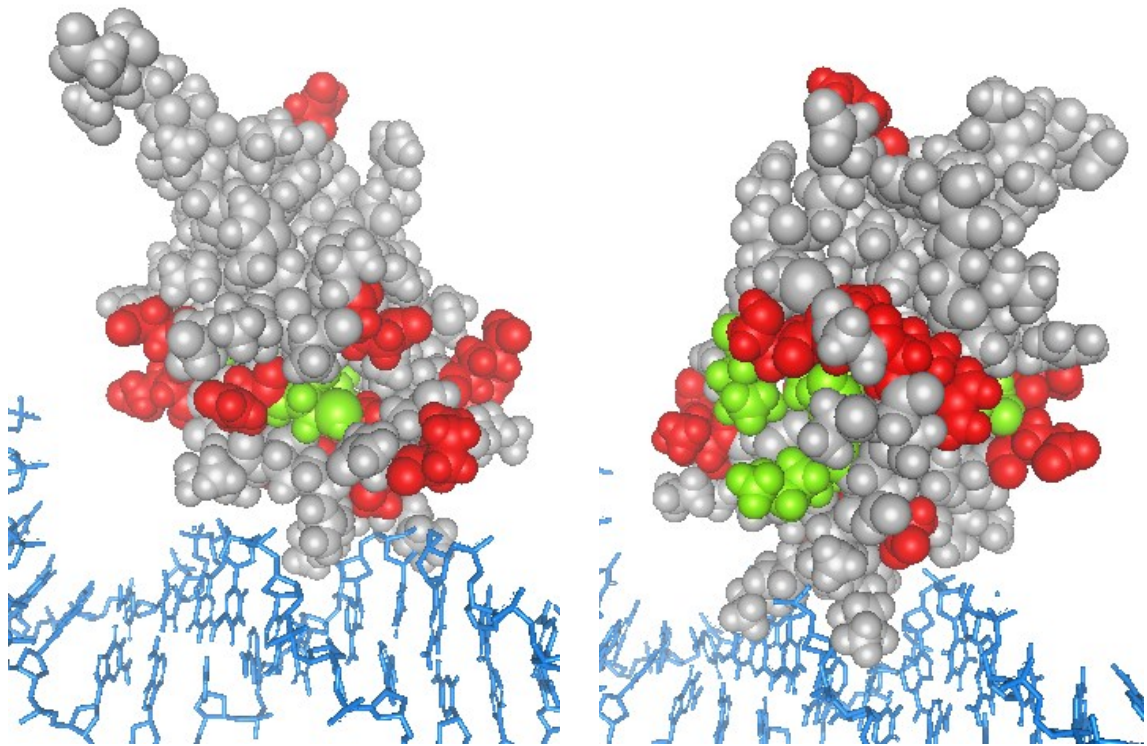
Figure 5.6 Model of αCTD indicating position of residues implicated in class II activation by NarL

Structure of αCTD (Jeon *et al.*, 1995)showing side chains identified as critical for class II activation by NarL. Two views of the structure, related by a 180° rotation on the vertical axis, are shown. Residues where alanine replacement impairs class II activation by NarL are coloured red and those that, when replaced, improve class II activation are coloured green.

potentially in order to recruit polymerase at the *ogt* promoter. Structural studies of NarL indicate that it has 2 distinct domains that are joined by a flexible linker and that the CTD is competent to bind DNA (see Figure 5.7, panel A) (Baikalov *et al.*, 1996, Maris *et al.*, 2002). To investigate what, if any, role the NarL NTD plays in activating transcription, full length NarL and the CTD of NarL were expressed from a plasmid and the activity of NarL-dependent promoters was measured in cells lacking chromosomal *narL* and *narP*.

The NarL CTD and full length NarL were amplified from the *E. coli* chromosome as NsiI-HindIII fragments (producing the NarL-CTD and NarL-Full fragments respectively). The NarL-CTD and NarL-Full fragments were cloned into the NsiI –HindIII sites of the high copy-number vector pJW15 Δ100 (see Figure 5.7, panel B). The plasmid pJW15 Δ100 carries an EcoRI-HindIII fragment that contains the *melR* promoter and the coding sequence of *melR* with an NsiI site engineered at codons 1 and 2. The NarL-CTD and NarL-Full fragments replaced the *melR* coding sequence and yielded plasmids pDLC1 and pDLC2 respectively (see Figure 5.7, panel C). This fused the NarL-CTD and NarL-Full fragments to the *melR* translation start.

JCB3884 cells, which lack NarL and NarP, were transformed with pRW50 carrying one of the ogt100, ogt1041 or ogt1052 fragments as promoter::*lacZ* fusions. Cells were also transformed with pJW15 Δ100, pDLC1 or pDLC2. Cells were then grown anaerobically in minimal salt medium with or without nitrate and expression of promoter::*lacZ* fusions was measured.

Results in Figure 5.8, panel A, show expression of the ogt100::*lacZ* fusion, which contains the class III activated *ogt* promoter. In cells that contain pJW15 Δ100, which does not carry *narL*, minimal expression of the ogt100::*lacZ* fusion is observed and nitrate fails to induce expression. In cells that contain pDLC1, which expresses the NarL-CTD fragment,

Figure 5.7 Structure of NarL and diagrams of NarL fragments

Panel A, The amino acid sequence of NarL. Shaded regions indicate residues involved in the NarL secondary structure and the identity of those structural elements is given above the alignment. Sequence underlined in red is that reported as constituting the NTD and that underlined in green constitutes the CTD. Linker helix and flexible linker are denoted. Adapted from Baikalov *et al.* (1996).

Panel B shows a map of the pJW15 Δ100 plasmid with the region between the EcoRI and HindIII sites enlarged and diagrams of the NarL-CTD fragment and NarL-Full fragment. A capital 'E' denotes an EcoRI site, a capital 'N' denotes an NsiI site and a capital 'H' denotes a HindIII site. The bent arrow indicates the *melR* promoter. Curved arrows indicate the *ampR* and *galK* ORFs carried by pJW15 Δ100. In the diagrams of fragments, the *narL* coding region is indicated by a labelled grey box and the specific domains marked with a labelled clear box. Diagram is not to scale

Panel C is a table of plasmids used in this section. The table lists the vector used in construction of the plasmid, the relevant DNA fragment carried in the plasmid and the promoter driving expression of the gene contained within the fragment.

A

β1　　*　　α1　　　　β2　　*　　α2　　　　β3　　*　　α3

NarL Eco　MSNQEPATILLIDDHPMLRTGVKQLISMAPDITVVGEASNGEQGIELAESLDPDLILLDLNMPG...MNGLETLDKLRE

NTD

β4　*　α4　　　β5　*　　α5　　　　　α6

NarL Eco　KSLSGRIVVFSVSNHEEDVVTALKRGADGYLLKDMEPEDLLKALHQAAAGEMVLSEALTPVLAASLRANRATTERDVN

Linker helix　**Flexible linker**

α7　　　　α8　　　　　α9　　　　　α10

NarL Eco　QLTPRERDILKLIAQGLPNKMIARRLDITESTVKVHVKHMLKKMKLKSRVEAAVWVHQERIF

CTD　　**DNA binding element**

B

E　P*melR*　N　　　　H

melR

N　　　　　H
narL
CTD
NarL-CTD
fragment

N　　　　　　H
NTD　*narL*　CTD
NarL-Full
fragment

AmpR
GalK
pJW15 Δ100

C
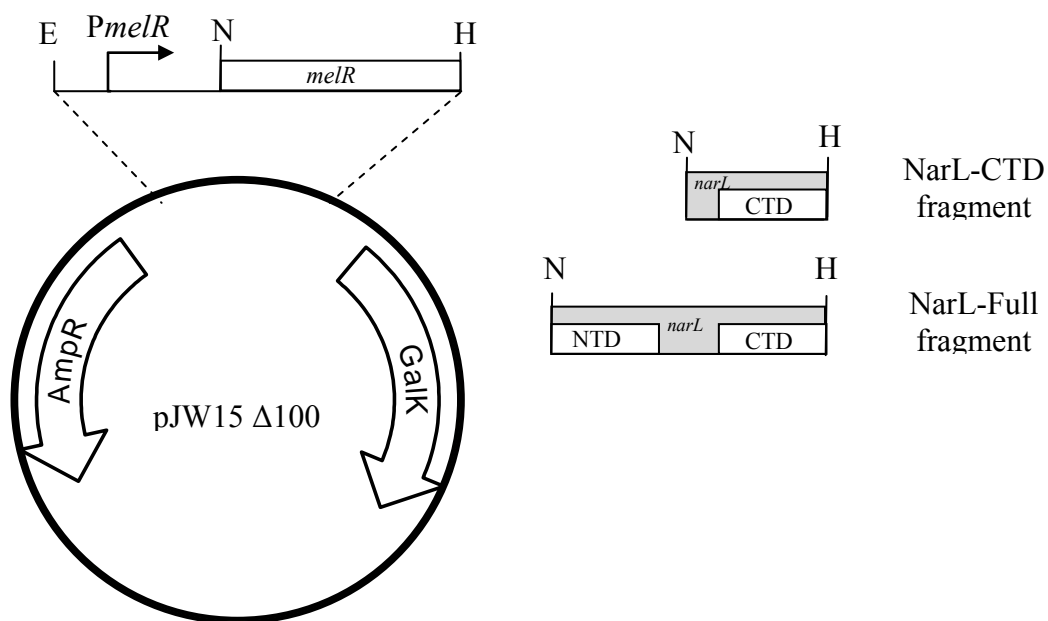
Plasmid	Background	Carries fragment	Promoter
pDLC1	pJW15 Δ100	NarL-CTD	P*melR*
pDLC2	pJW15 Δ100	NarL-Full	P*melR*

181
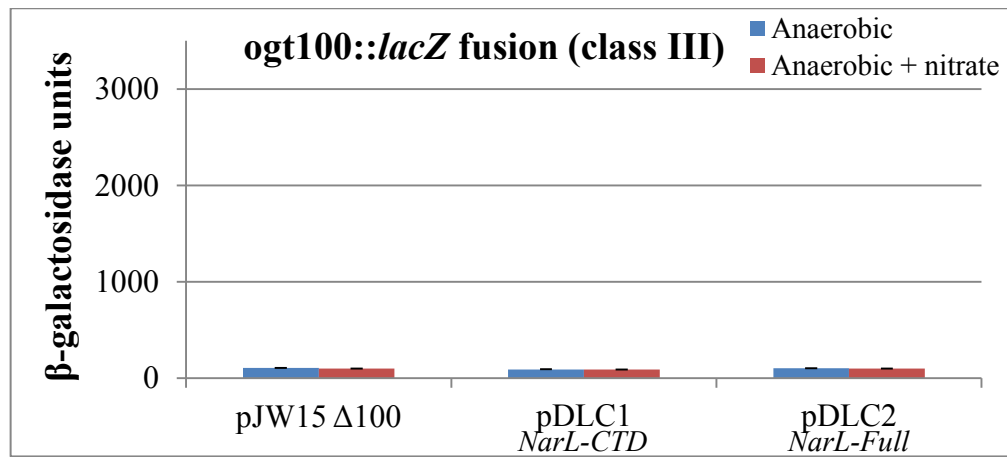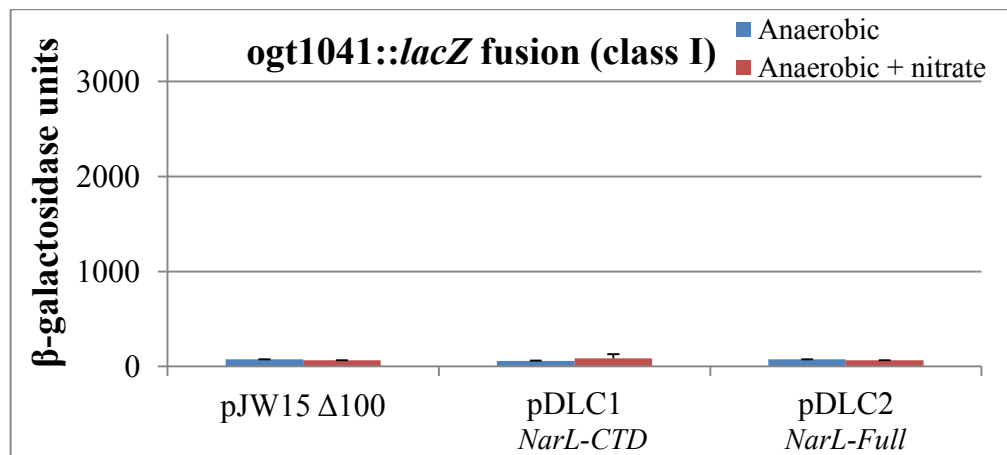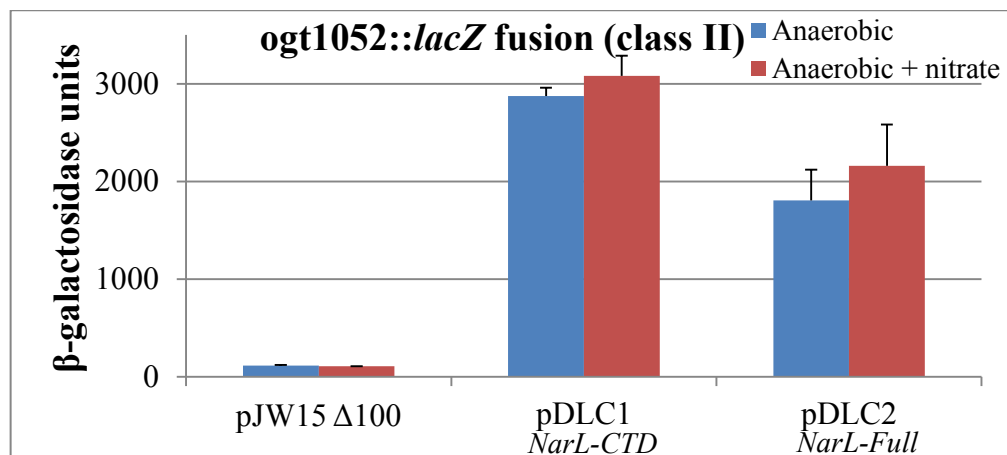
Figure 5.8 Expression of promoter::*lacZ* fusions with plasmid encoded NarL CTD and full length protein

Figure shows expression of promoter::*lacZ* fusions measured in JCB3884 cells containing empty vector pJW15 Δ100 or derivatives that express the NarL CTD or full-length protein. NarL fragment carried by plasmid is given below plasmid name in italicised text. Cells were grown anaerobically or anaerobically in minimal salt medium supplemented with 20 mM sodium nitrate. Data shown are averages from at least 3 biological repeats, and error bars show one standard deviation from the mean. Panel A shows expression of the ogt100::*lacZ* fusion, carried by pRW50, panel B shows expression of the ogt1031::*lacZ* fusion, carried by pRW50 and panel C shows expression of the ogt1052::*lacZ* fusion carried by pRW50.

minimal expression is observed and nitrate fails to induce expression. This suggests that the CTD of NarL is not competent to activate transcription by a class III mechanism. Minimal expression of the ogt100::*lacZ* fusion is observed in cells that contain pDLC2, which carries the NarL-Full fragment, and nitrate fails to induce expression. This indicates that the NarL expressed from the NarL-Full fragment is not able to activate transcription by a class III mechanism. This observation disagrees with the results of Squire *et al.* (2009) and previous results of this work, which indicate that the wild-type NarL is able to activate by a class III mechanism.

