Regulating Expression of Virulence Determinants in Enterotoxigenic *Escherichia coli* H10407

By

James Richard John Haycocks

A thesis submitted to the University of Birmingham

For the degree of

DOCTOR OF PHILOSOPHY

School of Biosciences

University of Birmingham

November, 2014

UNIVERSITYOF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

Abstract

Enterotoxigenic *Escherichia coli* (ETEC) is a major cause of traveller's and infantile diarrhoea in developing countries, and results in considerable mortality in under 5 year olds. Disease is mediated through adhesion of ETEC cells to the intestinal brush border and the secretion of the heat-stable and/or heat-labile toxin. Whilst the toxins have been well studied it is still not clear how their expression is regulated. Existing reports suggest a role for the cyclic AMP receptor protein (CRP) in this regulation.

This work has defined binding of CRP, H-NS and σ^{70} across the genome of the prototypical ETEC strain H10407. We demonstrate a central role for all three factors in regulating pathogenicity in ETEC H10407. Specifically, we show that although CRP binding across the chromosome of H10407 does not appear to differ significantly from that of non-pathogenic *Escherichia coli* K-12, CRP has been co-opted as a regulator of some plasmid-encoded virulence genes. Hence, we show that CRP directly regulates expression of both *estA2* and *estA1*, which encode the heat-stable toxins. Furthermore, CRP indirectly represses expression of the heat-labile toxin. This work also identifies a role for CRP in controlling transcription of a small open reading frame, imbedded within a gene, at the 3' end of an operon encoding a type I secretion system.

Intriguingly, CRP binding sites on the virulence plasmids p948 and p666 appear to be bound by H-NS rather than CRP *in vivo*. Since CRP and H-NS are known to respond to metabolic and osmotic flux, we suggest that toxin production is regulated by changes in these triggers.

Acknowledgements

Table of Contents

List of Figures

List of Tables

Abbreviations

A	Adenine
A (Ala)	Alanine
APS	Ammonium persulphate
AR	Activating region
Å	Angstrom
bp	Base pair
BSA	Bovine serum albumin
C	Cytosine
C (Cys)	Cysteine
ChIP-seq	Chromatin immunoprecipitation coupled with next generation sequencing
CIAP	Calf intestinal alkaline phosphatase
CRP	Cyclic-AMP receptor protein
cAMP	3'-5'-cyclic adenosine monophosphate
CTD	Carboxy-terminal domain
D (Asp)	Aspartic acid

ddH₂O	Distilled and de-ionised water
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	2'-deoxyribonucleoside 5'-triphosphate (N=A,C,G,T)
E	RNA polymerase core enzyme
E (Glu)	Glutamic acid
E. coli	*Escherichia coli*
EDTA	Diaminoethanetetra-acetic acid
EMSA	Electrophoretic mobility shift assay
ETEC	Enterotoxigenic *Escherichia coli*
F (Phe)	Phenylalanine
Fis	Factor for inversion stimulation
FNR	Fumarate and nitrate reductase
G	Guanosine
G (Gly)	Glycine

His (H)	Histidine
H-NS	Histone-like nucleoid structuring protein
HU	Histone-like protein from *E. coli* strain U93
IPTG	Isopropyl β-D-1- thiogalactopyranoside
K (Lys)	Lysine
kDa	Kilodalton
L (Leu)	Leucine
LB	Lysogeny broth
LT	Heat-labile toxin
M (Met)	Methionine
Mbp	Mega base pairs
MES	2- (*N*-morpholino)ethanesulphonic acid
Mg	Magnesium
MOPS	3-(*N*-morpholino)propanesulphonic acid
mRNA	Messenger ribonucleic acid
N (Asn)	Asparagine

NGS	Next generation sequencing
NTD	Amino-terminal domain
O	Operator
OD	Optical Density
ONPG	o-nitrophenyl-β-D-galactopyranoside
PCR	Polymerase chain reaction
pppGpp/ ppGpp	Guanosine penta/tetraphosphate
Q (Gln)	Glutamine
R (Arg)	Arginine
RNA	Ribonucleic acid
RNAP	RNA polymerase
RNase	Ribonuclease
S (Ser)	Serine
SDS	Sodium dodecyl sulphate
SELEX	Selective evolution of ligands by exponential enrichment
ST	Heat-stable toxin

StpA	Supression of td⁻ phenotype
T	Thymine
T (Thr)	Threonine
Taq	*Thermus aquaticus*
TEMED	N,N,N',N'-tetramethylethyene diamine
T_m	Melting temperature
Tris	Tris (hydroxymethyl) aminoethane
U	Uracil
V (Val)	Valine
v/v	Volume per volume
w/v	Weight per volume
W (Trp)	Tryptophan
WT	Wildtype
Y (Tyr)	Tyrosine
Zn	Zinc

Chapter 1

Introduction

1.1 Bacterial RNA polymerase

1.1.1 Gene regulation and the central dogma

The central dogma is a core concept of molecular biology. It describes the flow of genetic information from DNA to proteins via an RNA message. Briefly, DNA is organised into packets of information called genes that encode proteins. Genes are transcribed to form messenger RNA (mRNA) by RNA polymerase. These mRNAs are then translated by ribosomes into proteins (Figure 1.1). Transcription and translation are subject to regulation at various stages, with the earliest possible point of regulation being transcription initiation.

1.1.2 Transcription in *Escherichia coli*- RNA polymerase

Transcription in *Escherichia coli* (*E. coli*) requires a single 400 kDa DNA-dependent multi-subunit RNA polymerase (RNAP) (Darst, 2001, Browning and Busby, 2004). A high-resolution structure of RNAP is shown in Figure 1.2A, and a schematic is shown in Figure 1.2B. The core enzyme consists of five subunits; β, β', two α subunits, and ω, all of which have different functions (Zhang *et al.*, 1999). Early cryo-EM studies revealed that the RNA polymerase has an overall 'crab-claw' structure, where the DNA double helix is accommodated between the 'claws' (Darst, 2001). More recent high resolution studies of *Thermus aquaticus* (*Taq*) RNAP core (3.3 Å) and holoenzyme (4 Å), together with crystal structures of the *E. coli* RNAP holoenzyme, has revealed the fine molecular details of RNAP-DNA interactions during transcription initiation (Murakami *et al.,* 2002, Murakami, 2013, Zhang *et al.*, 1999, Zuo *et al.*, 2013).
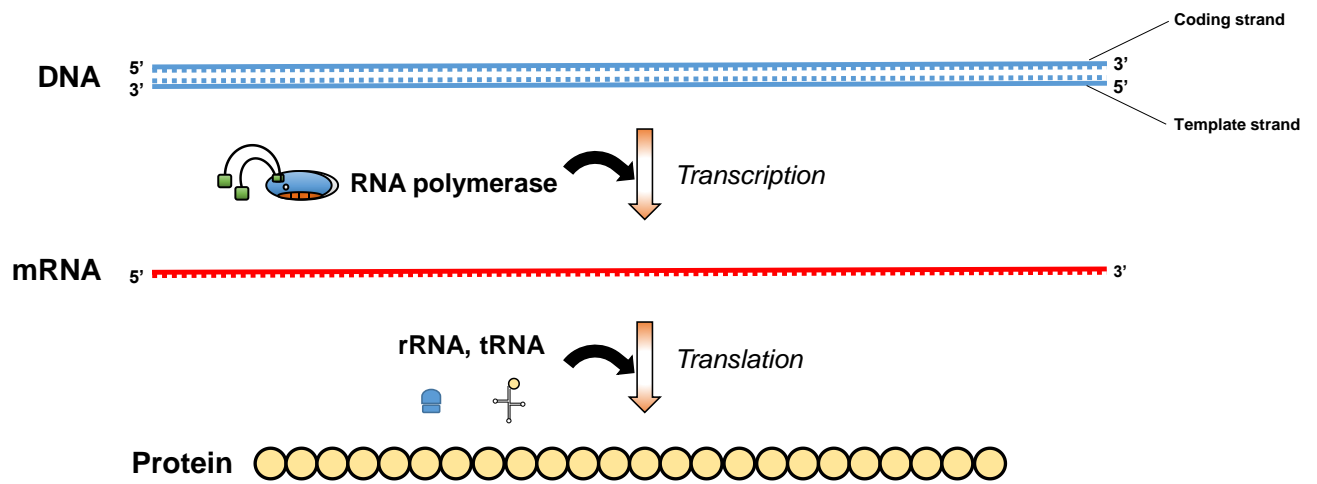
Figure 1.1 Gene expression in *E. coli*

DNA is shown in blue, RNA is shown in red. Dashed ladders represent bases, solid lines represent the sugar-phosphate backbone. Amino acids are shown as cream circles.

A)

- α
- σ
- β'
- β
- ω

α-CTD

Active centre

Secondary channel

σ4

β flap

σ1

σ2

β' main channel clamp

B)

α-CTD II

β-lobe 1

β-lobe 2

Secondary channel

α-CTD I

β' G-loop (trigger loop)

β' F-helix (bridge helix)

β flap

σ4

Front DNA-binding clamp

β' zinc-finger

σ3

β' zipper

σ2

σ1

β' lid

β' rudder

Main channel clamp
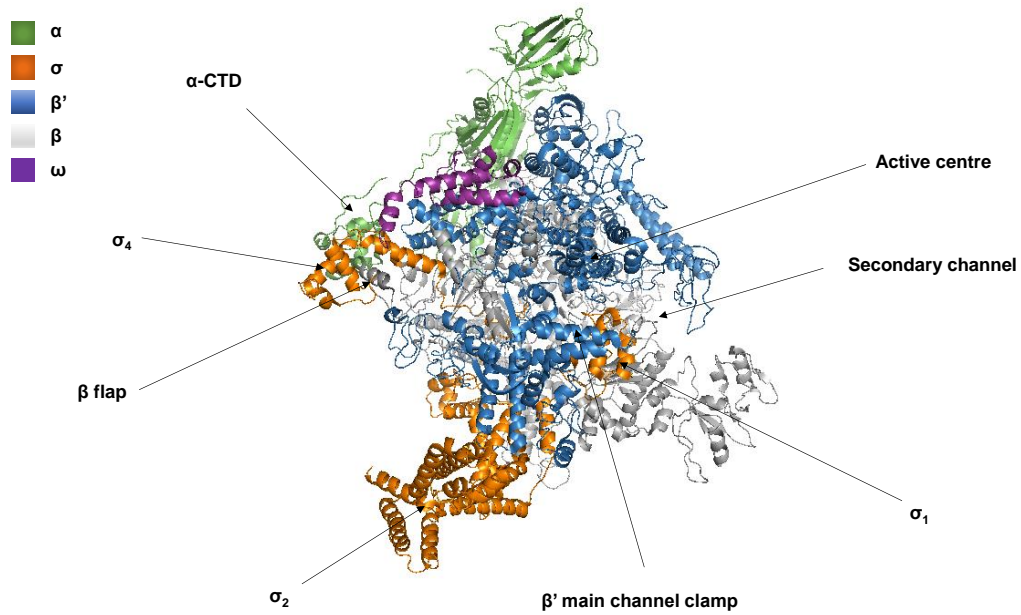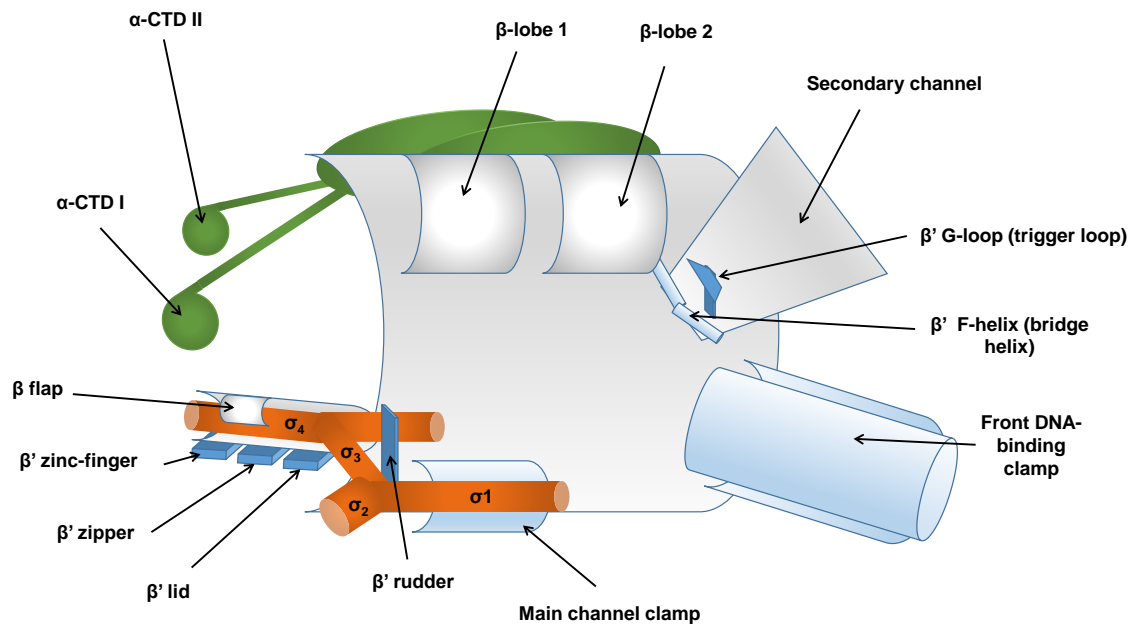
Figure 1.2 Structure of *E. coli* RNAP holoenzyme