Figure 5.8, panel B, shows expression of the ogt1041::*lacZ* fusion, which contains a NarL-dependent, class I activated promoter. In cells that contain pJW15 Δ100, which does not carry *narL*, minimal expression of the ogt100::*lacZ* fusion is observed and nitrate fails to induce expression. In cells that contain e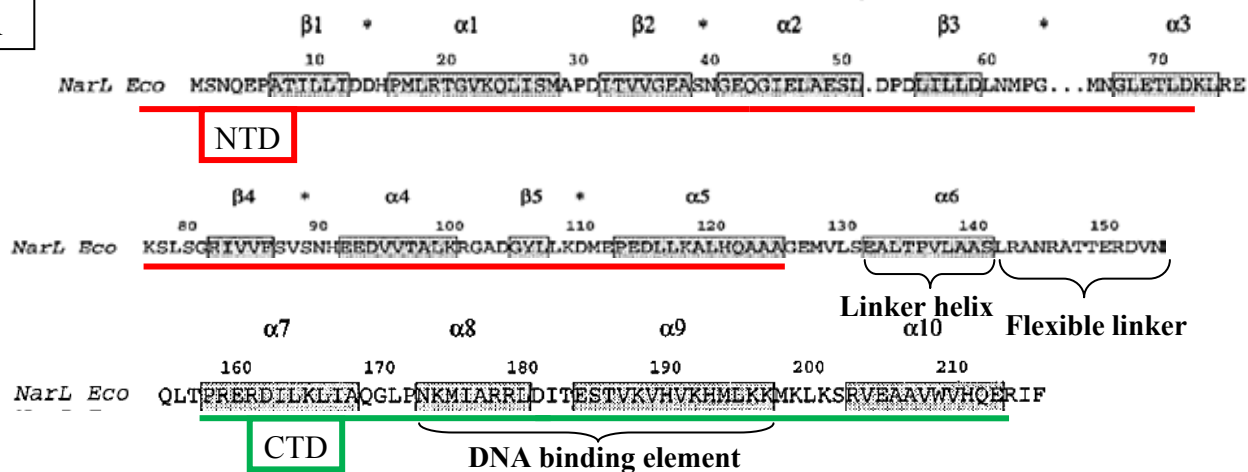ither pDLC1 or pDLC2, which carry the NarL-CTD and NarL-Full fragments respectively, minimal expression is observed and nitrate fails to induce expression. This indicates that neither the CTD of NarL, nor the NarL expressed from the NarL-Full fragment, is competent to activate transcription by a class I mechanism. This observation disagrees with the previous results of this work, which indicate that the wild-type NarL is able to activate by a class I mechanism.

Figure 8, panel C, shows the expression of the ogt1052::*lacZ* fusion, which is NarL-dependent and is activated by a class II mechanism. In cells that contain pJW15 Δ100, which does not carry *narL*, minimal expression is observed and nitrate fails to induce expression. In cells that contain pDLC1, which carries the NarL-CTD fragment, maximal expression of the ogt1052::*lacZ* fragment is observed although nitrate fails to induce expression further. This indicates that the CTD of NarL is competent to activate transcription by a class II mechanism and that the CTD does not itself require activation in response to nitrate in order to do so. A high level of expression is observed in cells that contain pDLC2, which carries the NarL-Full

fragment, and nitrate fails to induce further. This agrees with previous results that indicate that NarL can activate transcription by a class II mechanism. Previous results also indicate that, in JCB387 (L$^+$ P$^+$) cells, expression of the ogt1052::lacZ fusion is high in the absence of nitrate. This is possibly because the ogt1052 fragment contains a consensus DNA site for NarL at an optimal location, and therefore residually active NarL would bind and activate transcription.

These results indicate that the NarL CTD is competent to activate transcription by a class II mechanism but not by class I or class III mechanisms. These results also indicate that the NarL encoded by the NarL-Full fragment is able to activate by a class II mechanism but not by class I or class III mechanisms, an observation which is at odds with earlier results. There are two possible explanations for this discrepancy. First, in order to clone *narL*, the second codon was changed from AGT, which encodes Serine, to CAT, which encodes Threonine. It is possible that this modification, or another product of the cloning strategy used, prevents NarL from adopting its correct secondary and tertiary structure. Alternatively, it is possible that NarX does not function correctly in JCB3884 cells. This is because *narX* and *narL* are in an operon and so when JCB3875 (from which JCB3884 was derived) was constructed by disruption of *narL* with a transposon, *narX* expression may have been affected. Therefore NarL would not be phosphorylated in response to nitrate and would be unable to activate transcription.

5.4. Plasmid encoded NarL functions similarly when supplied *in trans*

To investigate the inability of plasmid-encoded full-length NarL to activate class I and III promoters, the experiment was repeated in the *narL$^+$ narP$^+$* strain JCB387. As before, expression of promoter::*lacZ* fusions was measured following anaerobic growth in minimal salt medium with or without nitrate.

Figure 5.9, panel A, shows expression of the ogt100::*lacZ* fusion, which contains the class
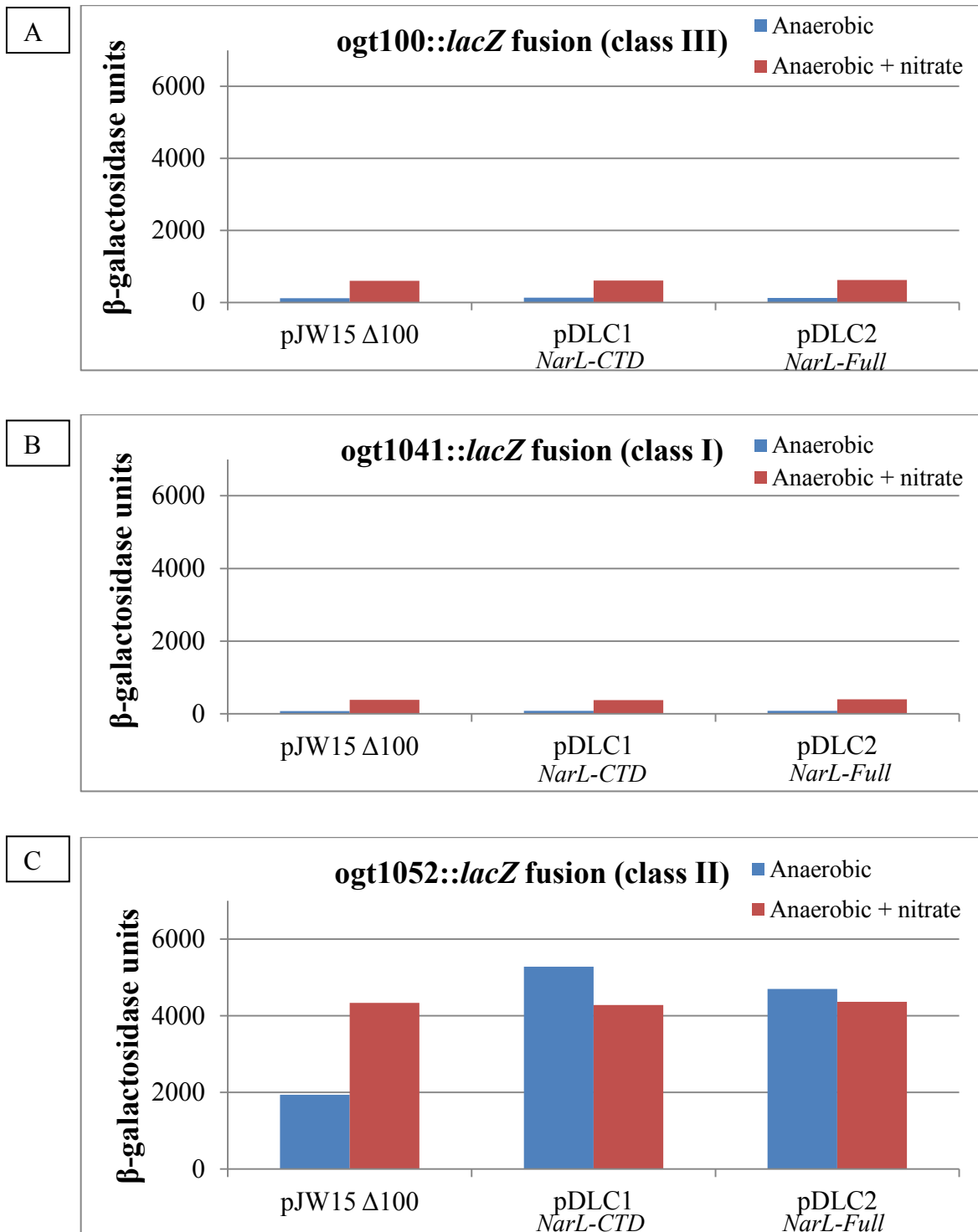
Figure 5.9 Expression of promoter::*lacZ* fusions with plasmid encoded NarL CTD and full length protein supplied in *trans*

Figure shows expression of promoter::*lacZ* fusions measured in JCB387 cells containing control plasmid pJW15 Δ100 or plasmids that express the NarL CTD or full-length protein. NarL fragment carried by plasmid is given below plasmid name in italicised text. Cells were grown anaerobically or anaerobically in minimal salt medium supplemented with 20 mM sodium nitrate. Panel A shows expression of the ogt100::*lacZ* fusion, carried by pRW50, Panel B shows expression of the ogt1041::*lacZ* fusion, carried by pRW50 and Panel C shows expression of the ogt1052::*lacZ* fusion carried by pRW50

III wild-type *ogt* promoter. In cells that contained the empty pJW15 Δ100 vector plasmid, minimal expression was observed in the absence of nitrate and expression was induced by nitrate as expected. In cells that contained pDLC1 or pDLC2, which carry the NarL-CTD fragment and the NarL-Full fragment respectively, the pattern and levels of expression of the ogt100::*lacZ* fusion was identical to that of cells that contain pJW15 Δ100. This indicates that neither the NarL CTD nor the NarL encoded by the NarL-Full fragment can activate transcription by a class III mechanism when supplied in *trans*.

Figure 5.9, panel B, shows expression of the ogt1041::*lacZ* fusion, which is activated by a class I mechanism. In cells that contain the empty pJW15 Δ100 vector plasmid, minimal expression was observed in the absence of nitrate and expression was induced by nitrate. In cells that contain pDLC1 or pDLC2, which carry the NarL-CTD fragment and the NarL-Full fragment respectively, the pattern and levels of expression of the ogt1041::*lacZ* fusion were identical to that of cells that contained pJW15 Δ100. This indicates that neither the NarL CTD nor the NarL encoded by the NarL-Full fragment can activate transcription by a class I mechanism when supplied in *trans*.

Figure 5.9, panel C, shows expression of the ogt1052::*lacZ* fusion, which is activated by a class II mechanism. In cells that contain empty pJW15 Δ100 vector plasmid, expression of the ogt1052::*lacZ* fusion was high and nitrate induced further. In cells that contain pDLC1, which carries the NarL-CTD fragment, expression of the ogt1052::*lacZ* fusion was maximal anaerobically and nitrate did not induce further. This indicates that the CTD of NarL contributes to activity of a class II promoter when supplied in *trans*. In cells that contain pDLC2, which carries the Narl-Full fragment, expression of the ogt1052::*lacZ* fusion was high and similar whether cells were grown in the presence or absence of nitrate. This indicates that the NarL encoded by the NarL-Full fragment can activate by a class II mechanism when supplied in *trans*. These results confirm that the NarL CTD can activate by a class II

mechanism when supplied in *trans,* but that it cannot activate by class I or III mechanisms. Results also show that the NarL encoded by the NarL-Full fragment is able to activate by a class II mechanism but not by class I or III mechanisms. As functional NarX was present in the JCB387 cells used, these results suggest that the NarL-Full fragment does not produce a functional NarL protein.

5.5. Plasmid encoded NarXL can activate by class I, II and III mechanisms

The above results indicate that the NarL-Full fragment in pDLC2 does not produce a NarL protein that can function similarly to the wild-type NarL. To mitigate any errors introduced by cloning *narL* downstream of the *melR* promoter and translation start, the *narXL* operon was ligated into a plasmid complete with its own promoter region.

The NarXL fragment was prepared by amplifying the *narXL* operon from the chromosome as an EcoRI-HindIII DNA fragment. Plasmid pJW15 Δ100 (see Figure 5.7, panel B) was prepared with EcoRI and HindIII to remove *melR* and its regulatory region and the NarXL fragment was ligated in its place. However, it was not possible to recover transformants and it has been suggested that such a construct is toxic to the cell because of the high-copy nature of pJW15 Δ100 (Douglas Browning, personal communication). Thus, the NarXL, NarL-CTD and NarL-Full fragments were cloned into the lower copy-number plasmid pLG339 (see Figure 5.10).

To make the tetracycline and kanamycin resistant pLG339 compatible with pRW50, it was necessary to clone into, and therefore disrupt, the tetracycline resistance gene. BamHI-EcoRI DNA fragments *be*NarL-CTD and *be*NarL-Full were constructed by amplifying the EcoRI-HindIII fragments from pDLC1 and pDLC2 (see Figure 5.7) and changing the restriction sites (see Figure 5.10, panel A). The BamHI-EcoRI fragments were then cloned into prepared pLG339 producing pDLC3 and pDLC4 (see Figure 5.10, panel B). The *be*NarXL fragment was prepared by amplification from the chromosome and was cloned into
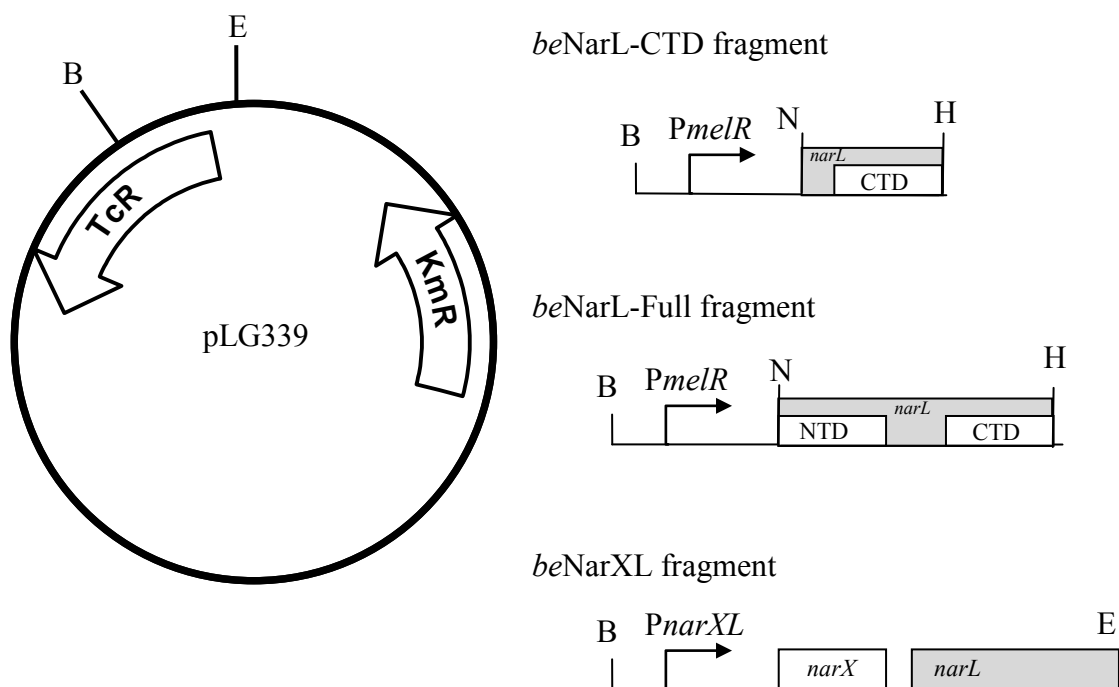
Figure 5.10 pLG339 and cloned NarL fragments

Panel A, shows a map of the pLG339 plasmid and diagrams of the *be*NarL-CTD, *be*NarL-Full and *be*NarXL DNA fragments. A capital 'B' denotes a BamHI site, a capital 'E' denotes an EcoRI site and a capital 'N' denotes an NsiI site. The bent arrow indicates a promoter, which is labelled. The curved arrows labelled TcR indicates a tetracycline resistance gene and the curved arrow labelled KmR indicates a kanamycin resistance gene. A shaded box indicates the *narL* coding region. Diagram is not to scale

Panel B is a table of plasmids used in this section. The table lists the vector used in construction of the plasmid, the relevant DNA fragment carried in the plasmid and the promoter driving expression of the gene or operon contained within the fragment.

prepared pLG339, producing pDLC5. JCB3884 cells, which do not contain NarL or NarP were transformed with either pDLC3, pDLC4 or pDLC5. Cells were then transformed with pRW50 carrying ogt100, ogt1041 or ogt1052 DNA fragments. Cells were grown anaerobically with or without nitrate and expression of promoter::*lacZ* fusions was measured

Figure 5.11, panel A, shows expression of the ogt100::*lacZ* fusion, which contains the wild-type *ogt* promoter and is activated by NarL by a class III mechanism. In cells that contained empty pLG339 ΔRS, expression of the ogt100::*lacZ* fusion was minimal and nitrate did not induce. In cells that contained pDLC3 or pDLC4, which carry the NarL-CTD and NarL-Full fragments respectively, expression of the ogt100::*lacZ* fusion was identical to that in cells that contained pLG339 ΔRS. This indicates that the NarL-CTD and the NarL encoded by the NarL-Full fragment are unable to activate by a class III mechanism when expressed from pLG339. In cells that contain pDLC5, which carries the NarXL fragment, expression of the ogt100::*lacZ* fusion is minimal when cells are grown anaerobically and nitrate induces expression. This indicates that the NarL encoded by the NarXL fragment in pDLC5 is able to activate by a class III mechanism and in a nitrate dependent manner.

Figure 5.11, panel B, shows expression of the ogt1041::*lacZ* fusion, which contains a NarL-dependent promoter that is activated by a class I mechanism. In cells that contained empty pLG339 ΔRS, expression of the ogt1041::*lacZ* fusion was minimal and nitrate did not induce expression further. In cells that contained pDLC3 or pDLC4, which carry the NarL-CTD and NarL-Full fragments respectively, expression of the ogt1041::*lacZ* fusion was identical to that in cells that contained pLG339 ΔRS. This indicates that the NarL-CTD and the NarL encoded by the NarL-Full fragment are unable to activate by a class I mechanism when expressed from pLG339. In cells that contained pDLC5, which carries the NarXL fragment, expression of the ogt1041::*lacZ* fusion was minimal when cells are grown anaerobically and nitrate induced expression. This indicates that the NarL encoded by the

189
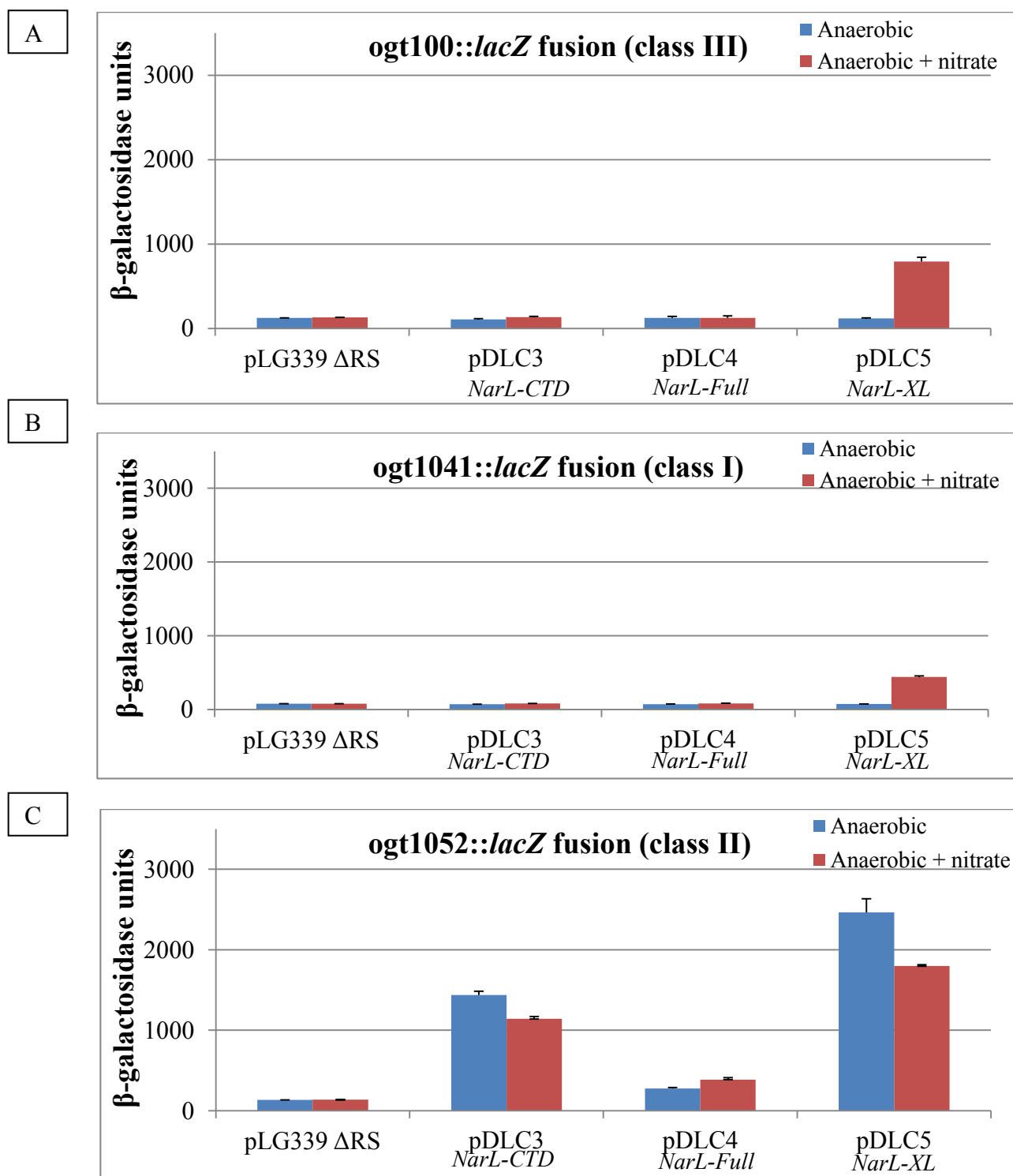
Figure 5.11 Expression of promoter::*lacZ* fusions with plasmid encoded NarL CTD and full-length protein

Figure shows expression of promoter::*lacZ* fusions measured in JCB3884 cells containing control plasmid pLG339 or plasmids that express the NarL CTD (pDLC3), full length NarL (pDLC4) or the *narXL* operon (pDLC5). NarL fragment carried by plasmid is given below plasmid name in italicised text. Cells were grown anaerobically or anaerobically in minimal salt medium supplemented with 20 mM sodium nitrate. Data shown are averages from at least 3 biological repeats, and error bars show one standard deviation from the mean. Panel A shows expression of the ogt100::*lacZ* fusion, carried by pRW50, panel B shows expression of the ogt1041::*lacZ* fusion, carried by pRW50 and panel C shows expression of the ogt1052::*lacZ* fusion carried by pRW50.

NarXL fragment is able to activate by a class I mechanism and in a nitrate dependent manner.

Figure 5.11, panel C, shows expression of the ogt1052::*lacZ* fusion, expression of which is activated by NarL by a class II mechanism. In cells that contained empty pLG339 ΔRS, expression of the ogt1052::*lacZ* fusion was minimal and nitrate failed to induce. In cells containing pDLC3, which carries the NarL-CTD fragment, expression of the ogt1052::*lacZ* fragment was high when the cells are grown anaerobically and nitrate did not induce expression further. Absolute levels of expression of the ogt1052::*lacZ* fusion were lower in cells containing pDLC3, which is a low copy number plasmid, than in cells containing pDLC1, which is a high copy number plasmid. This supports the hypothesis that the NarL CTD can activate by a class II mechanism. In cells that contained pDLC4, which carries the NarL-Full fragment, expression of the ogt1052::*lacZ* fusion was higher than in cells that contained pLG339 ΔRS, but lower than in cells that contained pDLC3. This further indicates that the NarL encoded by the NarL-Full fragment can activate by a class II mechanism. In cells that contained pDLC5, which carries the NarXL fragment, expression of the ogt1052::*lacZ* fusion was maximal anaerobically and nitrate failed to induce further. This indicates that the NarL encoded by the NarXL fragment is able to induce by a class II mechanism and that it can do this in the absence of nitrate.

These results show that it is possible to express NarL from a plasmid and that it acts similarly to wild-type, chromosomally encoded NarL. Furthermore, these results indicate that the NarL CTD is able to activate by a class II mechanism, while full length NarL is able to activate by class I, II and III mechanisms.

5.6. The NarP CTD can activate by a class II mechanism.

Previous results of this work indicate that NarP can activate by a class II mechanism but not by class I or III mechanisms. To investigate whether the CTD of NarP is able to activate transcription, the CTD was expressed from plasmid pJW15 Δ100. The structure of NarP has

not been solved so cloning was based on an alignment with NarL (see Figure 5.12, panel A). To ensure that a construct expressing a functional CTD was obtained, 3 fragments were prepared that start from different codons: NarP-CTD146, which starts from the same relative codon as the NarL-CTD fragment; NarP-CTD138 that starts from codon 138 and NarP-CTD148 that starts from codon 148. Fragments were amplified from the chromosome as NsiI-HindIII fragments and ligated into Nsi-HindIII prepared pJW15 Δ100 producing pDLC6 (NarP-CTD146), pDLC8 (NarP-CTD138) and pDLC9 (NarP-CTD148) (see, panel C). JCB3884 cells, which lack NarL and NarP, were transformed with NarL / NarP-dependent promoter::*lacZ* fusions and either plasmid pJW15 Δ100, pDLC6, pDLC8 or pDLC9. Expression of the promoter *lacZ* fusions was measured following anaerobic growth in minimal salt medium.

Results in Figure 5.13, panel A, show expression of the ogt100::*lacZ* fusion, which contains the class III activated wild-type *ogt* promoter. In cells that contained the empty pJW15 Δ100 vector, expression of the ogt100::*lacZ* fusion was minimal. In cells containing pDLC6, pDLC8 or pDLC9, which carry the NarP-CTD146, NarP-CTD138 and NarP-CTD148 fragments respectively, expression of the ogt100::*lacZ* fusion was also minimal. This indicates that none of the NarP CTDs expressed by the plasmids are able to activate by a class III mechanism.

Results in Figure 5.13, panel B, show expression of the ogt1041::*lacZ* fusion, which contains a NarL-dependent promoter that is activated by a class I mechanism. In cells that contained the empty pJW15 Δ100 vector, expression of the ogt1041::*lacZ* fusion was minimal. In cells containing pDLC6, pDLC8 or pDLC9, which carry the NarP-CTD146, NarP-CTD138 and NarP-CTD148 fragments respectively, expression of the ogt1041::*lacZ* fusion was also minimal. This indicates that none of the NarP CTDs expressed by the plasmids are able to activate by a class I mechanism.

Figure 5.12 Structure of NarP and DNA fragments used in this section

Panel A, Alignment of the amino acid sequence of NarL and NarP. Shaded regions indicate residues involved in the NarL secondary structure and the identity of those structural elements is given above the alignment. Sequence underlined in red is that reported as constituting the NTD and that underlined in green constitutes the CTD. Linker helix and flexible linker are denoted. Adapted from Baikalov *et al.* (1996).

anel B, shows a map of the pJW15 Δ100 plasmid with the region between the EcoRI and HindIII sites enlarged and diagrams of the NarP-CTD146, NarP-CTD138 and NarP-CTD148 fragments. A capital 'E' denotes an EcoRI site, a capital 'N' denotes an NsiI site and a capital 'H' denotes a HindIII site. The bent arrow indicates the *melR* promoter. Curved arrows indicate the *ampR* and *galK* ORFs carried by pJW15 Δ100. In the diagrams of fragments, the *narP* coding region is indicated by a labelled grey box and the specific domains marked with a labelled clear box. Diagram is not to scale

Panel C is a table of plasmids used in this section. The table lists the vector used in construction of the plasmid, the relevant DNA fragment carried in the plasmid and the promoter driving expression of the gene contained within the fragment.