A) Crystal structure of *E. coli* RNAP holoenzyme (Murakami, 2013). Diagram contructed using PyMOL. Subunits have been colour coded.

B) The cartoon shows RNAP containing σ[70] (shown in orange). Different parts of the enzyme are labelled. Subunits are colour coded; the α-subunits are shown in green, β' is shown in blue/dark blue, and β is shown in gray. ω is not shown. (Based on Borukhov and Nudler, 2007).

1.1.3 Promoters

Transcription initiates at regions of DNA known as promoters, which are located upstream of genes and transcription start sites. Throughout this work transcription start sites are referred to as "+1", whilst promoter sequences upstream are denoted by a "-" prefix. The core RNAP enzyme (which has the composition $\alpha_2\beta\beta'\omega$) is capable of catalysing RNA synthesis, but cannot initiate transcription specifically at promoters (Browning and Busby, 2004). In order to recognise promoter regions, the core enzyme must associate with a σ-factor, to form the holoenzyme ($\alpha_2\beta\beta'\omega\sigma$) (Ishihama, 1993). The holoenzyme is competent for promoter recognition, DNA strand separation (also known as 'promoter melting') and transcription initiation. To initiate transcription, the RNAP holoenzyme recognises a number of discrete sequences within the promoter region. Promoter recognition involves interactions between the σ-factor and the promoter DNA (Figure 1.3) (Feklistov *et al.,* 2014).

RNAP holoenzyme containing the σ^{70} ($E\sigma^{70}$) recognises two highly conserved sequences located upstream of the transcription start site; the -35 hexamer (consensus 5'-TTGACA-3') and the -10 hexamer (consensus 5'-TATAAT-3') (Saecker *et al.*, 2011, Shultzaberger *et al.*, 2006). These core promoter elements may be augmented with additional sequences. For example, the 'extended -10' element (present at 20 % of σ^{70} promoters) refers to the presence of a 'TGn' motif upstream of the -10 hexamer (Burr *et al.,* 2000, Mitchell *et al.,* 2003).
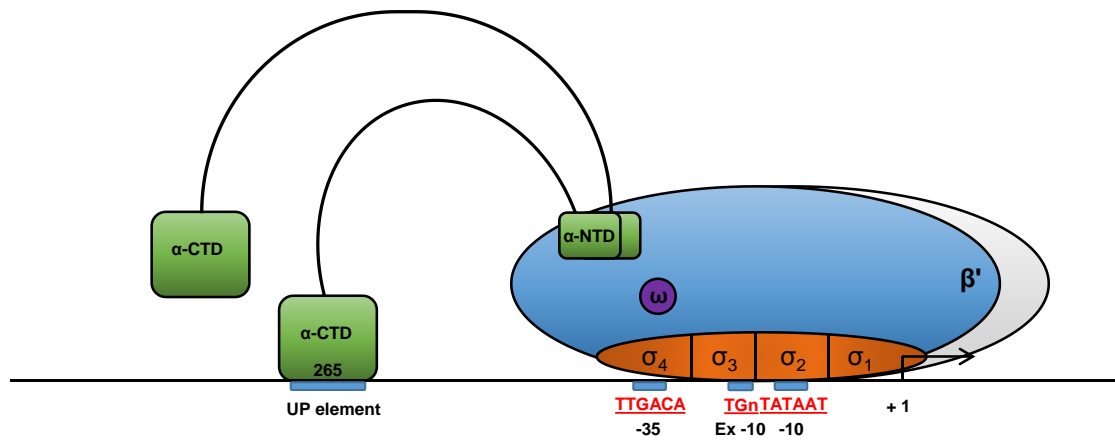
Figure 1.3 Promoter recognition by σ^{70}-containing RNAP holoenzyme

RNAP recognition of promoter DNA. Domain 4 of the σ-factor recognises the '-35 hexamer' (through a helix-turn-helix motif), domain 2 recognises the '-10 hexamer', and domain 3 recognises the 'extended -10' sequence. The "265" determinant of α-CTD recognises the UP element. The α-subunits are shown in green, the σ-factor is orange, the ω subunit is purple, the β is gray, and the β' subunit is shown in blue. Based on Browning and Busby (2004).

1.1.4 The α-subunit

RNAP contains two α subunits, which play a number of functional roles. Each α subunit consists of two domains, which fold independently (Blatter *et al.,* 1994, Gourse *et al.,* 2000). Each α subunit is 37 kDa in size, and 329 amino acids long. Residues 8-235 form the N-terminal domain, and residues 249-329 form the C-terminal domain. The N-terminus is required for dimerisation of the α subunits and the α-NTD dimer serves as a point of assembly for the rest of the polymerase (Figure 1.4). The α-NTD can interact with transcriptional regulators (Blatter *et al*, 1994, Ebright and Busby, 1995). The α-CTD also interacts with transcriptional regulators and is joined to α-NTD by a flexible 13 amino acid linker (Ebright, 1993, Estrem *et al*, 1998). The α-CTD is able to bind DNA. Residue Arg-265 of α-CTD is part of a helix-hairpin-helix motif and recognises the DNA minor groove (Benoff *et al.*, 2002, Ross *et al.*, 2003, Ross and Gourse, 2005). This interaction is favoured by AT-rich DNA sequences, known as 'UP' elements (Ross *et al*, 1993, Busby and Ebright, 1999). Naturally occurring UP elements have been shown to increase promoter activity 20-30 fold. (Ross *et al.*, 1993). Furthermore, synthetic UP elements can have even greater stimulatory effects. In a study of random sequences fused upstream of a *rrnB* P1 promoter-*lacZ* construct, Estrem *et al.* (1998) used SELEX (Systematic Evolution of Ligands by Exponential enrichment), to identify a 5'–nnAAA(A/T)(A/T)T(A/T)TTTTnnAAAnnn-3' consensus sequence for the UP elements. The α-CTD also has an important role in transcription activation at promoters which do not contain an UP element. For example, deletion of α-CTD results in a 10-fold reduction in RNAP binding at the *lacUV5* promoter, which lacks an UP element (Ross and Gourse, 2005).

7
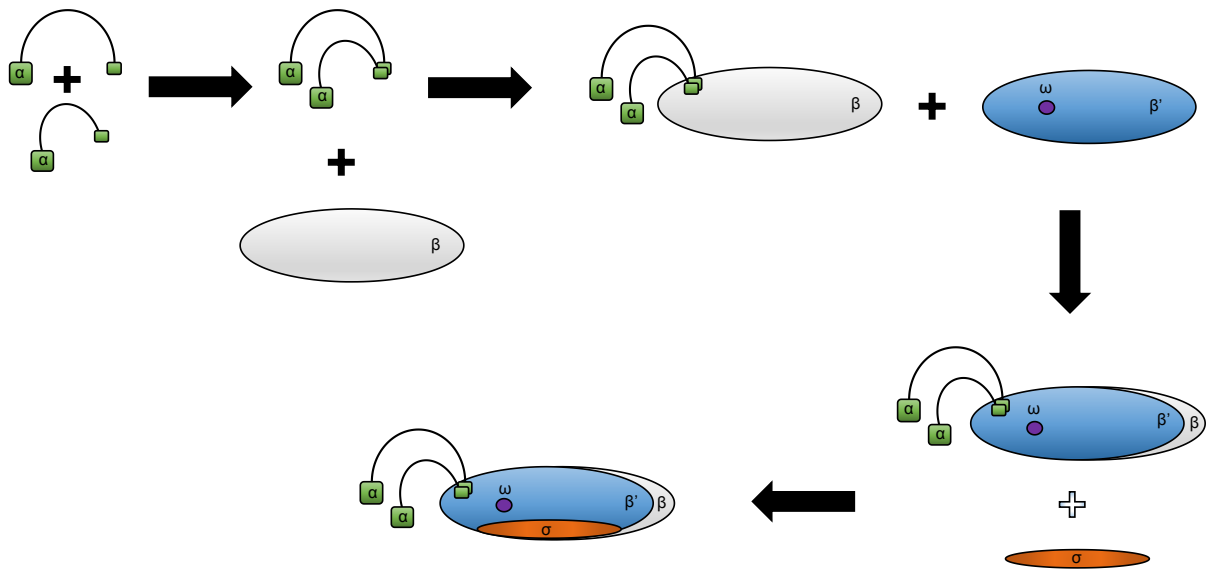
Figure 1.4 Assembly of RNAP

Assembly of the RNAP holoenzyme occurs in the order; α_2-$\alpha_2\beta$- $\alpha_2\beta\beta'\omega$- $\alpha_2\beta\beta'\omega\sigma$. The σ-factor associates before promoter recognition (Ishihama, 1990). The α-subunits are shown in green, the σ-factor is orange, the ω subunit is purple, the β is gray, and the β' subunit is shown in blue. Based on Mathew and Chatterji (2006).

1.1.5 β and β' subunits

The β and β' subunits form the claw and the active site of RNAP. The active centre consists of a Mg^{2+} co-factor, held in position by three aspartate residues. These residues are form a highly conserved amino acid motif (NADFDGD) in the β' subunit (Darst, 2001, Dieci *et al.,* 1995, Vassylyev *et al.*, 2002, Zhang *et al.*, 1999). There are two further important structural features on the active site; the F-bridge helix and the G-trigger loop (Figure 1.2B). These are critical mobile parts which play important roles in NTP selection and RNAP translocation (Vassylyev *et al.*, 2007). Briefly, RNAP elongates through a ratchet mechanism, where conformational changes in the F-bridge helix drive translocation of the elongation complex. The G-trigger loop plays an important role by controlling the conformation of the F-bridge helix. Single amino acid mutations in either the trigger loop or bridge helix can give rise to altered fidelity and processivity of RNAP (Bar-Nahum *et al.*, 2005). Structurally, RNAP can be thought of as having four mobile parts which move relative to the active site; these are the clamp, β flap, and the β1 and β2 lobes (shown in Figure 1.2B) (Murakami *et al.*, 2002, Murakami and Darst, 2003).

1.1.6 The ω subunit

The ω subunit is a small 91 amino acid protein that binds the β' subunit, and wraps around the β' CTD (Zhang *et al.*, 1999). The role of the ω subunit remains enigmatic, although some reports suggest ω may act as chaperone for the β' subunit (other subunits appear to be folded by the GroEL chaperone) (Minahkin *et al.,* 2001, Mathew & Chatterji, 2006). The best characterised action of ω is its ability to render RNAP sensitive to the alarmone ppGpp (Gentry and Burgess, 1989, Potrykus and Cashel, 2008, Zuo *et al.*, 2013).

1.1.7 σ-factors

The holoenzyme of *Taq* RNAP (containing σ^A, the housekeeping σ^{70} homologue) has been solved to a resolution of 4 Å, both in the absence and presence of fork-junction DNA (Murakami *et al.*, 2002, Murakami *et al.*, 2002). Other high resolution crystal structures (2.6 Å) have been solved for *Thermus thermophilus* RNAP holoenzyme (Vassylylev e*t al.*, 2002). In these structures, σ faces upstream, and lies on a surface of core RNAP, with which it makes multiple stable contacts.

The σ^{70} factor contains four domains (σ_4, σ_3, σ_2, and σ_1), as illustrated in Figure 1.5 (Campbell *et al.*, 2002). The domains are divided into regions (e.g. domain 4 contains regions 4.1 and 4.2). All four domains contact core RNAP but two key contacts are conserved; domain σ_4 binds an α-helix of the β subunit flap, and σ_2 forms an important interaction with the β' subunit coiled-coil (Campbell *et al.*, 2002, Campbell *et al.,* 2008, Murakami and Darst, 2003, Sharp *et al.,* 1999). The σ_4 domain accommodates the β flap helix in a hydrophobic pocket. Consequently, this region of the holoenzyme is flexible, which in turn facilitates promoter bending during transcription initiation (Murakami *et al.*, 2002). The σ_1 domain is specific to σ^{70}, and region $\sigma_{1.1}$ plays an important role in preventing non-specific initiation by occupying the main RNAP channel. Region $\sigma_{1.1}$ is displaced from the channel upon promoter recognition. To facilitate its movement, the $\sigma_{1.1}$ region is linked to the rest of σ through a 37 amino acid linker (Mekler *et al.,* 2002). Region $\sigma_{1.1}$ also has roles in transcription regulation, and can regulate DNA entry into the active site depending on features of the bound promoter. Domain σ_1 also inhibits free σ from binding to DNA (Hook-Barnard and Hinton, 2009, Murakami and Darst, 2003).
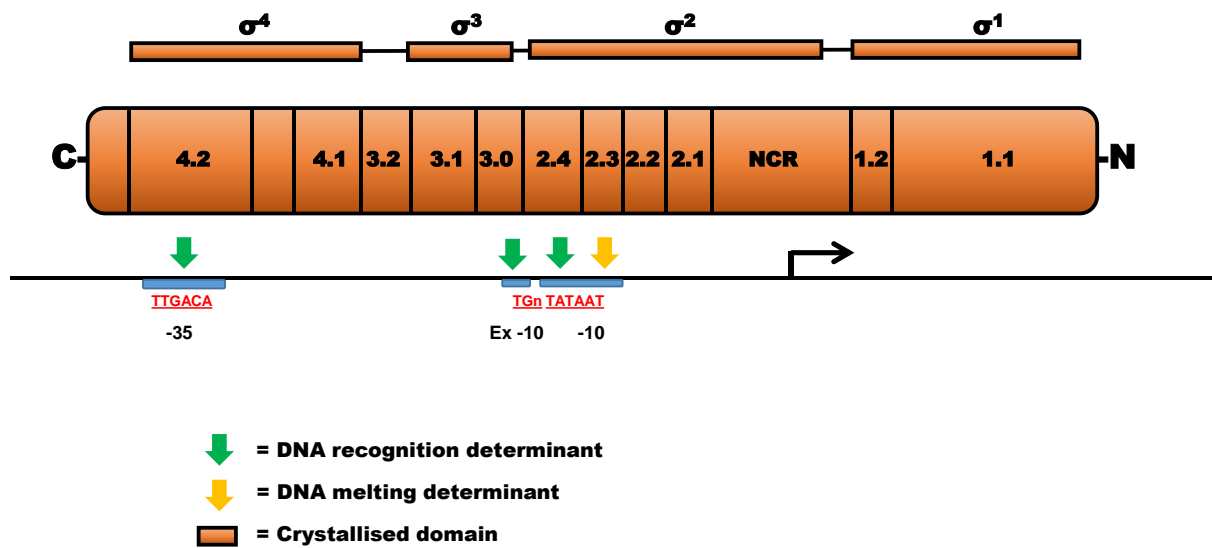
Figure 1.5 Generalised σ⁷⁰-family domain organisation

The figure shows the generalised structure of σ^{70}-family members. The upper schematic shows the domain structure, and the lower schematic shows the location of different regions. Green arrows show interactions between σ^{70} with DNA. Yellow arrows show regions involved in DNA melting. Note that σ_1 is present only in σ^{70} in *E. coli*. NCR is the non-conserved region. Based on Gruber and Gross (2003).

1.2 Transcription initiation

Initiation of transcription consists of four separate events; i) promoter recognition, ii) formation of an "open complex", iii) DNA scrunching (abortive initiation), and iv) promoter escape (Figure 1.6) (Browning and Busby, 2004, Murakami and Darst, 2003).

1.2.1 Promoter recognition

Domain σ_4 of σ^{70} recognises the -35 hexamer through a helix-turn-helix motif in region 4.2. Thus, region 4.2 residues Lys-593, Gln-589, Arg-588, Arg-586, Glu-585, Arg-584, Thr-583, Glu-574, Leu-573, Arg-562 and Arg-554 of *E. coli* σ^{70} are involved in recognition of the -35 element. Gln-589 contacts the base at positions -35, Arg-586 contacts the bases at positions -35 and -36, Glu-585 contacts the bases at positions -34 and -33, and Arg-584 makes direct contact with the bases at positions -31 and -30 (Campbell *et al.,* 2002, Murakami *et al.*, 2002). Domain σ_2 contacts the -10 hexamer of promoter DNA through a single helix. Important residues involved in recognising the -10 element include; Ser-428, Asn-383, Leu-386 and Leu-110, which interact with the base at position -7. The base at position -8 interacts with Ser-428, and Thr-432, whilst Phe-419 interacts with the base at position -10. The most important interaction occurs between sidechains of Tyr-430, Arg-423, Glu-420, Phe-419 and the base at position -11 (Feklistov and Darst, 2011). The extended -10 element is bound by residues Glu-458 and His-455 in a helix of domain σ_3 (Barne *et al.,* 1997). The -35 and UP elements are contacted first by RNAP, followed by the -10 element recognition (Feklistov and Darst, 2011, Lee *et al.* 2012).
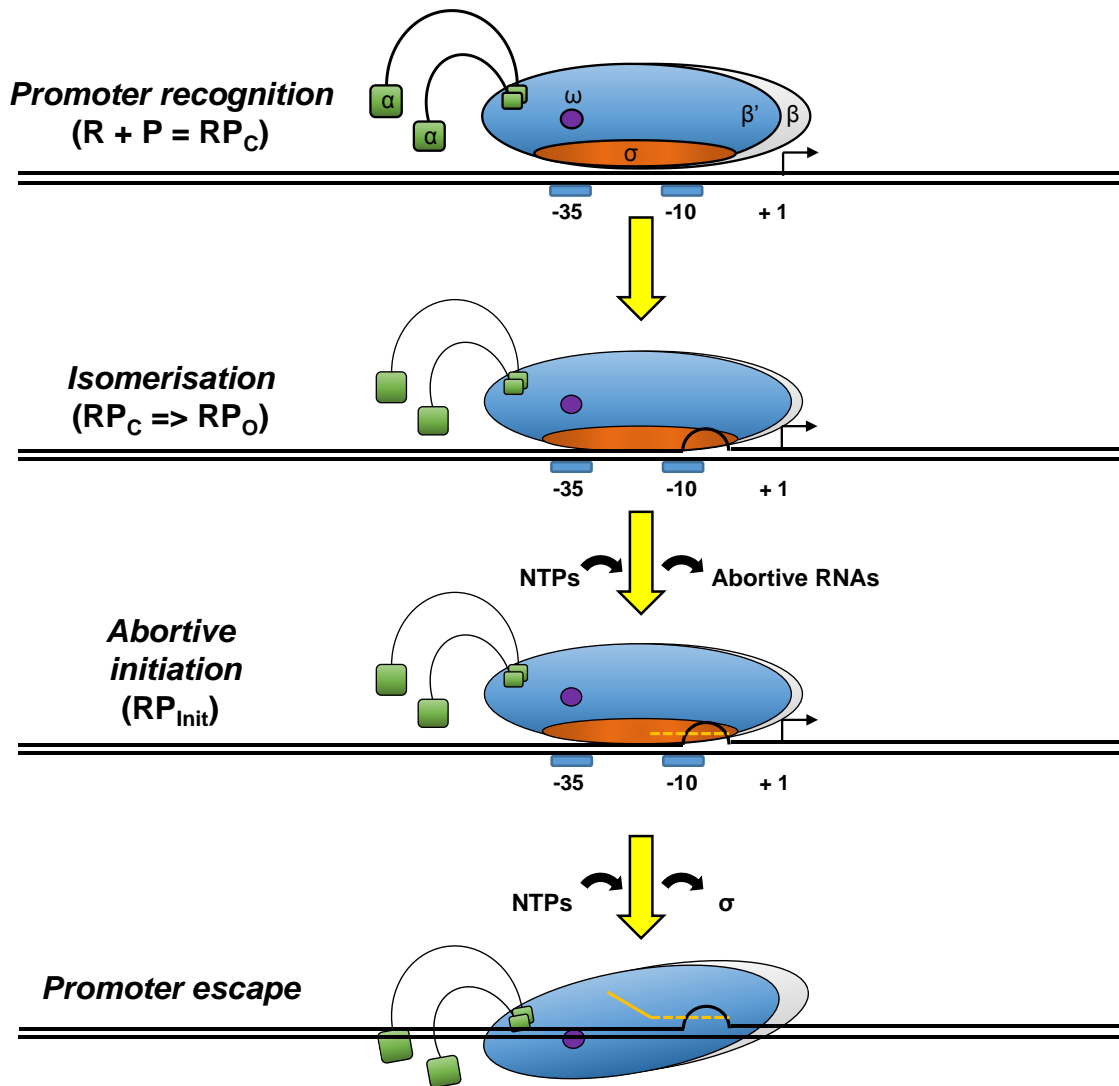
Figure 1.6 Transcription initiation in *E. coli*

Open complex formation at promoter regions follows four stages; binding/ promoter recognition, isomerisation, abortive initiation and promoter escape. DNA is depicted as solid black lines, the yellow line represents the nascent RNA chain. The α-subunits of RNAP are shown in green, the σ-factor is orange, the ω subunit is purple, the β is gray, and the β' subunit is shown in blue. Based on Browning and Busby (2004).

1.2.2 Isomerisation (open complex formation)

Isomerisation of the -10 element (open complex formation) occurs simultaneously with -10 recognition. Hence, bases A_{11} and T_7 are 'flipped out' of the DNA base stack and accommodated by conserved residues in region $\sigma_{2.3}$ (Feklistov and Darst, 2011). Consequently, these two bases are the best conserved in -10 hexamers (Shultzaberger *et al.,* 2006). Note that only position T_{12} in the -10 element is recognised when DNA is in a double-stranded form (Feklistov *et al.,* 2014, Fenton and Gralla, 2001). Hence, all contacts from A_{11} to T_7 are made only when DNA is single-stranded (Feklistov and Darst, 2011, Hook-Barnard and Hinton, 2007). Unwinding of the -10 element leads to formation of a 12/13 nucleotide 'transcription bubble'. The enzyme 'claw' closes around the DNA, which is fed into the active site, and the transcription bubble is maintained by the β' rudder and the β1 lobe (Figure 1.7) (Borukhov and Nudler, 2003, Borukhov and Nudler, 2007).