A

| β1 | * | α1 | β2 | * | α2 | β3 | * | α3 |

NarL Eco MSNQEPATILLIDDHPMLRTGVKQLISMAPDITVVGEASNGEQGIELAESL.DPDLILLDLNMPG...MNGLETLDKLRE
NarP Eco MPEATPFQVMIVDDHPLMRRGVRQLLELDPGSEVVAEAGDASAIDLANRL.DIDVILLDLNMKG...MSGLDTLNALRR

NTD

| β4 | * | α4 | β5 | * | α5 | α6 |

NarL Eco KSLSGRIVVFSVSNHEEDVVTALKRGADGYLLKDMPEDLLKALHQAAAGEMVLSEALTPVLAASLRANRATTERDVN
NarP Eco DGVTAQIIILTVSDASSDVPALIDAGADGYLLKDSDPEVLLEAIRAGAKGSKVFSERVNQYLREREMFGAEEDPFS..

Linker helix Flexible linker

| α7 | α8 | α9 | α10 |

NarL Eco QLTPRERDILKLIAQGLPNKMIARRLDITESTVKVHVKHMEKKMKLKSRVEAAVWVHQERIF
NarP Eco VLTERELDVLHELAQGLSNKQIASVLNISEQTVKVHIRNLLRKLNVRSRVAATILFLQQRGAQ

CTD DNA binding element

B

E P*melR* N H
melR

pJW15 Δ100
Amp*R*
Gal*K*

N H
narP CTD
NarP-CTD146 fragment

N H
narP CTD
NarP-CTD138 fragment

N H
narP CTD
NarP-CTD148 fragment

C

Plasmid	Background	Fragment	Promoter
pDLC6	pJW15 Δ100	NarP-CTD146	P*melR*
pDLC8	pJW15 Δ100	NarP-CTD138	P*melR*
pDLC9	pJW15 Δ100	NarP-CTD148	P*melR*

194
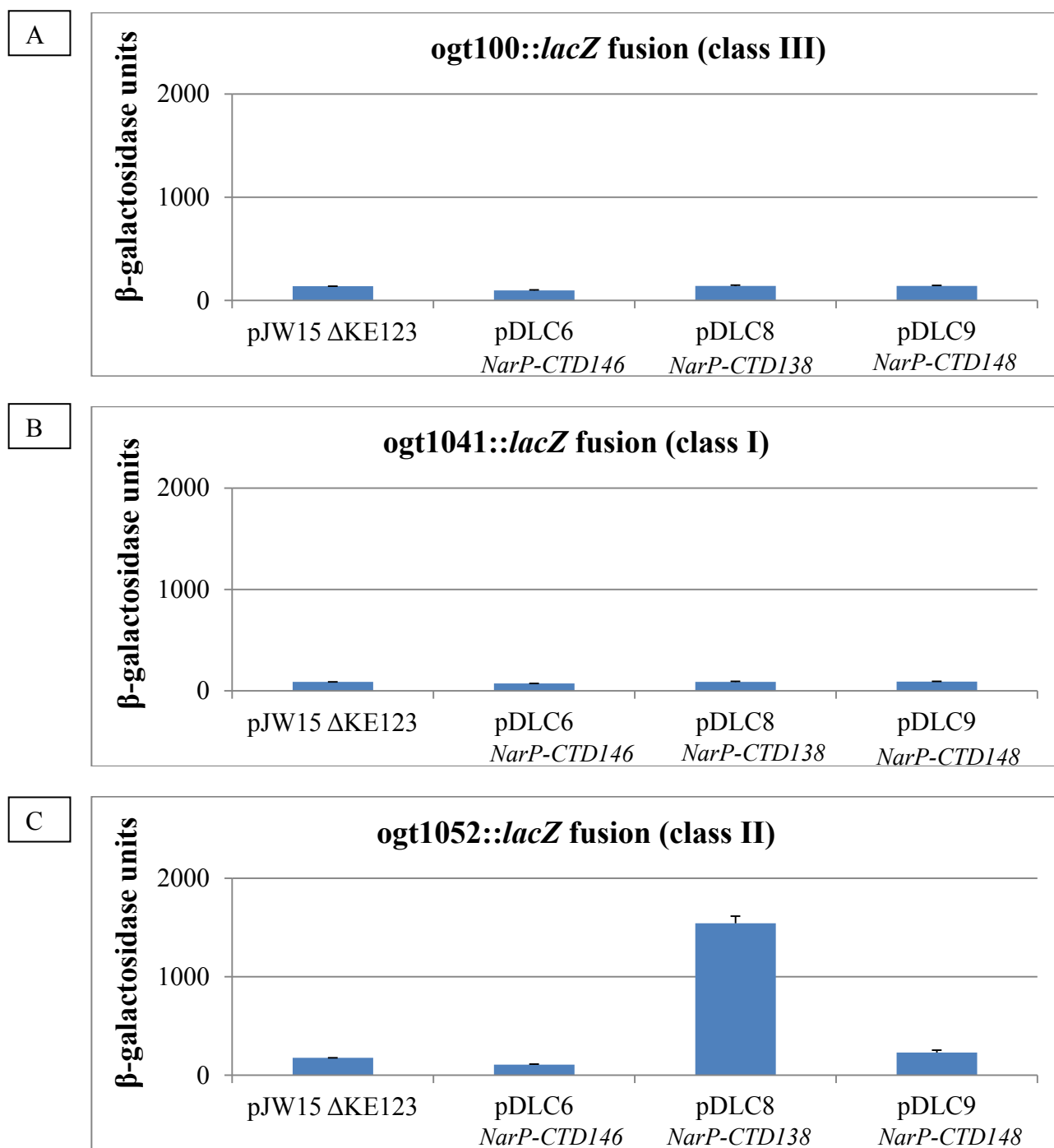
Figure 5.13 Expression of promoter::*lacZ* fusions with plasmids containing fragments of the NarP CTD

Figure shows expression of promoter::*lacZ* fusions measured in JCB3884 cells containing control plasmid pLG339 or plasmids that express fragments of the NarP CTD. NarP fragment carried by plasmid is given below plasmid name in italicised text. Cells were grown anaerobically in minimal salt medium. Data shown are averages from at least 3 biological repeats, and error bars show one standard deviation from the mean. Panel A shows expression of the ogt100::*lacZ* fusion, carried by pRW50, panel B shows expression of the ogt1041::*lacZ* fusion, carried by pRW50 and panel C shows expression of the ogt1052::*lacZ* fusion carried by pRW50.

Results in Figure 5.13, panel C, show expression of the ogt1052::*lacZ* fusion, which is activated by a class II mechanism. In cells that contained the empty pJW15 Δ100 vector, expression of the ogt1052::*lacZ* fusion was minimal. In cells that contained pDLC6, which carries the NarP-CTD146 fragment, minimal expression of the ogt1052::*lacZ* fusion was observed. In cells that contained pDLC8, which carries the NarP-CTD138 fragment, high expression of the ogt1052::*lacZ* fusion was observed. This indicates that the NarP CTD encoded by the NarP-CTD138 fragment is able to activate transcription by a class II mechanism. However, in cells that contained pDLC9, which carries the NarP-CTD148 fragment, expression of the ogt1052::*lacZ* fragment was minimal, indicating that the NarP-CTD encoded by the NarP-CTD148 fragment is not able to activate transcription by a class II mechanism.

Taken together, these results indicate that the CTD of NarP is able to activate transcription by a class II mechanism, but not by class I or II mechanisms. Results also suggest that the NarP CTD that is encoded by the NarP-CTD138 fragment is functional, whilst the NarP CTDs encoded by the NarP-CTD146 fragment and the NarP-CTD148 fragment are not.

5.7. Plasmid encoded full-length NarP can activate by a class II mechanism

Previous results indicate that chromosomally expressed NarP can only activate by a class II mechanism and that this activation requires only the CTD of NarP. To ensure that this is also true of plasmid-expressed full-length NarP, the *narP* gene, complete with its own promoter region, was cloned into pLG339 (see Figure 5.14, panel A). The BamHI-EcoRI NarP-op fragment was constructed by amplification of the *narP* gene from the chromosome. The NarP-op fragment was ligated into BamHI-EcoRI-cut pLG339, producing pDLC14 (see Figure 5.14, panel B). JCB3884 cells, which do not contain NarL or NarP, were transformed with either pDLC14 or empty pLG339 ΔRS. Transformants were then transformed with pRW50 carrying derivatives of the *ogt* promoter that are activated by class I, II or III
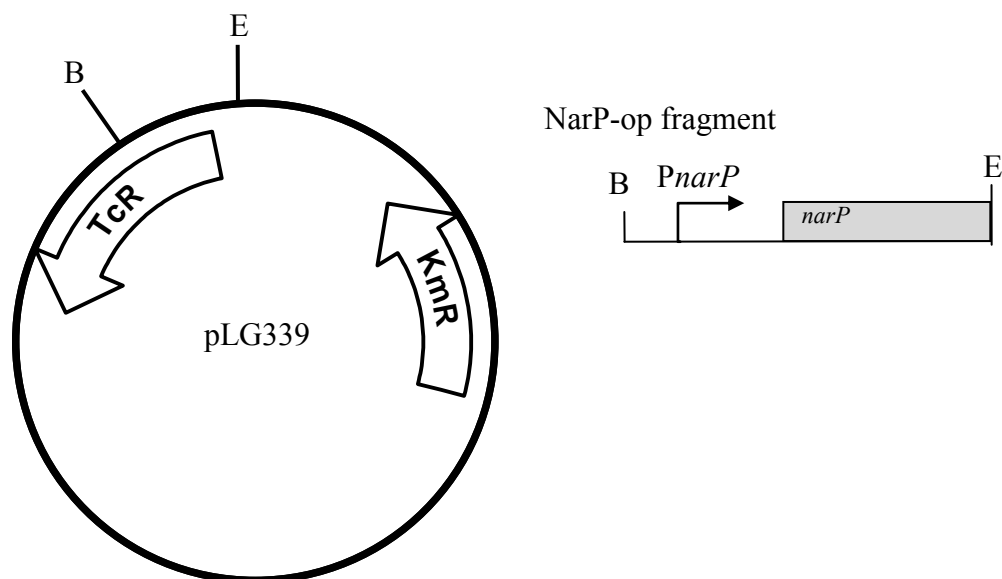
Figure 5.14 Map of pLG339 and NarP fragment used in this section

Panel A, shows a map of the pLG339 plasmid and a diagram of NarP-op DNA fragment. A capital 'B' denotes a BamHI site, a capital 'E' denotes an EcoRI site. The bent arrow indicates the *narP* promoter, which is labelled. The curved arrows labelled TcR indicates a tetracycline resistance gene and the curved arrow labelled KmR indicates a kanamycin resistance gene. A shaded box indicates the *narL* coding region. Diagram is not to scale

Panel B is a table of the plasmid used in this section. The table lists the vector used in construction of the plasmid, the relevant DNA fragment carried in the plasmid and the promoter driving expression of the gene or operon contained within the fragment.

mechanism. The β-Galactosidase activity was measured following cells growth in minimal salt medium, with or without nitrate.

Figure 5.15, panel A, shows expression of promoter::*lacZ* fusions in cells containing empty pLG339 ΔRS. Expression of all promoter::*lacZ* fusions was minimal, and, as expected, nitrate did not induce expression. Figure 5.15, panel B, shows the expression of promoter::*lacZ* fusions in cells that contain pDLC14, which carries the complete *narP* gene controlled by its own promoter. Expression of the ogt100::*lacZ* fusion, which contains the wild-type *ogt* promoter and is activated by a class III mechanism, was minimal in the absence of nitrate, and nitrate failed to induce. Expression of the ogt1041::*lacZ* fusion, which is activated by a class I mechanism, was minimal in the absence of nitrate and nitrate failed to induce. However, expression of the ogt1052::*lacZ* fusion, which is activated by a class II mechanism, was high when cells were grown anaerobically and was induced by nitrate.

These results agree with previous results and confirm that NarP is able to activate transcription only by a class II mechanism. These results also indicate that the NarP expressed by pDLC14 is functional.