1.2.3 DNA scrunching (abortive initiation)

Following isomerisation, a number of abortive initiation cycles commence, producing small abortive RNAs 2-15 nucleotides in length (Goldman *et al.,* 2009, Murakami *et al.,* 2002). During this initial RNA synthesis, the position of RNAP holoenzyme, relative to the promoter DNA, does not change. Rather, DNA 'scrunching' occurs, whereby DNA is taken up and accommodated inside the polymerase as abortive initiation cycles occur (Kapanidis *et al.*, 2006, Revyakin *et al.,* 2006,). The abortive initiation cycles likely represent a build-up of free energy in the polymerase, which allows the polymerase to release bound promoter elements and proceed to productive elongation (Feklistov *et al.,* 2014, Murakami *et al.,* 2003).

A

- α-CTD II
- α-CTD I
- β-lobe 1
- β-lobe 2
- Secondary channel
- β' G-loop (trigger loop)
- β' F-helix (bridge helix)
- Front DNA-binding clamp
- β flap
- −35
- σ₄
- σ₃
- σ₂
- σ1
- β' zinc-finger
- β' zipper
- β' lid
- −10
- β' rudder
- Main channel clamp
- +15

B

- α-CTD II
- α-CTD I
- β-lobe 1
- β-lobe 2
- Secondary channel
- β' G-loop (trigger loop)
- β' F-helix (bridge helix)
- Front DNA-binding clamp
- β flap
- −35
- σ₄
- σ₃
- σ₂
- σ1
- β' zinc-finger
- β' zipper
- β' lid
- −10
- β' rudder
- Main channel clamp
- +15

Figure 1.7 RNAP promoter recognition (A) and in open complex (B)

After promoter recognition, DNA is fed into the main channel and DNA-binding clamp. In both cartoons, the α-subunits of RNAP are shown in green, the σ-factor is orange, the ω subunit is purple, the β is gray, and the β' subunit is shown in blue. DNA is shown as a black line. Based on Borukhov and Nudler (2007).

1.2.4 Promoter escape

Once a nascent RNA reaches 15 nucleotides in length, elongation becomes productive, and the σ-factor is released. Release is driven by changes in RNAP that make it more processive and stably bound to the DNA (Murakami *et al.,* 2002, Vassylyev *et al.*, 2002). Some elongation complexes may retain (or rebind σ) and this has been implicated in transcriptional pausing (Mooney *et al.,* 2005, Reppas *et al.,* 2006, Wade and Struhl, 2004). Once the ternary elongation complex is formed transcription continues until termination, where the core polymerase is released and is available for σ-rebinding (Murakami and Darst, 2003).

1.3. Transcriptional regulation in *E. coli*

1.3.1 Modulation of RNA polymerase by small molecules- (p)ppGpp and ω

A number of small molecules are capable of binding to RNAP and regulating transcription. The best examples of this are guanosine 5'triphosphate, 3' diphosphate (pppGpp), and guanosine 5' diphosphate, 3' diphosphate (ppGpp). (p)ppGpp is produced by cells starved of nutrients and (p)ppGpp levels are regulated by RSH (Rel Spo Homologue) family proteins (Potrykus and Cashel, 2008). Proteins in this family can synthesise and degrade (p)ppGpp. Two RSH family proteins are known to control (p)ppGpp concentration in *E. coli*; RelA, which is only able to synthesize (p)ppGpp, and SpoT, which is important for degrading (p)ppGpp. Increased intracellular concentrations of (p)ppGpp trigger the 'stringent response'; a reprogramming of cellular transcription in poor growth conditions that includes a reduction in transcription of rRNA and tRNA genes, and increased expression of genes for amino acid biosynthesis. Regulation by (p)ppGpp occurs through direct interactions with RNAP (Traxler *et al.*, 2008).

Whether (p)ppGpp activates or represses transcription from a particular promoter depends of the half-life of the open complex formed by RNAP (Barker *et al.*, 2001). This is often

dependent on the sequence of the discriminator, the sequence downstream of the -10 element and upstream of the transcription start site (Figueroa-Bossi *et al.*, 1998, Barker *et al.*, 2001). For example, *rRNA* promoters contain GC-rich discriminators, and form open complexes with short half-lives which are sensitive to ppGpp inhibition (Paul *et al.*, 2005). Promoters for genes involved in amino acid biosynthesis tend to have low-GC content discriminator sequences, and form open complexes with longer half-lives, which ppGpp can activate in concert with the DksA protein (Paul *et al.*, 2004).

Early studies (Igarashi *et al.*, 1989) suggested a role for the ω subunit in responding to ppGpp. This requirement for the ω subunit can be complemented by DskA (Vrentas *et al.*, 2005). Recently, the crystal structure of ppGpp complexed with RNAP has been solved (Zuo *et al.*, 2013). In conjunction with cross-linking and mutation analysis (Ross *et al.*, 2013) this revealed the binding site of ppGpp. The interaction site lies between the RNAP β' subunit and the ω subunit. The phosphate groups of ppGpp interact with a pocket between β' and α-NTD, the guanosine base interacts with ω (Ross *et al.*, 2013). This interaction of ppGpp with RNAP reduces the mobility of the enzyme, which slows open complex formation (Zuo *et al.*, 2013).

1.3.2 Regulation by σ-factors

1.3.2.1 σ-factors produced by *E. coli*

Different σ-factors recognise different sets of genes. Consequently the transcriptional output of the cell can be rapidly altered to suit environmental conditions through the use of alternate σ-factors. *E. coli* produces seven σ-factors (σ^{70}, σ^{54}, σ^{38}, σ^{32}, σ^{28}, σ^{24}, and σ^{19}, summarised in Table 1.1). With the exception of σ^{54} all are members of the σ^{70} family and thus are structurally related (Gruber and Gross, 2003). σ^{70}, the 'housekeeping' σ-factor, is essential for viability, and has 1643 binding targets on the *E. coli* genome (Cho *et al.,* 2014, Helmann and Chamberlin,

σ-factor	Function	K_d for binding RNAP core enzyme (nM)	Number of binding sites
σ^{70} / RpoD	General constituative σ-factor	0.26	1643
σ^{54} / RpoN	Nitrogen metabolism	0.3	180
σ^{38} / RpoS	Stationary phase / general stress	4.26	903
σ^{32} / RpoH	Heat-shock	1.24	312
σ^{28} / RpoF	Flagellar assembly	0.74	51
σ^{24} / RpoE	Extracytoplasmic stress	2.43	65
σ^{18} / FecI	Iron starvation	1.73	7

Table 1.1 σ-factors produced by *E. coli*

Information is from Cho *et al.,* 2014, Gruber and Gross, 2003, and Maeda *et al.*, 2000.

1988). Genes which assist in survival during starvation are dependent on σ^{38}, the heat-shock response is activated by σ^{32}, motility is controlled by σ^{28}, the response to extracytoplasmic stress is mediated by σ^{24}, and σ^{19} controls transcription of the *fecABCDE* operon (Braun *et al.*, 2003, Fitzgerald *et al.,* 2014, Gruber and Gross, 2003, Soutourina *et al.*, 1999). These alternative σ-factors have between seven (σ^{19}) and 903 (σ^{38}) binding sites across the *E. coli* K-12 genome. The σ^{54} factor plays an important role in regulating nitrogen metabolism and has 180 binding sites (Cho *et al.,* 2014). σ^{54} activates transcription through a different mechanism similar to that of TATA-box binding proteins of eukaryotes. Hence, σ^{54} has an absolute requirement for activator proteins containing an AAA+ domain and requires ATP hydrolysis to activate transcription (Buck *et al.*, 2000). At σ^{54}-dependent promoters, DNA must be bent to allow contact between the AAA+ activator, and the RNAP holoenzyme containing σ^{54}. This often involves an ancillary transcription factor (Kustu *et al.*, 1989).

Different σ-factors re-direct RNAP to different subsets of promoters according to environmental conditions. However, these subsets frequently overlap. For example, σ^{24} and σ^{70} share 52 binding targets, and σ^{54} and σ^{70} share 180 binding targets. The largest overlap in binding sites is that between σ^{38} and σ^{70}, which share 805 binding sites (Cho *et al.,* 2014). The considerable overlap between σ-factors is thought to provide fine-tuning of gene expression in different conditions. Surprisingly, many of these targets for σ-factors lie within genes. For example, 25% of σ^{32} targets and 58 % of σ^{28} targets lie within genes. Only 17 % of intragenic targets for σ^{28} are associated with transcription of an RNA (Fitzgerald *et al.,* 2014, Wade *et al.,* 2006). Conversely, widespread intragenic transcription can be driven by σ^{70} and this is often suppressed by proteins like the nucleoid associated protein H-NS (Singh *et al.,* 2014).

1.3.2.2 Regulating σ-factor activity

The regulation of alternate σ-factor activity is complex and occurs at all stages of gene expression, including post-transcriptional and post-translational regulation. For instance, under some conditions, σ^{70} can be regulated by the anti-σ factor Rsd (Jishage *et al.*, 2001). Given the comparatively high affinity of σ^{70} for core enzyme, compared to other σ-factors, sequestration of σ^{70} by Rsd may increase the potential for σ-factor competition (Table 1.1). The *rpoS* mRNA (encoding σ^{38}) is post-transcriptionally regulated by H-NS (which down-regulates translation), HU and Hfq (which increase translation). Furthermore, RpoS can degraded by the RssB-ClpXP system (Gruber and Gross, 2003, Hengge-Aronis, 2002). The *rpoH* mRNA (encoding σ^{32}) contains two regions, A and B, in the vicinity of the translation start site. These regions base pair and result in a secondary structure that prevents ribosome binding. The mRNA secondary structure is lost at higher temperatures, which permits translation of σ^{32} (Morita *et al.*, 1999). In non-stressful conditions, σ^{24} is usually inactivated and sequestered by the membrane-tethered RseA protein. Upon detection of unfolded outer membrane proteins a proteolysis cascade is activated, which results in degradation of RseA by the DegS and YaeL proteases, releasing σ^{24} for association with RNAP core enzyme (Alba and Gross, 2004).

Three sets of genes are involved in motility in *E. coli*; early, middle and late genes. The only early transcription unit encodes the FlhD and the FlhC regulators. The middle genes are controlled by the $FlhD_4C_2$ complex, which activates transcription at 11 transcription units by the σ^{70} holoenzyme, including the *fliA* gene which encodes σ^{28}. FlgM, an anti σ-factor, can inhibit σ^{28} activity. Completion of the flagellar basal body and hook results in secretion of FlgM out of the cell, releasing σ^{28}, which activates transcription of the late genes (Chilcott and Hughes, 2000, Fitzgerald *et al.,* 2014).

1.3.3 Transcriptional activators

There are many mechanisms for transcription activation; these can involve direct interaction of transcription factors (TFs) with RNAP or can be indirect (Browning and Busby, 2004, Lee *et al.*, 2012). A well-defined example of a transcriptional regulator is the cyclic AMP receptor protein (CRP). CRP plays a key role in managing carbon metabolism in *E. coli*, and has become a regulatory paradigm. Hence, CRP can activate or repress transcription via many mechanisms (Busby and Ebright, 1999, Shimada *et al.*, 2011). CRP commonly activates transcription by binding upstream of, or overlapping, promoter -35 elements. Depending on the position of the CRP site the activation mechanism is known as class I, II or III (Browning and Busby, 2004). Depending on the class of activation CRP interacts with RNAP using different surfaces. These surfaces are known as activating regions (Busby and Ebright, 1999).

CRP binds DNA as a dimer and three defined activating regions are presented by each monomer (Lawson *et al.,* 2004). Activating region 1 (AR1) consists of amino acid residues 156-164 with Thr-158 being the most important residue. These amino acids form an exposed 11Å by 14Å patch on the surface of CRP, close to a canonical β-turn, adjacent to the helix-turn-helix DNA binding motif (Zhou *et al*, 1993, Niu *et al*, 1994). Activating region 2 (AR2) is located towards the N-terminus of CRP and consists of positively charged residues His-19, His-21 and Lys-101 (Niu *et al.,* 1996). Hence, AR2 is a positively charged surface of CRP. Activating region 3 (AR3) consists of Glu-58, Lys-57, Gly-56, Glu-55, Glu-54, Asp-53 and Lys-52. Hence, AR3 is a negatively charged surface on an exposed β-turn. Note that the positively charged lysine at position 52 hinders the activity of AR3 and subsitution with asparagine converts AR3 into a more potent activating region (Williams *et al.,* 1991, Rhodius and Busby, 2000, Rhodius and Busby, 2000). AR1 contacts RNAP at all classes of CRP activated promoters. At class II promoters (and some class III promoters) AR3 contacts σ^{70} and AR2 interacts with α-NTD (Rhodius *et al.,* 1997).

1.3.3.1 Class I promoters

At class I promoters CRP binds on the same face of the DNA as RNAP and contacts the α-CTD (Figure 1.8A). For example, at the *lac* promoter, CRP binds to a site centred 60.5 bases upstream of the transcription start site (Malan *et al.,* 1984). Note that the positioning of CRP is not fixed. Hence, Gaston *et al.* (1990) used a modified *melR* promoter to show that CRP is able to activate transcription from positions -60.5, -72.5, -82.5 and -92.5 at class I promoters. Thus, correct helical phasing is essential but the CRP site is not fixed. The flexibility in the position of the CRP site is thought to be permitted by the unstructured 44 Å linker of the α subunit, together with the DNA bend induced by CRP binding (Blatter *et al.,* 1994, Busby and Ebright, 1994, Schultz *et al.,* 1995). At some promoters "helical phasing" rules may be relaxed depending on the promoter elements present. Specifically, the presence of an UP element, or a consensus extended -10 element, can compensate for CRP mis-positioning (Zhou *et al.*, 2014). AR1 of the downstream subunit is the only determinant of CRP required at class I promoters, but there are three components of α-CTD which are important for activation at class I promoters (Zhou *et al*., 1993). Firstly, as discussed earlier, α-CTD residue Arg-265 is essential for DNA binding (Gaal *et al.,* 1996). Secondly, the α-CTD '287' determinant (Thr-285, Glu-286, Val-287, Glu-288 and Arg-317) forms a 14Å by 23Å surface that interacts with AR1 (Savery *et al.,* 2002). Thirdly, at some class I promoters, where the CRP site is centred at -60.5, one α-CTD is sandwiched between CRP and RNAP σ_4 (Figure 1.8A) (Benoff *et al.,* 2002). Here residues of the α-CTD 261 determinant (Asp-258, Asp-259 and Glu-261) contact residues in σ_4 (593-604). In fact, this interaction has been shown to be important at both activator-dependent and -independent promoters (Chen *et al*., 2003, Lee *et al.,* 2003, Ross and Gourse, 2005, Savery *et al.*, 2002).
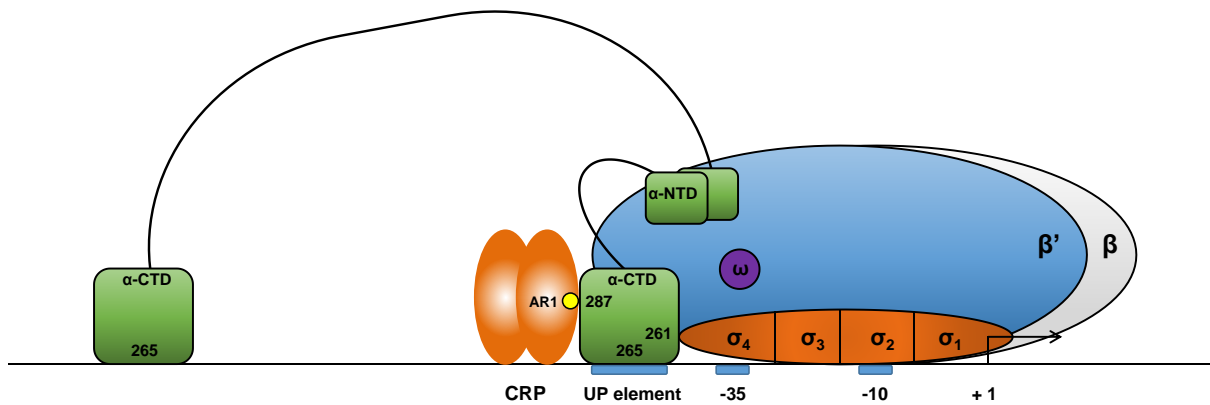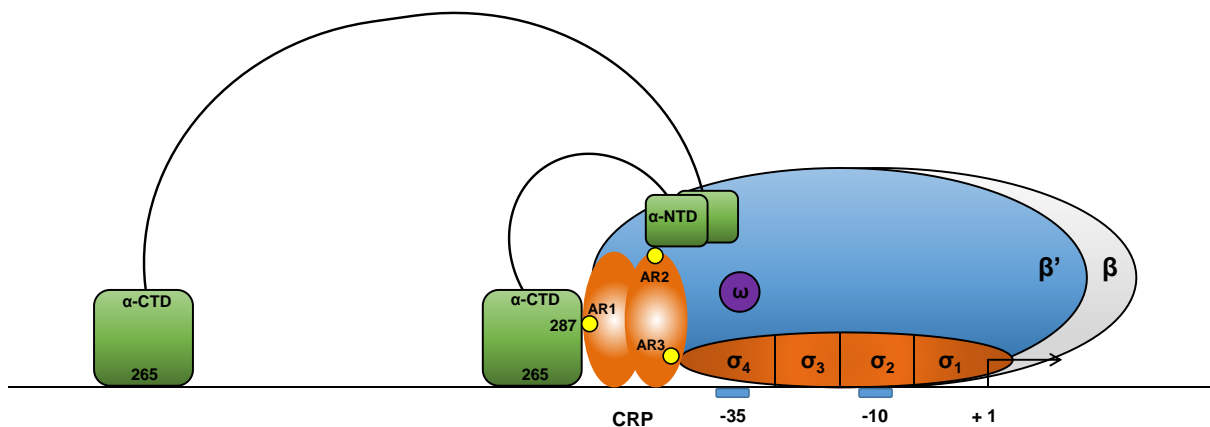
Figure 1.8 Transcription activation by CRP at Class I (A) and Class II (B) promoters

RNAP subunits are labelled. CRP activating regions are shown as yellow dots. DNA is shown as a black line. Based on Busby and Ebright (1999).

A. Class I activation by CRP at a -60.5 position. The '261' determinant on α-CTD can contact σ-factor region 4 (residues 573-604). Only the downstream subunit of CRP is contacted by α-CTD '287' determinant (the other α-CTD binds non-specifically at upstream DNA). The '265' determinant is involved in binding DNA, at UP elements, if present.

B. Class II activation. RNAP contacts both subunits of CRP, which is sited at a -41.5 position. (The other α-CTD binds non-specifically at upstream DNA). AR1 of CRP contacts the '287' determinant of α-CTD, AR2 contacts α-NTD. An UP element is not essential but does enhance transcription. The '265' determinant is involved in binding DNA, at UP elements, if present.

1.3.3.2 Class II promoters

At class II promoters, CRP binds to a site overlapping the -35 hexamer, centred at position -41.5, and interacts with the σ and α subunits of RNAP (Figure 1.8B). AR1, on the upstream CRP subunit, contacts the '287' determinant on one α-CTD (Savery *et al.,* 1998). This interaction serves to recruit RNAP to the promoter since mutation of AR1 reduces initial binding by RNAP 8-fold (West *et al.*, 1993, Niu *et al.,* 1996). AR2 of the downstream subunit interacts with residues Glu-165, Asp-164, Glu-163 and Glu-162 on α-NTD (Niu *et al.*, 1996). These interactions stimulate isomerisation and deletion of AR2 reduces open complex formation 10-fold. If unmasked, AR3 of the downstream subunit interacts with σ^{70} region 4.2 (residues 593-603) (Rhodius *et al.*, 2000). At both class I and class II promoters, the '265' determinant of α-CTD interacts with the DNA minor groove and this is enhanced by UP elements (Gaal *et al.,* 1996).

1.3.3.3 Class III promoters

Many bacterial promoters respond to multiple environmental signals and have a more complex organisation than the simplistic examples described above. Hence, they may contain sites for different TFs or multiple sites for the same TF (Figure 1.9) (Browning and Busby, 2004, Busby and Ebright, 1999). At the artificial *CC(-93.5) CC (-61.5)* promoter, activation is dependent on occupation of CRP sites centred at positions -93.5 and -60.5 (Tebbutt *et al*., 2002). Both sites must be bound by CRP, with each CRP dimer being contacted by an α-CTD (Lloyd *et al.,* 2002). Activation of the *acs*P2 is achieved through occupation of a CRP sites centred at positions -69.5 and -122.5. At both sites, CRP contacts α-CTD via the 287 determinant (Beatty *et al.,* 2003).
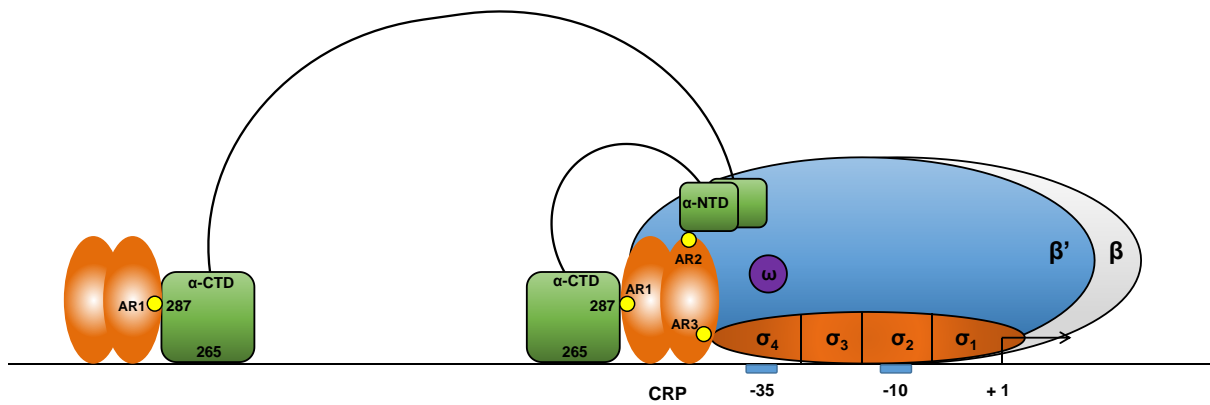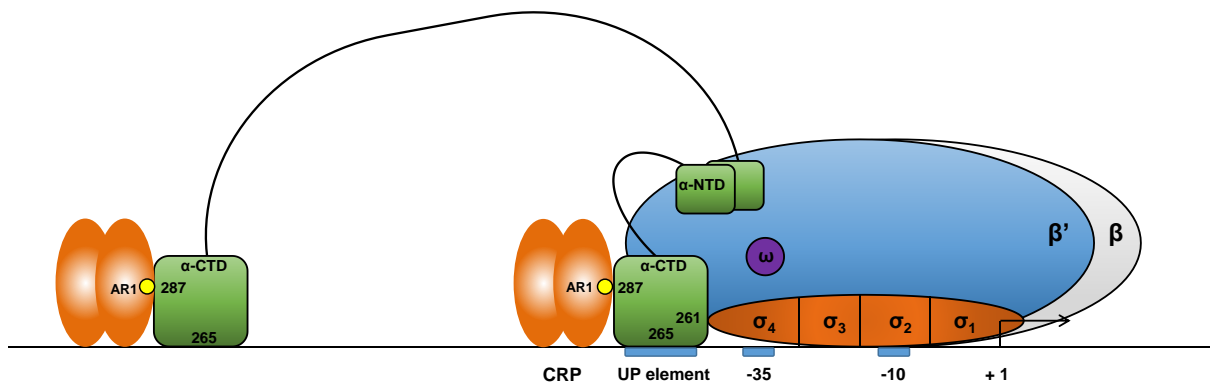
Figure 1.9 Transcription activation by CRP at class III promoters

RNAP subunits are labelled, DNA is shown as black line, activating regions are shown in yellow. Class III activation by CRP. Both α-CTD subunits are involved in binding multiple activators at different positions (A and B). Based on Busby and Ebright (1999).