5.8. Discussion

5.8.1. *NarL interacts directly with the αCTD*

There are many examples in the literature of promoters where NarL regulates transcription. However, proving a mechanism of regulation can be difficult and molecular mechanisms of NarL have only been detailed in a few cases. An example of such a case is that of the *E. coli nirB* promoter, where transcription is induced by NarL. Investigation showed that NarL does not activate in the classical sense, but instead overcomes the repressors IHF and Fis by 'remodelling the nucleoprotein assembly'(Browning *et al.*, 2004). NarL is likely able to do this because when it binds to DNA it both bends the DNA by 40° and distorts the
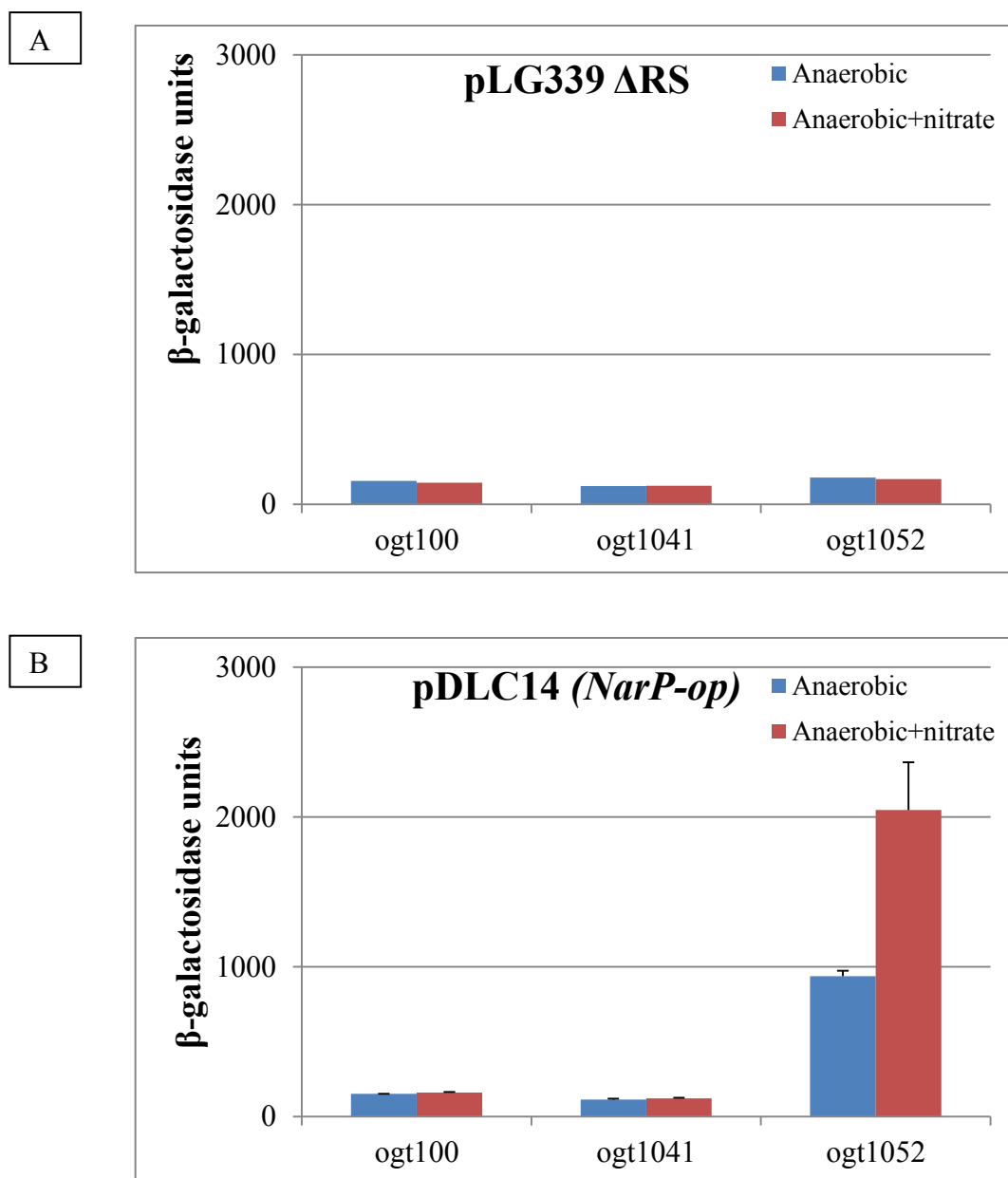
Figure 5.15 Expression of promoter::*lacZ* fusions in cells containing a plasmid carrying the *narP* operon

Figure shows expression of promoter::*lacZ* fusions, carried by pRW50, measured in JCB3884 cells containing control plasmid pLG339 (panel A) or pDLC14 (panel B), which carries the NarP-OP fragment. Cells were grown anaerobically or anaerobically with nitrate in minimal salt. Data shown are averages from at least 3 biological repeats, and error bars show one standard deviation from the mean.

local helical structure (Maris *et al.*, 2002). However, at some promoters, NarL binds to DNA sites in positions where it would be expected to be able to contact the αCTD of RNAP directly. Furthermore, other members of the protein family to which NarL belongs are able to interact with the αCTD, which would suggest that NarL is able to as well.

The recently characterised *ogt* promoter presents the perfect tool for such a study. The wild-type promoter is activated by a class III mechanism and requires NarL bound at 2 separate DNA targets upstream of the transcript start. Derivatives of the *ogt* promoter have been constructed that are activated either by class I or class II mechanisms. Furthermore, NarL has been shown to activate the *ogt* promoter independently of FNR. FNR is known to interact directly with the αCTD of RNAP and so the presence of FNR at a promoter would complicate such a study.

The results show that there are residues in the αCTD that are important for NarL-dependent activation at the *ogt* promoter and at the class I and class II derivatives. The positions identified in the αCTD are surface exposed, suggesting that they may be involved in interaction with another protein. The identified positions fall in a band around the outside of the αCTD and many of the positions were also identified by Savery *et al.* (1998, 2002) as important for promoter activation by CRP.

Lin and Stewart (2010) recently investigated the NarL CTD by mutating surface exposed residues in order to identify residues that might interact with polymerase. They identified several residues that, when mutated to alanine, gave rise to a NarL protein that was able to bind its target DNA site, but was not able to activate transcription. The positions identified by Lin and Stewart (Lin and Stewart, 2010) are shown in Figure 5.16 and can be seen to group on one face of NarL in a possible 'Activation Region' (Rhodius *et al.*, 1997). It should be noted however that their study was not exhaustive, and instead focused on residues implicated in an alignment with TraJ. The possibility also remains that an activation region may also be
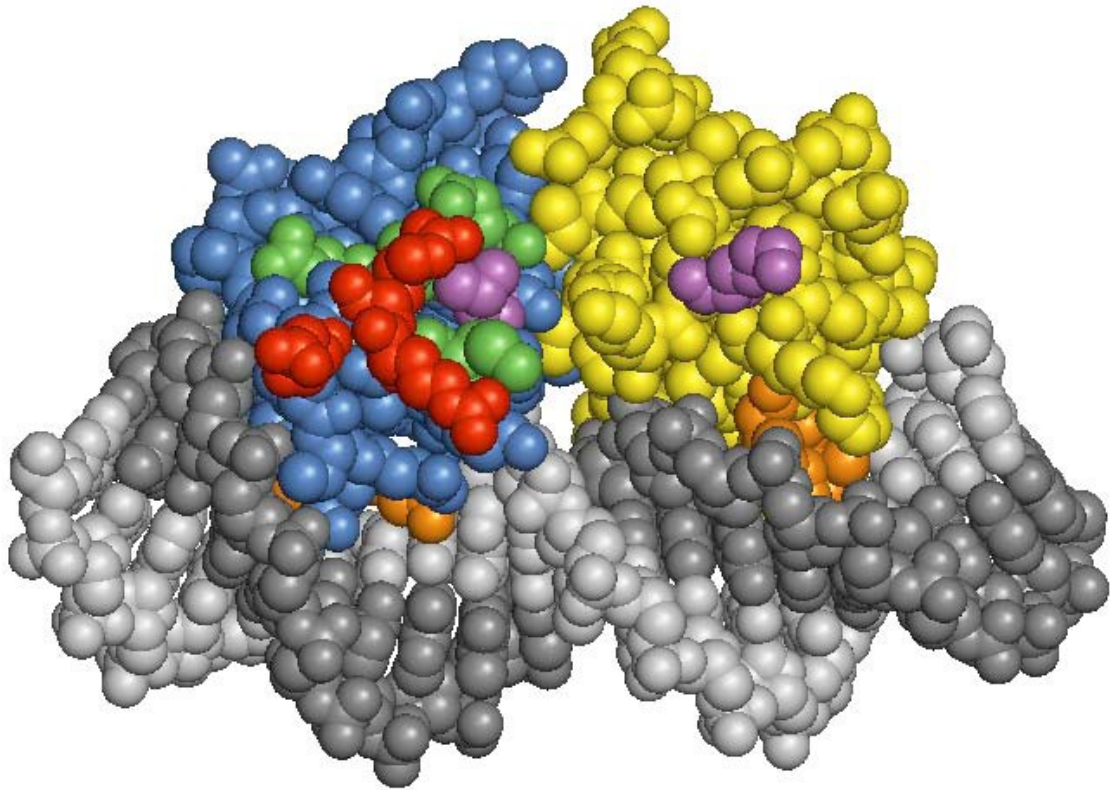
Figure 5.16 Residues of NarL CTD implicated in transcription activation

From Lin and Stewart (2010) X-ray model of a NarL-CTD dimer in complex with its DNA-binding site from the *nirB* operon control region. Structure from Maris *et al.* (2005). The two protomers are colored blue and gold, and the two DNA strands are shaded light and dark gray. Residues are highlighted according to their phenotype: PC, red (Arg-178, Arg-179, Asp-181); functional, green (Asp-162, Leu-166, Gln-169, Met-175); deficient, magenta (Leu-171, Lys-199); DNA-binding, orange (Lys-188, Val-189, Lys-192).

found in the NTD.

5.8.2. The NTD of NarL plays a role in transcription activation

This work and the work of Lin and Stewart (2010) has shown that the full length NarL protein is able to activate by class I, II and III mechanisms, whilst the CTD is only able to activate by a class II mechanism. This suggests that the NTD of NarL contains determinants of activation by class I and III mechanisms, and possibly contains an activation region that interacts with RNAP.

A comparison can be made with CRP-dependent activation by class I and class II mechanisms. CRP requires determinants in its NTD for class II activation that are not required for activation by a class I mechanism because it contacts the NTD of α during class II activation (see Figure 1.7) (Niu *et al.*, 1996). Conversely, NarL does not require its NTD for class II activation but it does require it for class I activation. This suggests that, unlike CRP, NarL makes no contact with the αNTD.

To elucidate residues of the NarL NTD that are involved in transcription activation by class I and III mechanisms, an alanine scan of the NarL NTD could be done. Alternatively, random mutagenesis by error prone PCR could be used to identify key residues.

5.8.3. The CTD of NarL may contact σ when activating by a class II mechanism

An interesting observation from the results of the alanine scan is that, when the expression of a class II activated promoter::*lacZ* fusion is measured (See Figure 5.5), there is less deviation from the wild-type level overall than there is when expression of class I or class III activated promoter::*lacZ* fusions are measured (Figure 5.3). This could simply be due to the relative placement of the RNA polymerase α subunit and NarL, or potentially indicates that NarL makes less contact with the αCTD at class II promoters than it does at class I promoters. CRP has been found to interact with the σ subunit of RNAP and it is possible that NarL does as well (Hu and Gross, 1985).

5.8.4. *NarP can activate by fewer mechanisms than can NarL*

This work and the work of Lin and Stewart (2010) indicate that NarP can activate by a class II mechanism only, whilst NarL can activate by all 3 classes. The results of both studies also indicate that the CTD of NarL and that of NarP are able to activate by a class II mechanism. This suggests that distinctions between the NTDs of NarL and NarP give rise to the different abilities of NarL and NarP.

NarP is activated by NarQ in response to nitrite and NarL is activated by NarX in response to nitrate. However, recent studies have shown that there is some crosstalk between partners and that NarL can be activated by both NarQ and NarX whilst NarP is activated only by NarQ (Noriega *et al.*, 2009). The NTD of both NarL and NarP is known to be the site of phosphorylation by the sensor regulator partner proteins, which supports the idea that the NTDs of NarL and NarP differ functionally. Because of this, it might be expected that the CTDs of NarL and NarP would be more similar than are the NTDs, but alignments show similar levels of resemblance (data not shown). A potentially interesting experiment would be to fuse the NTD of NarL to the CTD of NarP and vice versa. This could produce a non-functional protein (see below) or could further demonstrate that class I and class III mechanisms of activation are dictated by the NTD.

One theory as to the evolution of NarL / NarP is that NarP is the ancestral regulator because it appears in more species than does NarL and NarL has not been reported in a species that does not also encode NarP (Stewart, 2003). The originator of this theory does however point out a flaw in it: that it is difficult to distinguish by sequence alone whether an ORF codes for a NarL or a NarP, although a key indicator might be its proximity to *narG*. This work supports, but by no means proves, the theory that NarL evolved from NarP by gaining functions and becoming differentially phosphorylated.

5.8.5. *Cloning difficulties encountered and significance of problems*

Various cloning difficulties were encountered in the construction of plasmid-encoded NarL and NarP. First, the NarL-Full fragment was designed to produce a full-length NarL protein, but the second codon was mutated to aid cloning, which appears to produce a non-functioning protein. This protein was not able to activate by class I or III mechanisms but was able to activate by a class II mechanism. This pattern of activation is identical to that of the NarL CTD. NarL contains both a linker helix and an unordered flexible linker between the domains and so it is possible that this linker region acts as a buffer, possible because the CTD folds independently of the NTD. The mutated second codon therefore potentially causes a malformed NTD whilst not affecting the CTD.

A similar problem was encountered with cloning the NarP CTD. Two constructs, the NarP-CTD146 fragment and the NarP-CTD148 fragment (see Figure 5.13), failed to produce a functional NarP CTD. Interestingly, the NarP-CTD148 fragment starts from the same position as the relevant fragment used by Lin and Stewart (2010), which was shown to be functional. However, constructs used by Lin and Stewart (2010) contained a 20 amino acid leader, derived from LacZ, fused upstream. This leader may aid correct folding and benefits of such a leader have been observed with GFP (personal communication from Jack Bryant, University of Birmingham, UK).

Finally, it was found to be impossible to obtain transformants with a high copy number plasmid carrying the *narXL* operon under the control of its own promoter. This is likely to be because such a construct is toxic to the cell. If so, it would indicate that NarL is an important protein and that overproduction causes undesirable regulatory effects that the cell cannot tolerate.

5.8.6. Summary

This work has produced evidence that NarL interacts with RNAP directly, via the α subunit. This work also demonstrates that the NarL CTD is able to activate class II promoters,

whilst the full-length protein is required for class I and II mechanisms. The results of this work agree with those of Lin and Stewart (2010) but have the benefit that only promoters where NarL activates independently of other factors were used in the current study. In contrast, Lin and Stewart (2010) study used promoters where FNR also activates, which may have interfered with results.

Chapter 6: Closing comments

Escherichia coli is able to survive in anaerobic conditions by utilising alternative electron acceptors such as nitrate and nitrite. This ability may aid the survival of *E. coli* in the more oxygen-poor regions of the digestive system and / or outside the body (in soils for example). However, with the benefits of utilising nitrate / nitrite come the threats posed by the toxic by-products of nitrate / nitrite reduction, reactive nitrogen species. RNS can cause bacteriostasis or even cell death by damaging the iron-sulphur clusters of certain proteins or by damaging DNA. Therefore, it is expected that *E. coli* cells that synthesize proteins to utilise nitrate / nitrite would also synthesize proteins to protect cells from damage by RNS, using common regulation mechanisms.

The Hybrid Cluster Protein (HCP) has been previously implicated in the cellular response to reactive nitrogen species and work described in chapter 3 examined the regulation of *hcp* expression. My work confirms that the *hcp* expression is activated by FNR in response to anaerobiosis and is repressed by NsrR in the absence of RNS. Contrary to previous reports, my work shows that NarL / NarP do not have a direct effect on *hcp* expression, despite binding to the promoter at a location from where they would be expected to activate. Random mutagenesis of the *hcp* promoter failed to reveal novel activator binding sites. This work also confirms that the *hcp* promoter is highly active when fully activated, and indicates that activity may be driven by core promoter elements and FNR activation. This work has contributed to the understanding of *hcp* expression, although the function of HCP remains enigmatic.