Two distinct TFs may also work together to activate transcription. For example, at the *ansB* promoter, a CRP site is found 91.5 bp upstream of the transcription start site, and an FNR site is located 41.5 bp upstream of the transcription start site. Activation is dependent on occupation of both sites (Scott *et al.*, 1995).

1.3.4 Other mechanisms of transcriptional activation

The mechanisms of transcriptional activation described above require direct interactions between CRP and RNAP. However, transcription can be activated by re-modelling promoter DNA topology. A well-studied example of this is regulation by MerR at the *mer* operon promoter (Hobman, 2007). Transcription is activated by binding of MerR to the spacer region between the -35 and the -10 elements, which leads to unwinding of the DNA (Figure 1.10A). This unwinding results in correct alignment of the promoter -35 and -10 elements for recognition by RNAP (Ansari *et al.,* 1992).

At some promoters, activation is dependent on co-operative binding by two transcription factors (Figure 1.10B). For example, at the *melAB* promoter, CRP binds to a degenerate site 81.5 bases upstream of the transcription start site. Binding to this site requires MelR, a TF that responds to melibiose. Co-operativity is dependent on direct MelR-CRP interactions. Hence, MelR recruits CRP to its target (Grainger *et al.,* 2004, Wade *et al.,* 2001).

In other scenarios, transcription activation can be activated by re-positioning an activator from a non-stimulatory binding site (Figure 1.10C). This is exemplified by MalT and CRP at the *malK* promoter. Here, MalT binds with high affinity to three non-activatory sites upstream of the promoter. Activation requires i) binding of MalT to two distal binding sites further upstream and ii) CRP binding at three sites in between the distal and proximal set of MalT binding sites. When all sites are occupied the three promoter proximal MalT molecules are repositioned, which results in promoter activation (Richet *et al.,* 1991).
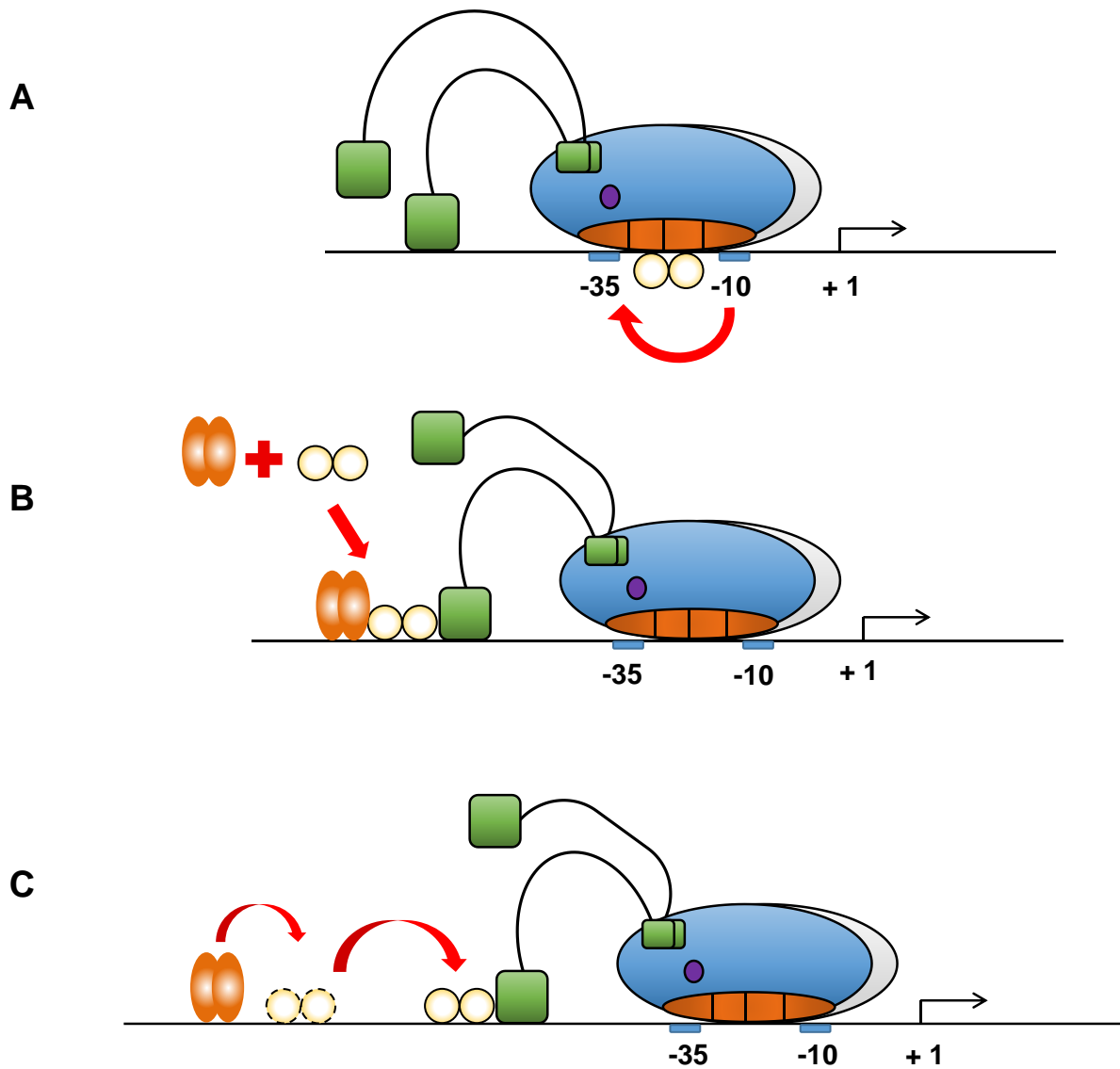
Figure 1.10 Additional mechanisms of transcription activation

 The figure shows other mechanisms of transcriptional activation through:

A) altering the conformation of the core promoter,

B) co-operative binding to DNA by one of more TFs and

C) re-positioning of an primary activator by binding a secondary activator.

In each panel, the primary activator is shown in yellow, the secondary activator is shown in orange. DNA is shown as a straight black line. Based on Browning and Busby (2004).

1.3.5 CRP structure

The structure of CRP (Figure 1.11) has been solved to increasing resolution using X-ray crystallography (McKay *et al.,* 1982, Parkinson *et al*., 1996, Passner and Steitz, 1997, Passner e*t al.,* 2000) and NMR spectroscopy (Popvych *et al.*, 2009). The studies show that CRP consists of an N-terminal domain, which binds cAMP through a series of β-sheets (residues 1-136), and a C-terminal domain (residues 139-209) that contains the helix-turn-helix DNA binding determinant (Busby and Ebright 1999, Passner *et al.,* 2000). Each monomer contains 12 β-sheets, and 6 α-helices (A-F) (Passner *et al*, 2000). Helix C (residues 112-133) is involved in CRP dimerisation (Weber and Steitz, 1987) whilst helices E (residues 168-176) and F (residues 180-191) comprise the DNA binding helix-turn-helix motif (Passner and Steitz, 1997, Fic *et al.*, 2009). Each CRP monomer can bind two molecules of cAMP; one in a high-affinity *anti*-conformation in the N-terminal domain (Weber and Steitz, 1987), and one in a lower-affinity *syn* conformation in the C-terminal domain (Passner and Steitz, 1997, Tutar, 2008). Binding of CRP in the *syn* conformation occurs only at millimolar concentrations of cAMP and is unlikely to be physiologically relevant (Fic *et al.*, 2009, Harman, 2001).

NMR spectroscopy (Popovych *et al.,* 2009) revealed the mechanism of cAMP-induced modulation of CRP activity. Binding of cAMP to CRP in the *anti*-conformation stabilises an extension of the dimerization helix by three turns, which subsequently rotates the DNA-recognition helices by 60°, bringing them closer together. This positions the F-helix optimally for sequence-specific interactions with the DNA major groove. The hinge formed by Leu-137 and Asp-138 links the N-terminal domain and C-terminal domain, and facilitates this conformation change. The extension of the 'C' helix occurs through Val-126 to Phe-136, which is unstructured in the absence of cAMP but is stabilised by binding of cAMP to CRP.
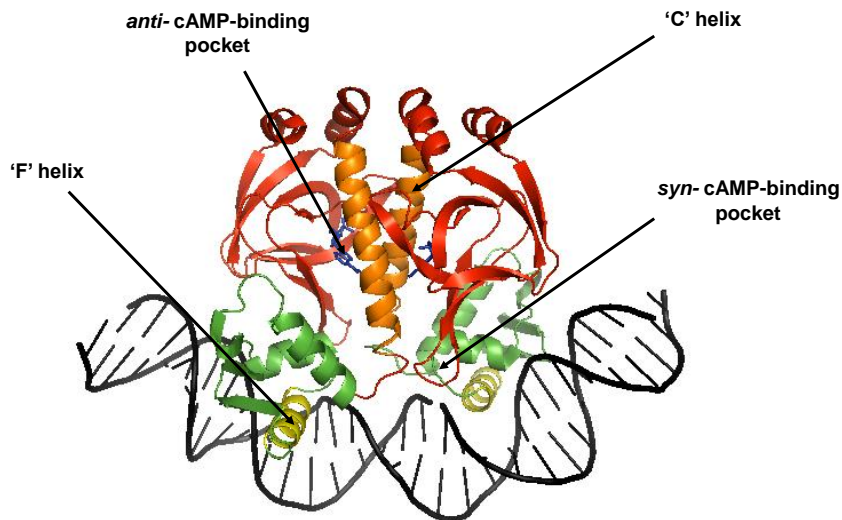
Figure 1.11 Crystal structure of CRP complexed with adenosine 3'5' cyclic monophosphate and DNA

The N-terminal domain of each monomer is shown is shown in red, the C-terminal domain is shown in green. The dimerisation helices (helix "C") is shown in orange, the DNA-binding helices (helix "F") are shown in yellow. cAMP bound to the high-affinity *anti*-cAMP binding sites is shown in blue. DNA is shown in black. The image was generated using PyMOL (Parkinson *et al.,* 1996).

Three interactions are made between cAMP and CRP. Firstly, the adenine base of cyclic AMP contacts Thr-127 and Ser-128, the latter of which is rotated by 120° relative to the cAMP binding pocket. Secondly, the sugar-phosphate group of cAMP interacts with amino acids Gly-71 to Ala-84. Thirdly, binding of cAMP removes Trp-85 from the cAMP binding site. (Popovych *et al*., 2009). Interestingly, CRP can be bound by cGMP. However, this does not trigger the conformational change described above (Fic *et al.,* 2009). When not bound by cAMP, CRP is not able to bind DNA specifically.