Regulation of the *hcp* promoter is complex and, even with the contributions of this work, is not fully understood. Reports of regulation of *hcp* expression by OxyR need to be investigated in the context of regulation by FNR / NsrR. The exact mechanism(s) producing the high level of promoter activity also needs to be better understood. This work shows that

NarL has an indirect effect on *hcp* expression that is relieved if NsrR is unable to bind to the *hcp* promoter. This strongly suggests that NarL plays a crucial role in the generation of reactive nitrogen species from nitrate / nitrate. The *hcp* promoter can be used as a reporter of RNS *in vivo* and, as is being used to investigate endogenous sources of RNS in *E. coli* (Claire Vine, University of Birmingham, unpublished results).

The Ogt protein is a DNA repair enzyme whose expression is activated by NarL in response to nitrate. NarL typically co-activates with FNR although is the sole (discovered) activator of Ogt expression. The work described in chapter 4 examines the positions and sequences of DNA sites from which NarL can activate *ogt* expression. The work shows that NarL can activate by class I, II and III mechanisms, although activation by class I or II requires DNA sites that closely match the consensus binding sequence, and that NarP is able to activate by a class II mechanism only, although this activation requires a DNA site that closely resembles the consensus binding sequence.

The work in chapter 5 further examined activation by the NarL and NarP proteins. Previous studies of positive regulation by NarL at the *nirB* and *nrfA* promoters has shown that NarL activates by displacing a repressor bound to the promoter (Browning *et al.*, 2004,2006). This work presents the first evidence that NarL is able to activate transcription by directly contacting RNA polymerase. Plasmids were made that express NarL, NarL CTD, NarP or the NarP CTD and transcription activation of the expressed proteins was studied. The NarL CTD, full-length NarP and NarP CTD are able to activate transcription by class II mechanisms whilst the full-length NarL can activate by class I, II and III mechanisms. The CTD of NarL is also able to activate the *nirB* promoter, which indicates that it can displace bound IHF (data not shown and Lin *et al.*, 2010)). The plasmids constructed in this study will be useful for future studies of activation by NarL. Future studies should focus on mechanisms of activation and explore the regions of NarL and NarP that contact RNA polymerase although some work

towards this goal has been reported in Lin *et al.* (2010). Because the NarL CTD can displace bound repressors and activate by a class II mechanism, it might be useful to do a transcriptomics study of cells containing either NarL or the NarL CTD. This might identify promoters that are directly activated by class I or III mechanisms. If the PC mutations identified in Lin *et al.* (2010) are included, it might be possible to identify promoters that are activated by a class II mechanism as well. Such techniques may rapidly increase our understanding of the important NarL regulon.

Rather than considering the expression of HCP and Ogt as regulated in response to nitrate, it may be useful to regard their expression as regulated in response to RNS. In this model, expression of HCP is regulated in response to NO directly, while expression of Ogt is regulated in response to a precursor of RNS, nitrate. NsrR therefore acts as an immediate regulator of HCP expression in response to RNS and NarL acts as a prophylactic regulator of Ogt expression in response to RNS. This difference in regulation is reflected in promoter activity. When maximally induced, the *hcp* promoter is highly active, possibly because it is up-regulated in response to an already present threat. The *ogt* promoter is not highly active, even when maximally induced. Because expression of Ogt is up-regulated in response to a precursor, the promoter does not need to be highly active as there would be time for Ogt to be synthesized before RNS are produced in the cell.

HCP has been implicated in survival during infection of a human host, which may be because HCP protects the cell from RNS (or ROS) produced by the host immune system to kill invading cells. This study suggests that the *hcp* promoter would be activated by exogenous RNS, although for optimal promoter activity the cell would have to be in an anaerobic environment. However, Ogt is unlikely to be involved in protection from the host immune system. By responding to a metabolic precursor of NO rather than NO directly, expression of Ogt would not be expected to be up-regulated by exogenous RNS such as that

produced by macrophages. DNA damage resulting from exogenous RNS is likely to be repaired by the adaptive response protein Ada.

Whilst expression of Ogt is induced by nitrate, it is curious that it is not induced by nitrite. Reduction of nitrite is also believed to give rise to NO and so Ogt would be expected to be produced in response to the presence of nitrite. A possible explanation for this is that expression of Ogt might be optimally regulated by NarL and Fis and a promoter organisation that NarP could bind to may not be fully Fis regulated. That expression of Ogt is not regulated by nitrite may therefore be a trade-off. As with DNA damage caused by exogenous RNS, DNA damage from by-products of nitrite reduction is likely dealt with by Ada.

One topic this work addresses is mechanisms of regulating the expression of a gene by multiple factors. Assuming a simple promoter that is regulated by 2 factors, there are 3 possibilities for how it may be regulated: by two repressors; by an activator and a repressor or by two activators. With regards to regulating, a repressor is less efficient for a cell than is an activator, because the cell must synthesise the repressor in the absence of a signal. An activator can be synthesised and up-regulate expression of itself and other genes when a particular signal is present. This prevents some wasted energy on protein production, but conversely may mean the cellular response takes longer to occur as the activator needs to be synthesised. The *hcp* promoter is an example of a promoter that is regulated by an activator and a repressor, FNR and NsrR respectively. Synthesising NsrR in the absence of RNS may appear to be an unnecessary burden on the cell but is necessary if HCP is involved in protection from RNS. This is because when RNS are present (having been produced either endogenously or exogenously), the cell needs to rapidly commence synthesis of protective proteins to offset the damage that will be caused. If NsrR were an activator, precious minutes may be added to the cells response time. As expression of *ogt* is regulated by 2 copies of an

activator (that bind to poor sites), it is likely that *ogt* does not need to be synthesised in an especially time-sensitive fashion.

In summary, this work has contributed to the understanding of transcription regulation by NarL and NarP and provided tools and initial data to aid further characterisation of these important transcription factors. My work also made significant progress in the understanding of regulation of the *hcp* promoter, that (combined with future functional studies) may help to elucidate the cellular role of HCP.

Chapter 7: Bibliography

Abbondanzieri, E. A., W. J. Greenleaf, J. W. Shaevitz, R. Landick and S. M. Block, (2005) Direct observation of base-pair stepping by RNA polymerase. *Nature* **438**: 460-465.

Allison, L. A., M. Moyle, M. Shales and C. J. Ingles, (1985) Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell* **42**: 599-610.

Artsimovitch, I. and R. Landick, (2000) Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc Natl Acad Sci U S A* **97**: 7090-7095.

Baikalov, I., I. Schroder, M. Kaczor-Grzeskowiak, K. Grzeskowiak, R. P. Gunsalus and R. E. Dickerson, (1996) Structure of the *Escherichia coli* response regulator NarL. *Biochemistry* **35**: 11053-11061.

Baranova, N. N., A. Danchin and A. A. Neyfakh, (1999) Mta, a global MerR-type regulator of the Bacillus subtilis multidrug-efflux transporters. *Mol Microbiol* **31**: 1549-1559.

Bell, A. I., J. A. Cole and S. J. Busby, (1990) Molecular genetic analysis of an FNR-dependent anaerobically inducible *Escherichia coli* promoter. *Mol Microbiol* **4**: 1753-1763.

Belyaeva, T. A., J. T. Wade, C. L. Webster, V. J. Howard, M. S. Thomas, E. I. Hyde and S. J. Busby, (2000) Transcription activation at the *Escherichia coli melAB* promoter: the role of MelR and the cyclic AMP receptor protein. *Mol Microbiol* **36**: 211-222.

Bijlsma, J. J. and E. A. Groisman, (2003) Making informed decisions: regulatory interactions between two-component systems. *Trends Microbiol* **11**: 359-366.

Blake, T., A. Barnard, S. J. Busby and J. Green, (2002) Transcription activation by FNR: evidence for a functional activating region 2. *J Bacteriol* **184**: 5855-5861.

Blatter, E. E., W. Ross, H. Tang, R. L. Gourse and R. H. Ebright, (1994) Domain organization of RNA polymerase alpha subunit: C-terminal 85 amino acids constitute a domain capable of dimerization and DNA binding. *Cell* **78**: 889-896.

Blattner, F. R., G. Plunkett, 3rd, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau and Y. Shao, (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* **277**: 1453-1462.

Bodenmiller, D. M. and S. Spiro, (2006) The *yjeB* (*nsrR*) gene of *Escherichia coli* encodes a nitric oxide-sensitive transcriptional regulator. *J Bacteriol* **188**: 874-881.

Bonnefoy, V. and J. A. Demoss, (1994) Nitrate reductases in *Escherichia coli*. *Antonie Leeuwenhoek* **66**: 47-56.

Borukhov, S. and K. Severinov, (2002) Role of the RNA polymerase sigma subunit in transcription initiation. *Res Microbiol* **153**: 557-562.

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

Browning, D. F., J. A. Cole and S. J. Busby, (2000) Suppression of FNR-dependent transcription activation at the *Escherichia coli nir* promoter by Fis, IHF and H-NS: modulation of transcription initiation by a complex nucleo-protein assembly. *Mol Microbiol* **37**: 1258-1269.

Browning, D. F., J. A. Cole and S. J. Busby, (2004) Transcription activation by remodelling of a nucleoprotein assembly: the role of NarL at the FNR-dependent *Escherichia coli nir* promoter. *Mol Microbiol* **53**: 203-215.

Browning, D. F., D. J. Lee, S. Spiro and S. J. Busby, (2010) Down-regulation of the *Escherichia coli* K-12 *nrf* promoter by binding of the NsrR nitric oxide-sensing transcription repressor to an upstream site. *J Bacteriol* **192**: 3824-3828.

Browning, D. F., D. J. Lee, A. J. Wolfe, J. A. Cole and S. J. Busby, (2006) The *Escherichia coli* K-12 NarL and NarP proteins insulate the *nrf* promoter from the effects of integration host factor. *J Bacteriol* **188**: 7449-7456.

Burgess, R. R., (1969) Separation and characterization of the subunits of ribonucleic acid polymerase. *J Biol Chem* **244**: 6168-6176.

Burgess, R. R. and L. Anthony, (2001) How sigma docks to RNA polymerase and what sigma does. *Curr Opin Microbiol* **4**: 126-131.

Burgess, R. R., A. A. Travers, J. J. Dunn and E. K. Bautz, (1969) Factor stimulating transcription by RNA polymerase. *Nature* **221**: 43-46.

Burmann, B. M., K. Schweimer, X. Luo, M. C. Wahl, B. L. Stitt, M. E. Gottesman and P. Rosch, (2010) A NusE:NusG complex links transcription and translation. *Science* **328**: 501-504.

Butala, M., S. J. Busby and D. J. Lee, (2009) DNA sampling: a method for probing protein binding at specific loci on bacterial chromosomes. *Nucleic Acids Res* **37**: e37.

Campbell, E. A., O. Pavlova, N. Zenkin, F. Leon, H. Irschik, R. Jansen, K. Severinov and S. A. Darst, (2005) Structural, functional, and genetic analysis of sorangicin inhibition of bacterial RNA polymerase. *EMBO J* **24**: 674-682.

Campbell, G. R., J. Prosser, A. Glover and K. Killham, (2001) Detection of *Escherichia coli* O157:H7 in soil and water using multiplex PCR. *J Appl Microbiol* **91**: 1004-1010.

Cases, I., V. de Lorenzo and C. A. Ouzounis, (2003) Transcription regulation and environmental adaptation in bacteria. *Trends Microbiol* **11**: 248-253.

Chismon, D. L., D. F. Browning, G. K. Farrant and S. J. Busby, (2010) Unusual organization, complexity and redundancy at the *Escherichia coli hcp-hcr* operon promoter. *Biochem J* **430**: 61-68.

Choi, S. H. and E. P. Greenberg, (1991) The C-terminal region of the *Vibrio fischeri* LuxR protein contains an inducer-independent lux gene activating domain. *Proc Natl Acad Sci U S A* **88**: 11115-11119.

Clegg, S. J., W. Jia and J. A. Cole, (2006) Role of the *Escherichia coli* nitrate transport protein, NarU, in survival during severe nutrient starvation and slow growth. *Microbiology* **152**: 2091-2100.

Condon, C., C. Squires and C. L. Squires, (1995) Control of rRNA transcription in *Escherichia coli*. *Microbiol Rev* **59**: 623-645.

Constantinidou, C., J. L. Hobman, L. Griffiths, M. D. Patel, C. W. Penn, J. A. Cole and T. W. Overton, (2006) A reassessment of the FNR regulon and transcriptomic analysis of the effects of nitrate, nitrite, NarXL, and NarQP as *Escherichia coli* K12 adapts from aerobic to anaerobic growth. *J Biol Chem* **281**: 4802-4815.

Cooper, S. J., C. D. Garner, W. R. Hagen, P. F. Lindley and S. Bailey, (2000) Hybrid-cluster protein (HCP) from *Desulfovibrio vulgaris* (Hildenborough) at 1.6 A resolution. *Biochemistry* **39**: 15044-15054.

Corker, H. and R. K. Poole, (2003) Nitric oxide formation by *Escherichia coli*. Dependence on nitrite reductase, the NO-sensing regulator Fnr, and flavohemoglobin Hmp. *J Biol Chem* **278**: 31584-31592.

Crack, J. C., J. Green, M. R. Cheesman, N. E. Le Brun and A. J. Thomson, (2007) Superoxide-mediated amplification of the oxygen-induced switch from [4Fe-4S] to [2Fe-2S] clusters in the transcriptional regulator FNR. *Proc Natl Acad Sci U S A* **104**: 2092-2097.

Cronan, J. E., (2001) *Escherichia coli* as an Experimental Organism. In: Encyclopedia of Life Sciences. John Wiley & Sons, Ltd, pp.

Cruz-Ramos, H., J. Crack, G. Wu, M. N. Hughes, C. Scott, A. J. Thomson, J. Green and R. K. Poole, (2002) NO sensing by FNR: regulation of the Escherichia coli NO-detoxifying flavohaemoglobin, Hmp. *Embo J* **21**: 3235-3244.

D'Autreaux, B., N. P. Tucker, R. Dixon and S. Spiro, (2005) A non-haem iron centre in the transcription factor NorR senses nitric oxide. *Nature* **437**: 769-772.

Darst, S. A., (2001) Bacterial RNA polymerase. *Curr Opin Struct Biol* **11**: 155-162.

Darwin, A. J., J. Li and V. Stewart, (1996) Analysis of nitrate regulatory protein NarL-binding sites in the fdnG and narG operon control regions of *Escherichia coli* K-12. *Mol Microbiol* **20**: 621-632.

Darwin, A. J., K. L. Tyson, S. J. Busby and V. Stewart, (1997) Differential regulation by the homologous response regulators NarL and NarP of *Escherichia coli* K-12 depends on DNA binding site arrangement. *Mol Microbiol* **25**: 583-595.

Demple, B., (1996) Redox signaling and gene control in the *Escherichia coli* soxRS oxidative stress regulon--a review. *Gene* **179**: 53-57.

Dibden, D. P. and J. Green, (2005) In vivo cycling of the *Escherichia coli* transcription factor FNR between active and inactive states. *Microbiology* **151**: 4063-4070.

Ebright, R. H. and S. Busby, (1995) The *Escherichia coli* RNA polymerase alpha subunit: structure and function. *Curr Opin Genet Dev* **5**: 197-203.

Ellinger, T., D. Behnke, H. Bujard and J. D. Gralla, (1994) Stalling of *Escherichia coli* RNA polymerase in the +6 to +12 region in vivo is associated with tight binding to consensus promoter elements. *J Mol Biol* **239**: 455-465.

Estrem, S. T., W. Ross, T. Gaal, Z. W. Chen, W. Niu, R. H. Ebright and R. L. Gourse, (1999) Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase alpha subunit. *Genes Dev* **13**: 2134-2147.

Fang, F. C., (2004) Antimicrobial reactive oxygen and nitrogen species: concepts and controversies. *Nat Rev Microbiol* **2**: 820-832.

Filenko, N., S. Spiro, D. F. Browning, D. Squire, T. W. Overton, J. Cole and C. Constantinidou, (2007) The NsrR regulon of *Escherichia coli* K-12 includes genes encoding the hybrid cluster protein and the periplasmic, respiratory nitrite reductase. *J Bacteriol* **189**: 4410-4417.

Filenko, N. A., D. F. Browning and J. A. Cole, (2005) Transcriptional regulation of a hybrid cluster (prismane) protein. *Biochem Soc Trans* **33**: 195-197.

Flynn, J. M., S. B. Neher, Y. I. Kim, R. T. Sauer and T. A. Baker, (2003) Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. *Mol Cell* **11**: 671-683.

Freter, R. and G. D. Abrams, (1972) Function of various intestinal bacteria in converting germfree mice to the normal state. *Infect Immun* **6**: 119-126.

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

Gentry, D., H. Xiao, R. Burgess and M. Cashel, (1991) The omega subunit of *Escherichia coli* K-12 RNA polymerase is not required for stringent RNA control in vivo. *J Bacteriol* **173**: 3901-3903.

Gilberthorpe, N. J., M. E. Lee, T. M. Stevanin, R. C. Read and R. K. Poole, (2007) NsrR: a key regulator circumventing *Salmonella enterica* serovar Typhimurium oxidative and nitrosative stress in vitro and in IFN-{gamma}-stimulated J774.2 macrophages. *Microbiology* **153**: 1756-1771.

Goh, E. B., P. J. Bledsoe, L. L. Chen, P. Gyaneshwar, V. Stewart and M. M. Igo, (2005) Hierarchical control of anaerobic gene expression in *Escherichia coli* K-12: the nitrate-responsive NarX-NarL regulatory system represses synthesis of the fumarate-responsive DcuS-DcuR regulatory system. *J Bacteriol* **187**: 4890-4899.

Gourse, R. L. and W. R. T. Gaal, (2000) UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. *Mol Microbiol* **37**: 687-695.

Green, J. and F. A. Marshall, (1999) Identification of a surface of FNR overlapping activating region 1 that is required for repression of gene expression. *J Biol Chem* **274**: 10244-10248.

Grossman, L. and O. Kovalsky, (2001) Nucleotide Excision Repair in Bacteria. In: Encyclopedia of Life Sciences. John Wiley & Sons, Ltd, pp.

Gusarov, I. and E. Nudler, (1999) The mechanism of intrinsic transcription termination. *Mol Cell* **3**: 495-504.

Hawley, D. K. and W. R. McClure, (1983) Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucleic Acids Res* **11**: 2237-2255.

Hengge, R., (2009) Proteolysis of sigma-S (RpoS) and the general stress response in *Escherichia coli*. *Res Microbiol* **160**: 667-676.

Herbert, K. M., W. J. Greenleaf and S. M. Block, (2008) Single-molecule studies of RNA polymerase: motoring along. *Annu Rev Biochem* **77**: 149-176.

Hidalgo, E. and B. Demple, (1997) Spacing of promoter elements regulates the basal expression of the *soxS* gene and converts SoxR from a transcriptional activator into a repressor. *EMBO J* **16**: 1056-1065.

Hoch, J. A., (2000) Two-component and phosphorelay signal transduction. *Curr Opin Microbiol* **3**: 165-170.

Holt, A. K. and D. F. Senear, (2010) An unusual pattern of CytR and CRP binding energetics at *Escherichia coli cddP* suggests a unique blend of class I and class II mediated activation. *Biochemistry* **49**: 432-442.

Hsu, L. M., (2002) Promoter clearance and escape in prokaryotes. *Biochim Biophys Acta* **1577**: 191-207.

Hu, J. C. and C. A. Gross, (1985) Mutations in the sigma subunit of *E. coli* RNA polymerase which affect positive control of transcription. *Mol Gen Genet* **199**: 7-13.

Igarashi, K., N. Fujita and A. Ishihama, (1991) Identification of a subunit assembly domain in the alpha subunit of *Escherichia coli* RNA polymerase. *J Mol Biol* **218**: 1-6.

Islam, M. S., L. E. H. Bingle, M. J. Pallen and S. J. W. Busby, (2010) Organization of the LEE1 operon regulatory region of enterohaemorrhagic *Escherichia coli* O157:H7 and activation by GrlA. *Mol Microbiol* **79**: 468–483.

Izu, H., S. Ito, M. D. Elias and M. Yamada, (2002) Differential control by IHF and cAMP of two oppositely oriented genes, *hpt* and *gcd*, in *Escherichia coli:* significance of their partially overlapping regulatory elements. *Mol Genet Genomics* **266**: 865-872.

Jayaraman, P. S., J. A. Cole and S. J. Busby, (1989) Mutational analysis of the nucleotide sequence at the FNR-dependent *nirB* promoter in *Escherichia coli*. *Nucleic Acids Res* **17**: 135-145.

Jeon, Y. H., T. Negishi, M. Shirakawa, T. Yamazaki, N. Fujita, A. Ishihama and Y. Kyogoku, (1995) Solution structure of the activator contact domain of the RNA polymerase alpha subunit. *Science* **270**: 1495-1497.

Jin, D. J., R. R. Burgess, J. P. Richardson and C. A. Gross, (1992) Termination efficiency at rho-dependent terminators depends on kinetic coupling between RNA polymerase and rho. *Proc Natl Acad Sci U S A* **89**: 1453-1457.

Jones, S. A., F. Z. Chowdhury, A. J. Fabich, A. Anderson, D. M. Schreiner, A. L. House, S. M. Autieri, M. P. Leatham, J. J. Lins, M. Jorgensen, P. S. Cohen and T. Conway, (2007) Respiration of *Escherichia coli* in the mouse intestine. *Infect Immun* **75**: 4891-4899.

Joung, J. K., D. M. Koepp and A. Hochschild, (1994) Synergistic activation of transcription by bacteriophage lambda cI protein and *E. coli* cAMP receptor protein. *Science* **265**: 1863-1866.

Joung, J. K., L. U. Le and A. Hochschild, (1993) Synergistic activation of transcription by *Escherichia coli* cAMP receptor protein. *Proc Natl Acad Sci U S A* **90**: 3083-3087.

Justino, M. C., C. C. Almeida, M. Teixeira and L. M. Saraiva, (2007) *Escherichia coli* di-iron YtfE protein is necessary for the repair of stress-damaged iron-sulfur clusters. *J Biol Chem* **282**: 10352-10359.

Kahramanoglou, C., C. L. Webster, M. S. El-Robh, T. A. Belyaeva and S. J. Busby, (2006) Mutational analysis of the *Escherichia coli melR* gene suggests a two-state concerted model to explain transcriptional activation and repression in the melibiose operon. *J Bacteriol* **188**: 3199-3207.

Kim, C. C., D. Monack and S. Falkow, (2003) Modulation of virulence by two acidified nitrite-responsive loci of *Salmonella enterica* serovar Typhimurium. *Infect Immun* **71**: 3196-3205.

Kobayashi, M., K. Nagata and A. Ishihama, (1990) Promoter selectivity of *Escherichia coli* RNA polymerase: effect of base substitutions in the promoter -35 region on promoter strength. *Nucleic Acids Res* **18**: 7367-7372.

Kolb, A., D. Kotlarz, S. Kusano and A. Ishihama, (1995) Selectivity of the *Escherichia coli* RNA polymerase E sigma 38 for overlapping promoters and ability to support CRP activation. *Nucleic Acids Res* **23**: 819-826.

Laver, J. R., T. M. Stevanin, S. L. Messenger, A. D. Lunn, M. E. Lee, J. W. Moir, R. K. Poole and R. C. Read, (2010) Bacterial nitric oxide detoxification prevents host cell S-nitrosothiol formation: a novel mechanism of bacterial pathogenesis. *FASEB J* **24**: 286-295.

Lawley, P. D. and D. H. Phillips, (1996) DNA adducts from chemotherapeutic agents. *Mutat Res* **355**: 13-40.

Lesnik, E. A., R. Sampath, H. B. Levene, T. J. Henderson, J. A. McNeil and D. J. Ecker, (2001) Prediction of rho-independent transcriptional terminators in *Escherichia coli*. *Nucleic Acids Res* **29**: 3583-3594.

Lewis, M., (2005) The lac repressor. *C R Biol* **328**: 521-548.

Li, Z. and B. Demple, (1994) SoxS, an activator of superoxide stress genes in *Escherichia coli*. Purification and interaction with DNA. *J Biol Chem* **269**: 18371-18377.

Lin, A. V. and V. Stewart, (2010) Functional roles for the GerE-family carboxyl-terminal domains of nitrate response regulators NarL and NarP of *Escherichia coli* K-12. *Microbiology* **156**: 2933-2943.

Lin, H. Y., P. J. Bledsoe and V. Stewart, (2007) Activation of *yeaR-yoaG* operon transcription by the nitrate-responsive regulator NarL is independent of oxygen- responsive regulator Fnr in *Escherichia coli* K-12. *J Bacteriol* **189**: 7539-7548.

Lodge, J., J. Fear, S. Busby, P. Gunasekaran and N. R. Kamini, (1992) Broad host range plasmids carrying the *Escherichia coli* lactose and galactose operons. *FEMS Microbiol Lett* **74**: 271-276.

Maniatis, T., E. F. Fritsch and J. Sambrook, (1982) *Molecular cloning: A laboratory manual*. CSHL, New York.

Maris, A. E., M. R. Sawaya, M. Kaczor-Grzeskowiak, M. R. Jarvis, S. M. Bearson, M. L. Kopka, I. Schroder, R. P. Gunsalus and R. E. Dickerson, (2002) Dimerization allows DNA target site recognition by the NarL response regulator. *Nat Struct Biol* **9**: 771-778.

Marshall, F. A., S. L. Messenger, N. R. Wyborn, J. R. Guest, H. Wing, S. J. Busby and J. Green, (2001) A novel promoter architecture for microaerobic activation by the anaerobic transcription factor FNR. *Mol Microbiol* **39**: 747-753.

Martinez-Antonio, A. and J. Collado-Vides, (2003) Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr Opin Microbiol* **6**: 482-489.

Maxam, A. M. and W. Gilbert, (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol* **65**: 499-560.

Memisoglu, A. and L. D. Samson, (2001) DNA Repair by Reversal of Damage. In: Encyclopedia of Life Sciences. John Wiley & Sons, Ltd, pp.

Metheringham, R. and J. A. Cole, (1997) A reassessment of the genetic determinants, the effect of growth conditions and the availability of an electron donor on the nitrosating activity of *Escherichia coli* K-12. *Microbiology* **143**: 2647-2656.

Miller, G. and S. Hahn, (2006) A DNA-tethered cleavage probe reveals the path for promoter DNA in the yeast preinitiation complex. *Nat Struct Mol Biol* **13**: 603-610.

Minakhin, L., S. Bhagat, A. Brunning, E. A. Campbell, S. A. Darst, R. H. Ebright and K. Severinov, (2001) Bacterial RNA polymerase subunit omega and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. *Proc Natl Acad Sci U S A* **98**: 892-897.

Mitchell, J. E., D. Zheng, S. J. Busby and S. D. Minchin, (2003) Identification and analysis of 'extended -10' promoters in *Escherichia coli*. *Nucleic Acids Res* **31**: 4689-4695.

Montet, Y., P. Amara, A. Volbeda, X. Vernede, E. C. Hatchikian, M. J. Field, M. Frey and J. C. Fontecilla-Camps, (1997) Gas access to the active site of Ni-Fe hydrogenases probed by X-ray crystallography and molecular dynamics. *Nat Struct Biol* **4**: 523-526.

Mooney, R. A., S. A. Darst and R. Landick, (2005) Sigma and RNA polymerase: an on-again, off-again relationship? *Mol Cell* **20**: 335-345.

Murakami, K. S. and S. A. Darst, (2003) Bacterial RNA polymerases: the wholo story. *Curr Opin Struct Biol* **13**: 31-39.

Murakami, K. S., S. Masuda, E. A. Campbell, O. Muzzin and S. A. Darst, (2002) Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* **296**: 1285-1290.

Niu, W., Y. Kim, G. Tau, T. Heyduk and R. H. Ebright, (1996) Transcription activation at class II CAP-dependent promoters: two interactions between CAP and RNA polymerase. *Cell* **87**: 1123-1134.

Nohno, T., S. Noji, S. Taniguchi and T. Saito, (1989) The *narX* and *narL* genes encoding the nitrate-sensing regulators of *Escherichia coli* are homologous to a family of prokaryotic two-component regulatory genes. *Nucleic Acids Res* **17**: 2947-2957.

Noll, D. M., T. M. Mason and P. S. Miller, (2006) Formation and repair of interstrand cross-links in DNA. *Chem Rev* **106**: 277-301.

Nonaka, G., M. Blankschien, C. Herman, C. A. Gross and V. A. Rhodius, (2006) Regulon and promoter analysis of the *E. coli* heat-shock factor, sigma-32, reveals a multifaceted cellular response to heat stress. *Genes Dev* **20**: 1776-1789.