The cAMP-CRP complex binds to DNA at a 22bp sequence, which has the consensus 5'-AAA<u>TGTGA</u>TCTAGA<u>TCACA</u>TTT-3'. Sequence recognition occurs through 'direct readout', where amino acids directly recognise the nucleotide bases of binding site, and indirect readout, which involves recognition of the DNA sugar-phosphate backbone. Critical amino acid-base interactions are made within two 5 bp half-sites, which are underscored above. Key amino acid residues of CRP involved in interactions with nucleotide bases are Arg-185 (which contacts the guanosine base at position 7, relative to the 5' end of the site, and the thymine base at position 8 on the opposite strand), Glu-181 (which contacts the cytosine base at position 7 on the opposite strand) and Arg-180 (which contacts the guanosine base at position 5) (Busby and Ebright 1999, Ebright *et al.,* 1989, Lawson *et al.*, 2004). Binding of CRP results in an 80° bend occurring in DNA (Schultz *et al.,* 1991, Lawson *et al.,* 2004).

1.3.6 Catabolite repression

Carbon catabolite repression (CCR) is a phenomenon observed in most heterotrophic bacteria, and is a metabolic strategy for regulating genes that facilitate the use of alternate carbon sources. Often, this regulation involves changes in intracellular cyclic AMP concentrations (Botsford and Harman, 1992). In the *E. coli* glucose phoshophotransferase system, glucose is

phosphorylated to glucose 6-phosphate by a phosphorelay, known as a phosphotransferase system (PTS) as it enters the cell (Saier, 1998) (Figure 1.12). Phospho-enol pyruvate (PEP) supplies the phosphate group at the beginning of the phosphorelay, and is converted to pyruvate. The phosphate group supplied is eventually transferred to EIIA^{Glc} (enzyme IIA) through EI (enzyme I) and HPr (histidine protein). EIIA transfers the phosphate group from a membrane-bound complex consisting of EIIBC, which is responsible for phosphorylating glucose upon import (Görke and Stűlke, 2008). If glucose is not imported, EIIA^{Glc}, a subunit of EII^{Glc}, remains phosphorylated.

The presence of phosphorylated EIIA^{Glc}, together with an unknown factor ('Factor X'), results in the activation of membrane-bound adenyl cyclase (encoded by *cya*), which produces cyclic adenosine 3',5' monophosphate (cAMP) (Hogema *et al.,* 1998). When grown in media containing different glucose concentrations, Notley-McRobb *et al.,* (1997) have shown that cAMP concentrations are not directly linked to glucose concentrations, but instead exhibit two distinct upshifts when glucose concentrations decrease. The first increase occurs when glucose levels fall below 0.3 mM, and the second occurs when glucose concentrations fall below 1 μM, although the precise threshold varies between *E. coli* strains. Growth in minimal media resulted in intracellular cyclic AMP concentrations increasing 10-fold from 20 μM up to 180 μM (Green *et al.,* 2014, Notley-McRobb *et al.,* 1997). Hence, the cAMP-CRP complex activates expression of many genes involved in the use of alternate carbon sources and these genes are subject to CCR (Busby and Ebright, 1999, Görke and Stűlke, 2008).

EIIA^{Glc} also plays a noteworthy role in CCR. Unphosphorylated EIIA^{Glc} inhibits transporters which may import secondary carbon sources, when a preferred carbon source (usually glucose) is available (Inada *et al*, 1996). For example, at the *lac* operon, binding of allolactose (a product of lactose metabolism) to the Lac repressor (LacI) inhibits LacI DNA- binding (Bell and Lewis, 2000).
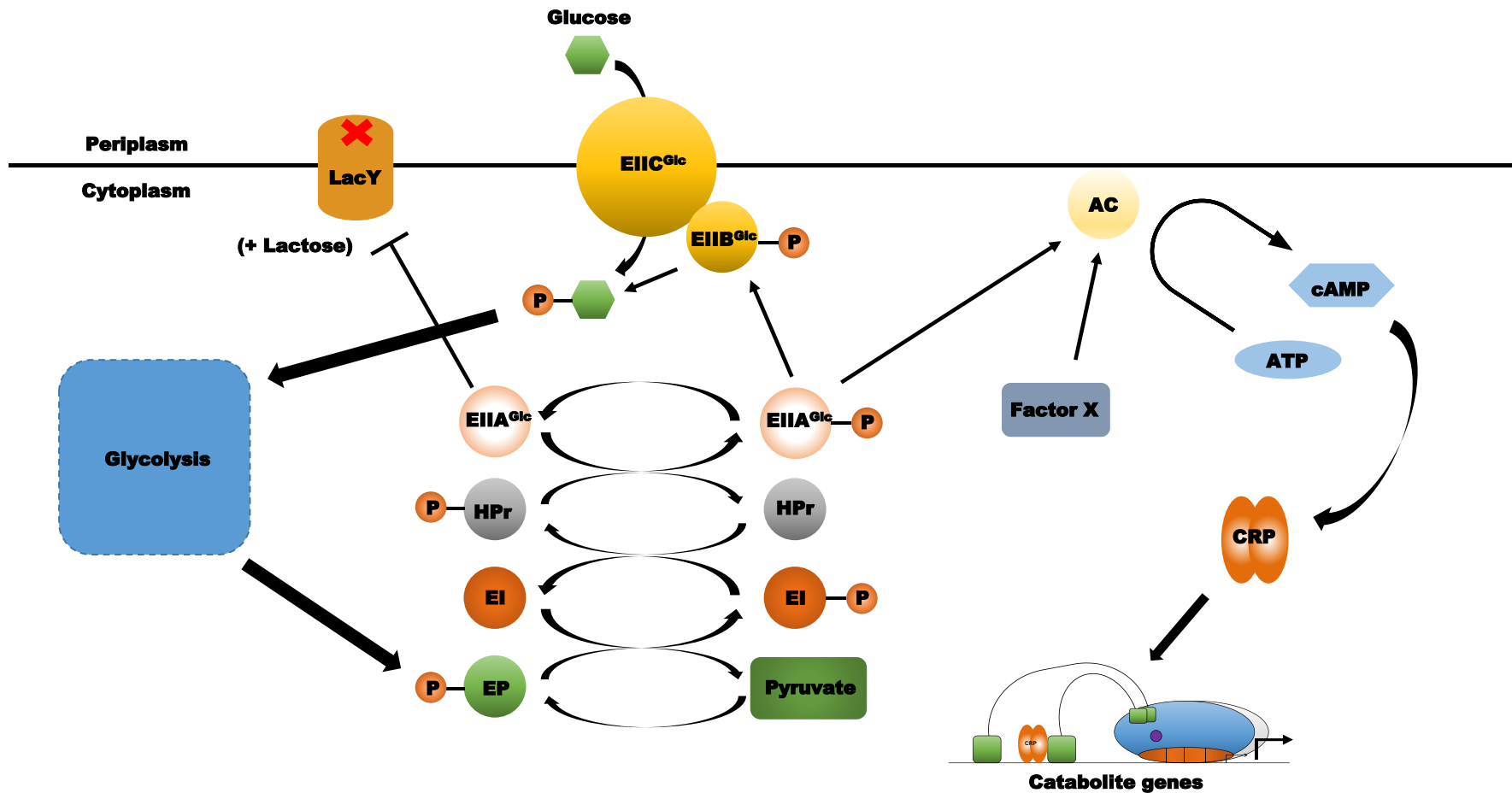
Figure 1.12 Carbon catabolite repression (CCR) in *E. coli*

Glucose enters the cell via the glucose specific EII^{Glc} complex in the inner membrane and is phosphorylated through the phosphotransferase system. The phosphorylation state of EIIA^{Glc} is a key factor in determining the phosphorelay, which is controlled by the ratio of phosphoenolpyruvate (PEP) to pyruvate. High concentrations of phosphorylated EIIA^{Glc} lead to activation of adenyl cyclase (AC), in the presence of an unknown factor (factor 'X'). HPr (Histidine Protein) and EIIA^{Glc} can interact with a number of additional transcription factors not shown here. Based on Görke and Stulke (2008).

In the presence of glucose, EIIA^{Glc} becomes unphosphorylated, and in this state is able to inhibit transporters for these inducers, for example, LacY, the lactose permease (Nelson *et al.,* 1983). This inhibition occurs only in the presence of lactose, which must be bound to the transporter (Hogema *et al.,* 1999). This phenomenon is known as "Inducer Exclusion", and prevents the import of many other secondary catabolites in the presence of glucose (Görke and Stűlke, 2008). cAMP-CRP plays an important role in this process by regulating the expression of the glucose PTS EIIBC complex encoded by the *ptsG* gene (Kimata *et al.*, 1997). The correlation between intracellular cAMP concentrations and glucose concentrations has been subject to controversy. The current prevailing opinion is that intracellular cAMP concentrations are markers of low energy levels in the cell. CRP (complexed with cAMP) activates catabolic genes and represses anabolic genes, hence increased cAMP concentrations lead to increased energy production through increased catabolism (Narang, 2009).

1.3.7 The CRP regulon

As discussed above, CRP (complexed with cAMP) facilitates the utilisation of alternative carbon sources by controlling gene expression (Busby and Ebright, 1999, Gosset *et al.,* 2004, Narang *et al.,* 2009). Thus, many studies have tried to identify the complete set of genes regulated by CRP in *E. coli*. Shimada *et al.* (2011) have provided the most comprehensive study of CRP-dependent promoters to date. This work shows that CRP regulates at least four irreversible, rate-limiting reactions in the glycolysis pathway by controlling expression of i) Glk (which phosphorylates glucose upon import into the cell) ii) YggF (a type II fructose 1,6-bisphosphatase) iii) Pps (phosphoenolpyruvate synthase) and iv) PckA (phophoenolpyruvate carboxykinase). The activation of PckA expression implicates CRP in regulating switching between glycolysis and gluconeogenesis. In addition, CRP also activates expression of several "last step" enzymes in the pentose phosphate pathway that shuttle D-glyceraldehyde-3-phosphate back into central metabolism. Shimada *et al.* also confirmed two other important

roles for CRP; the regulation of metabolic transporters and the regulation of "slave" TFs. Thus, CRP regulates the expression of 16 phosphotransferase systems, 11 major superfamily transporters, all 9 ABC transporters involved in carbohydrate import and 70 transcription factors involved in many aspects of metabolism (Shimada *et al.*, 2011).

The regulatory role of CRP is not limited to central metabolism. For instance, CRP affects the expression of the stationary phase σ-factor, σ^S (Hengge-Aronis, 2002). Motility is also controlled by CRP, in conjunction with H-NS; both proteins are required to activate transcription from the *flhDC* promoter (Soutourina *et al*, 1999). The chaperones GroEL/GroES are indirectly regulated by CRP as are DnaJ, DnaK and several proteases (including ClpPBX and Lon) (Gosset *et al*, 2004). Thus, the CRP regulon is extensive, and contains many indirect pathways.