Noriega, C. E., H. Y. Lin, L. L. Chen, S. B. Williams and V. Stewart, (2009) Asymmetric cross regulation between the nitrate-responsive NarX-NarL and NarQ-NarP two-component regulatory systems from *Escherichia coli* K-12. *Mol Microbiol*.

Nudler, E. and M. E. Gottesman, (2002) Transcription termination and anti-termination in *E. coli*. *Genes Cells* **7**: 755-768.

Oshima, T., H. Aiba, Y. Masuda, S. Kanaya, M. Sugiura, B. L. Wanner, H. Mori and T. Mizuno, (2002) Transcriptome analysis of all two-component regulatory system mutants of *Escherichia coli* K-12. *Mol Microbiol* **46**: 281-291.

Overton, T. W., M. C. Justino, Y. Li, J. M. Baptista, A. M. Melo, J. A. Cole and L. M. Saraiva, (2008) Widespread distribution in pathogenic bacteria of di-iron proteins that repair oxidative and nitrosative damage to iron-sulfur centers. *J Bacteriol* **190**: 2004-2013.

Page, L., L. Griffiths and J. A. Cole, (1990) Different physiological roles of two independent pathways for nitrite reduction to ammonia by enteric bacteria. *Arch Microbiol* **154**: 349-354.

Partridge, J. D., D. M. Bodenmiller, M. S. Humphrys and S. Spiro, (2009) NsrR targets in the *Escherichia coli* genome: new insights into DNA sequence requirements for binding and a role for NsrR in the regulation of motility. *Mol Microbiol* **73**: 680-694.

Perrin, S. and G. Gilliland, (1990) Site-specific mutagenesis using asymmetric polymerase chain reaction and a single mutant primer. *Nucleic Acids Res* **18**: 7433-7438.

Pierik, A. J., R. B. Wolbert, P. H. Mutsaers, W. R. Hagen and C. Veeger, (1992) Purification and biochemical characterization of a putative [6Fe-6S] prismane-cluster-containing protein from *Desulfovibrio vulgaris* (Hildenborough). *Eur J Biochem* **206**: 697-704.

Plumbridge, J., (2002) Regulation of gene expression in the PTS in *Escherichia coli:* the role and interactions of Mlc. *Curr Opin Microbiol* **5**: 187-193.

Poock, S. R., E. R. Leach, J. W. Moir, J. A. Cole and D. J. Richardson, (2002) Respiratory detoxification of nitric oxide by the cytochrome c nitrite reductase of *Escherichia coli*. *J Biol Chem* **277**: 23664-23669.

Potter, P. M., M. C. Wilkinson, J. Fitton, F. J. Carr, J. Brennand, D. P. Cooper and G. P. Margison, (1987) Characterisation and nucleotide sequence of ogt, the O6-alkylguanine-DNA-alkyltransferase gene of *E. coli*. *Nucleic Acids Res* **15**: 9177-9193.

Pribnow, D., (1975) Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. *Proc Natl Acad Sci U S A* **72**: 784-788.

Proshkin, S., A. R. Rahmouni, A. Mironov and E. Nudler, (2010) Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* **328**: 504-508.

Pullan, S. T., M. D. Gidley, R. A. Jones, J. Barrett, T. M. Stevanin, R. C. Read, J. Green and R. K. Poole, (2007) Nitric oxide in chemostat-cultured *Escherichia coli* is sensed by Fnr and other global regulators: unaltered methionine biosynthesis indicates lack of S nitrosation. *J Bacteriol* **189**: 1845-1855.

Rappas, M., D. Bose and X. Zhang, (2007) Bacterial enhancer-binding proteins: unlocking sigma-54-dependent gene transcription. *Curr Opin Struct Biol* **17**: 110-116.

Ren, B., N. Zhang, J. Yang and H. Ding, (2008) Nitric oxide-induced bacteriostasis and modification of iron-sulphur proteins in *Escherichia coli*. *Mol Microbiol* **70**: 953-964.

Rhodius, V. A. and V. K. Mutalik, (2010) Predicting strength and function for promoters of the *Escherichia coli* alternative sigma factor, sigma-E. *Proc Natl Acad Sci U S A* **107**: 2854-2859.

Rhodius, V. A., W. C. Suh, G. Nonaka, J. West and C. A. Gross, (2006) Conserved and variable functions of the sigma-E stress response in related genomes. *PLoS Biol* **4**: e2.

Rhodius, V. A., D. M. West, C. L. Webster, S. J. Busby and N. J. Savery, (1997) Transcription activation at class II CRP-dependent promoters: the role of different activating regions. *Nucleic Acids Res* **25**: 326-332.

Richardson, D. J., (2000) Bacterial respiration: a flexible process for a changing environment. *Microbiology* **146 (Pt 3)**: 551-571.

Rodionov, D. A., I. L. Dubchak, A. P. Arkin, E. J. Alm and M. S. Gelfand, (2005) Dissimilatory metabolism of nitrogen oxides in bacteria: comparative reconstruction of transcriptional networks. *PLoS Comput Biol* **1**: e55.

Rosenberg, M. and D. Court, (1979) Regulatory sequences involved in the promotion and termination of RNA transcription. *Annu Rev Genet* **13**: 319-353.

Russell, D. R. and G. N. Bennett, (1982) Construction and analysis of *in vivo* activity of *E. coli* promoter hybrids and promoter mutants that alter the -35 to -10 spacing. *Gene* **20**: 231-243.

Sanderson, A., J. E. Mitchell, S. D. Minchin and S. J. Busby, (2003) Substitutions in the *Escherichia coli* RNA polymerase sigma-70 factor that affect recognition of extended -10 elements at promoters. *FEBS Lett* **544**: 199-205.

Schwartz, C. J., O. Djaman, J. A. Imlay and P. J. Kiley, (2000) The cysteine desulfurase, IscS, has a major role in in vivo Fe-S cluster formation in *Escherichia coli*. *Proc Natl Acad Sci U S A* **97**: 9009-9014.

Scott, C., J. D. Partridge, J. R. Stephenson and J. Green, (2003) DNA target sequence and FNR-dependent gene expression. *FEBS Lett* **541**: 97-101.

Scott, S., S. Busby and I. Beacham, (1995) Transcriptional co-activation at the *ansB* promoters: involvement of the activating regions of CRP and FNR when bound in tandem. *Mol Microbiol* **18**: 521-531.

Sedgwick, B., (1997) Nitrosated peptides and polyamines as endogenous mutagens in O6-alkylguanine-DNA alkyltransferase deficient cells. *Carcinogenesis* **18**: 1561-1567.

Shaw, D. J., D. W. Rice and J. R. Guest, (1983) Homology between CAP and Fnr, a regulator of anaerobic respiration in *Escherichia coli*. *J Mol Biol* **166**: 241-247.

Shevell, D. E. and G. C. Walker, (1991) A region of the Ada DNA-repair protein required for the activation of *ada* transcription is not necessary for activation of *alkA*. *Proc Natl Acad Sci U S A* **88**: 9001-9005.

Spek, E. J., L. N. Vuong, T. Matsuguchi, M. G. Marinus and B. P. Engelward, (2002) Nitric oxide-induced homologous recombination in *Escherichia coli* is promoted by DNA glycosylases. *J Bacteriol* **184**: 3501-3507.

Squire, D. J., M. Xu, J. A. Cole, S. J. Busby and D. F. Browning, (2009) Competition between NarL-dependent activation and Fis-dependent repression controls expression from the *Escherichia coli yeaR* and *ogt* promoters. *Biochem J* **420**: 249-257.

Stewart, V., (1993) Nitrate regulation of anaerobic respiratory gene expression in *Escherichia coli*. *Mol Microbiol* **9**: 425-434.

Stewart, V., (2003) Biochemical Society Special Lecture. Nitrate- and nitrite-responsive sensors NarX and NarQ of proteobacteria. *Biochem Soc Trans* **31**: 1-10.

Stoebel, D. M., A. M. Dean and D. E. Dykhuizen, (2008) The cost of expression of *Escherichia coli lac* operon proteins is in the process, not in the products. *Genetics* **178**: 1653-1660.

Stoker, N. G., N. F. Fairweather and B. G. Spratt, (1982) Versatile low-copy-number plasmid vectors for cloning in *Escherichia coli*. *Gene* **18**: 335-341.

Sullivan, S. L. and M. E. Gottesman, (1992) Requirement for *E. coli* NusG protein in factor-dependent transcription termination. *Cell* **68**: 989-994.

Summers, A. O., (1992) Untwist and shout: a heavy metal-responsive transcriptional regulator. *J Bacteriol* **174**: 3097-3101.

Taverna, P. and B. Sedgwick, (1996) Generation of an endogenous DNA-methylating agent by nitrosation in *Escherichia coli*. *J Bacteriol* **178**: 5105-5111.

Teo, I., B. Sedgwick, M. W. Kilpatrick, T. V. McCarthy and T. Lindahl, (1986) The intracellular signal for induction of resistance to alkylating agents in *E. coli*. *Cell* **45**: 315-324.

Todorovic, S., M. C. Justino, G. Wellenreuther, P. Hildebrandt, D. H. Murgida, W. Meyer-Klaucke and L. M. Saraiva, (2008) Iron-sulfur repair YtfE protein from Escherichia coli: structural characterization of the di-iron center. *J Biol Inorg Chem* **13**: 765-770.

Toledo, J. C., Jr., C. A. Bosworth, S. W. Hennon, H. A. Mahtani, H. A. Bergonia and J. R. Lancaster, Jr., (2008) Nitric oxide-induced conversion of cellular chelatable iron into macromolecule-bound paramagnetic dinitrosyliron complexes. *J Biol Chem* **283**: 28926-28933.

Tucker, N. P., M. G. Hicks, T. A. Clarke, J. C. Crack, G. Chandra, N. E. Le Brun, R. Dixon and M. I. Hutchings, (2008) The transcriptional repressor protein NsrR senses nitric oxide directly via a [2Fe-2S] cluster. *PLoS One* **3**: e3623.

Tyson, K. L., J. A. Cole and S. J. Busby, (1994) Nitrite and nitrate regulation at the promoters of two *Escherichia coli* operons encoding nitrite reductase: identification of common target heptamers for both NarP- and NarL-dependent regulation. *Mol Microbiol* **13**: 1045-1055.

Ulrich, L. E., E. V. Koonin and I. B. Zhulin, (2005) One-component systems dominate signal transduction in prokaryotes. *Trends Microbiol* **13**: 52-56.

Uptain, S. M., C. M. Kane and M. J. Chamberlin, (1997) Basic mechanisms of transcript elongation and its regulation. *Annu Rev Biochem* **66**: 117-172.

van den Berg, W. A., W. R. Hagen and W. M. van Dongen, (2000) The hybrid-cluster protein ('prismane protein') from *Escherichia coli*. Characterization of the hybrid-cluster protein, redox properties of the [2Fe-2S] and [4Fe-2S-2O] clusters and identification of an associated NADH oxidoreductase containing FAD and [2Fe-2S]. *Eur J Biochem* **267**: 666-676.

Vassylyev, D. G., S. Sekine, O. Laptenko, J. Lee, M. N. Vassylyeva, S. Borukhov and S. Yokoyama, (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. *Nature* **417**: 712-719.

Vine, C. E., M. C. Justino, L. M. Saraiva and J. Cole, (2010) Detection by whole genome microarrays of a spontaneous 126-gene deletion during construction of a *ytfE* mutant: confirmation that a *ytfE* mutation results in loss of repair of iron-sulfur centres in proteins damaged by oxidative or nitrosative stress. *J Microbiol Methods* **81**: 77-79.

Vogel, U. and K. F. Jensen, (1994) The RNA chain elongation rate in *Escherichia coli* depends on the growth rate. *J Bacteriol* **176**: 2807-2813.

von Hippel, P. H. and O. G. Berg, (1989) Facilitated target location in biological systems. *J Biol Chem* **264**: 675-678.

Wade, J. T., D. C. Roa, D. C. Grainger, D. Hurd, S. J. Busby, K. Struhl and E. Nudler, (2006) Extensive functional overlap between sigma factors in *Escherichia coli*. *Nat Struct Mol Biol* **13**: 806-814.

Walker, M. S. and J. A. DeMoss, (1994) NarL-phosphate must bind to multiple upstream sites to activate transcription from the *narG* promoter of *Escherichia coli*. *Mol Microbiol* **14**: 633-641.

Weiss, B., (2006) Evidence for mutagenesis by nitric oxide during nitrate metabolism in *Escherichia coli*. *J Bacteriol* **188**: 829-833.

Welch, R. A., V. Burland, G. Plunkett, 3rd, P. Redford, P. Roesch, D. Rasko, E. L. Buckles, S. R. Liou, A. Boutin, J. Hackett, D. Stroud, G. F. Mayhew, D. J. Rose, S. Zhou, D. C. Schwartz, N. T. Perna, H. L. Mobley, M. S. Donnenberg and F. R. Blattner, (2002) Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci U S A* **99**: 17020-17024.

Wing, H. J., S. M. Williams and S. J. Busby, (1995) Spacing requirements for transcription activation by *Escherichia coli* FNR protein. *J Bacteriol* **177**: 6704-6710.

Wolfe, M. T., J. Heo, J. S. Garavelli and P. W. Ludden, (2002) Hydroxylamine reductase activity of the hybrid cluster protein from *Escherichia coli*. *J Bacteriol* **184**: 5898-5902.

Wosten, M. M., (1998) Eubacterial sigma-factors. *FEMS Microbiol Rev* **22**: 127-150.

Yan, A. and P. J. Kiley, (2009) Techniques to Isolate O_2-Sensitive Proteins: [4Fe-4S]-FNR as an Example. In: Methods in Enzymology. R. B. Richard & P. D. Murray (eds). Academic Press, pp. 787-805.

Yukl, E. T., M. A. Elbaz, M. M. Nakano and P. Moenne-Loccoz, (2008) Transcription Factor NsrR from *Bacillus subtilis* Senses Nitric Oxide with a 4Fe-4S Cluster. *Biochemistry* **47**: 13084-13092.

Zhang, G., E. A. Campbell, L. Minakhin, C. Richter, K. Severinov and S. A. Darst, (1999) Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 A resolution. *Cell* **98**: 811-824.

Zhou, Y., T. J. Merkel and R. H. Ebright, (1994) Characterization of the activating region of *Escherichia col*i catabolite gene activator protein (CAP). II. Role at Class I and class II CAP-dependent promoters. *J Mol Biol* **243**: 603-610.

Zuker, M., (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406-3415.