1.4. Repression

i) Steric inhibition:

The simplest mechanism of repression is steric inhibition, whereby a repressor binds over promoter elements and prevents RNAP binding (Figure 1.13A). A good example of such regulation is the Lac repressor, LacI, at the *lac* operon. LacI binds to the $lacO_1$ operator site (at position +11) to repress transcription of the downstream genes *lacZ*, *lacY* and *lacA* (Sellitti *et al.,* 1987). Binding of LacI to two further operators, $lacO_2$ (centred at position +412) and $lacO_3$ (centred at position -82) further stimulates repression (Wilson *et al.,* 2007).

ii) DNA looping:

At the *gal* P1 promoter, GalR binds two operators at positions -60.5 (O_E) and +53.5 (O_I) relative to the transcription start site (Figure 1.13B). Interactions between the bound GalR molecules form a loop (Lewis and Adhya, 2002). Since GalR:GalR interactions are weak, an additional
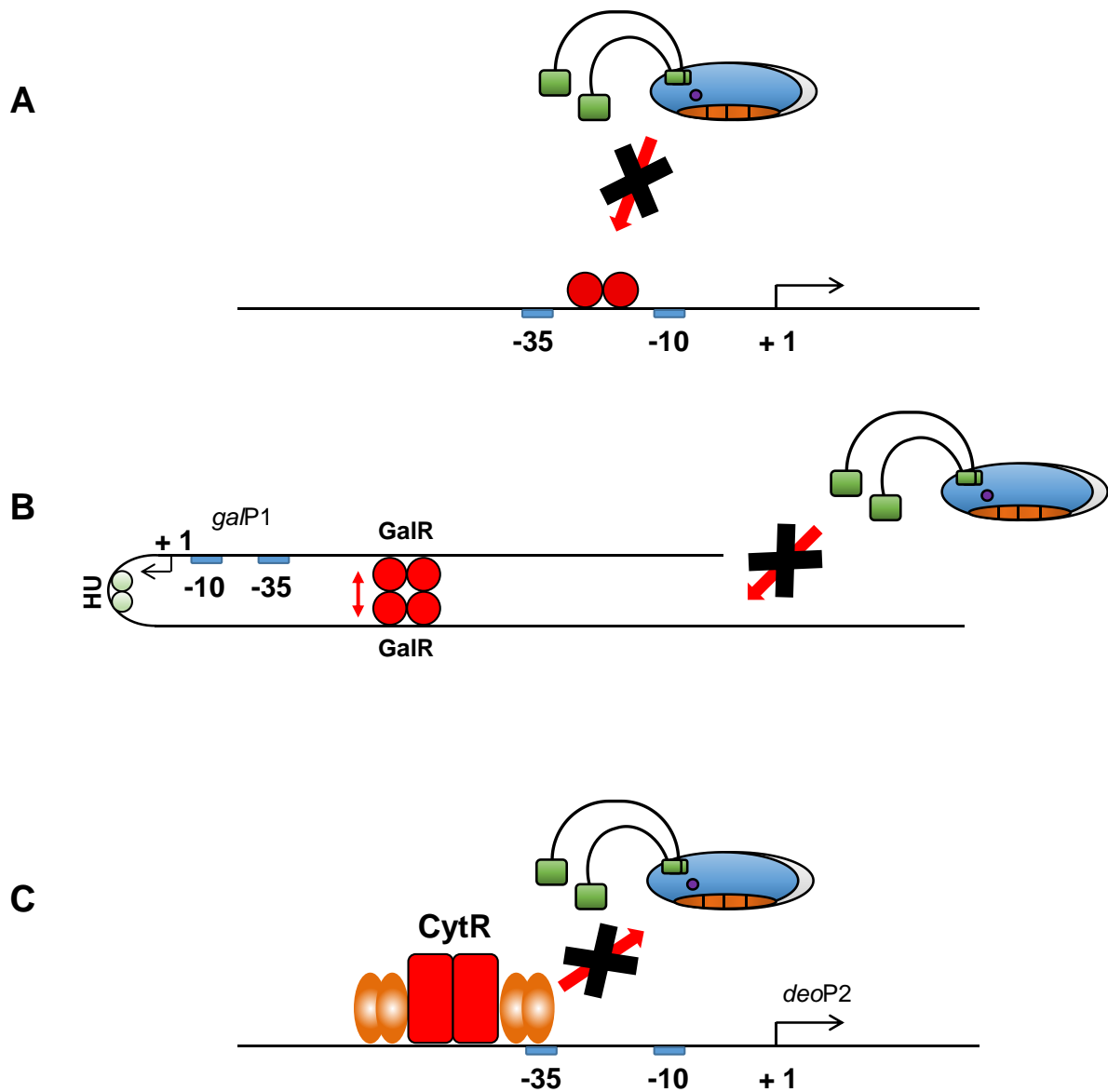
Figure 1.13 Mechanisms of transcriptional repression

Mechanisms of transcriptional repression by (A) steric hinderance, (B) DNA looping, and (C) 'anti-activation' by CytR at the *deo*P2 promoter. Red dimers represent repressors, light green circles represent additional factors, orange dimers represent activators, and DNA is shown as a black line. Based on Browning and Busby (2004), and Busby and Ebright (1999).

factor (HU), which binds at a position +6.5, is required to stabilise the loop. This facilitates DNA looping and prevents RNAP from accessing *gal*P1 (Choy and Adhya, 1992, Semsey *et al.*, 2002). DNA looping also occurs at the *lac* promoter; LacI dimers bound to operator regions *lacO₁* and *lacO₂*, or *lacO₁* and *lacO₃* form DNA loops. This looping mechanism represses the *lac* operon 600-fold as opposed to 20-fold repression due to steric hinderance by occupancy of *lacO₁* (Oehler *et al.*, 1990).

iii) Anti-activation:

Another mechanism of repression is anti-activation. A well-documented example is the interaction of the repressor CytR and CRP at promoters controlling genes involved pyrimidine metabolism. At the *deo*P2 promoter, CytR binds between two CRP dimers and denies the RNAP α-CTD access to the AR1 (Figure 1.13C) (Shin *et al.*, 2001).

1.5 Transcriptional regulation by Nucleoid-Associated Proteins (NAPs)

Most genes are also subject to regulation by one or more small basic proteins called nucleoid-associated proteins (NAPs). At least 12 have been identified in Gram negative bacteria. These include MvaT, EbfC, CbpA, CbpB, H-NS, StpA, Dps, IHF, Fis, MukB, Lrp and HU (Dorman and Dillon, 2010). The relative abundance of NAPs inside the cell alters in response to changes in growth rate and the environment. Hence, NAPs often regulate transcription in response to growth rate.

1.5.1 Fis

Fis (factor for inversion stimulation) is a protein involved in preparing cells for exponential growth. This is reflected in the abundance of Fis; it is present in exponentially growing cells at 60,000 copies per cell, but is not detectable in stationary phase cells (Ali Azam *et al.*, 1999). Globally, Fis has an important chromosome structuring role and stimulates DNA "branching", which refers to increased formation of apical loops (Schneider e*t al.,* 2001). However, Fis has

a specific regulatory role at some promoters. For example, Fis binds upstream of the promoter for the *rrnB* operon, and stimulates transcription initiation from the P1 promoter (Zhang and Bremer, 1996). Thus Fis is an important regulator of ribosome biogenesis.

1.5.2 H-NS

H-NS is found in most bacteria and binds AT-rich regions of the genome (Bertin *et al.,* 1999, Navarre *et al.*, 2006, Navarre *et al.,* 2007). H-NS (137 amino acids) contains two domains; the N-terminal domain is involved in dimerisation and multimerisation, the C-terminal domain binds DNA (Dorman, 2004). In *E. coli* H-NS, Ala-116, Thr-114, Arg-113, Glu-101 and Thr-109 are important DNA-binding determinants (Sette *et al.,* 2009). An H-NS-like protein, Lsr2, has been identified in *Mycobacterium tuberculosis* and other closely related Gram positive bacteria. Similarly Lsr2 targets AT-rich DNA, including some virulence genes, and recognises target DNA in the same way (Gordon *et al.,* 2010). Strikingly, *E. coli* H-NS and *Mycobacterium smegmatis* Lsr2 are only 20 % identical, but can functionally substitute for one another (Gordon *et al.* 2008).

Repression by H-NS is mediated by binding to a 'nucleation site', followed by multimerisation across adjacent DNA (Lang *et al.,* 2007, Rimsky *et al.,* 2001). Binding by H-NS can occur in two different ways; 'stiffening' and 'bridging', and the switch between these forms is mediated by divalent cations (Liu *et al.,* 2010). Thus, there are three models for transcriptional repression by H-NS i) occluding RNAP from promoters through filamentation, ii) forming DNA loops that trap RNAP iii) impeding RNAP elongation and hence promoting transcriptional termination (Dame *et al.,* 2002, Dole *et al.,* 2004, Lim *et al.,* 2012, Lucchini *et al.,* 2006, Peters *et al.,* 2012).

1.6 Defining regulons in *E. coli* K-12

A recent aim of many studies has been the identification of the complete set of binding sites for TFs, like CRP, and NAPs like H-NS. Strategies to identify sites for these factors are discussed below.

1.6.1 Bioinformatic prediction of transcript TFs binding sites

The first global studies of TF binding made use of newly available genome sequences and relied on *in silico* approaches (Robison *et al*, 1998, Tan *et al*, 2001). In particular, the application of Position Weight Matrices (PWM) played an important role. PWM are generated by aligning already characterised TFs binding sites. This allows a consensus sequence, describing the probability of a given nucleotide occurring at a given position in the binding site, to be generated. Using this approach, Robison *et al.* (1998) predicted sites for 55 transcription factors across the *E. coli* genome, including a total of 220 high-affinity CRP sites (81 % in non-coding regions), and 9097 low-affinity sites (34 % in non-coding regions). Tan *et al* (2001), used a similar approach, but combined datasets from *E. coli* and *Hemophilus influenzae* to improve the predictive power of the PWM. This study predicted 117 binding sites for CRP in *E. coli*, of which over 90 % were located upstream of genes. As shown by Robison *et al.,* Tan *et al.* also found weaker CRP sites within genes. Thus, bioinformatic approaches are useful for estimating the number and distribution of TF binding sites. However, there are limitations; all sites require experimental verification. For example, some sites may not be bound *in vivo* or may require cooperative binding with another TFs (Hollands *et al.,* 2007). Moreover, bioinformatic predictions can only be made if the TF binding site is already defined.

1.6.2 *In vivo* transcriptional profiling and ROMA

The development of DNA microarrays allowed the effects of TFs on global gene expression to be determined for the first time. Briefly, RNA is extracted from wildtype cells and cells lacking

the TF of interest (Figure 1.14A). Logically, some of the resulting changes in gene expression will be a consequence of removing the TF and, hence, TF regulated genes can be identified (Gosset *et al.,* 2004, Mockler and Ecker, 2005). However, this approach is problematic because it is impossible to separate direct regulatory events from changes in gene expression caused by i) altered growth rates and ii) downstream gene regulatory effects mediated by slave TFs. Many of these problems can be avoided by reconstituting the TF regulatory network *in vitro*. For example, Zheng *et al.* (2004) used ROMA (Run-Off transcription and Microarray Analysis) to define a CRP regulon. ROMA uses purified genomic DNA as a template for *in vitro* transcription, the products of which are subjected to reverse transcription and analysed using DNA microarrays (MacLellan *et al.,* 2009) (Figure 1.14B). Typically, two experiments are performed, one using RNAP in the absence of a TF, and another where the TF is present. This study identified 152 targets for CRP-dependent regulation. However, 17 operons known to be regulated by CRP were not identified by ROMA. This is expected because many promoters rely on interplay between several TFs. For instance, the *melAB* promoter requires binding of MelR to sites flanking CRP to activate transcription (Wade *et al*, 2001).

1.6.3 ChIP-chip and ChIP-seq

Chromatin immunoprecipitation (ChIP) coupled with microarray analysis (ChIP-chip) has provided a powerful tool for direct mapping of TF binding sites *in vivo*. ChIP involves rapid formaldehyde crosslinking of protein-DNA complexes in cells, followed by cell lysis and shearing of chromatin. Immunoprecipitation of protein-DNA complexes using antibodies isolates DNA bound by a specific TF. The precipitated DNA is labelled and hybridised to a microarray. This allows profiling of bound sites (Figure 1.15A). Often, this requires that DNA from a 'mock' immunoprecipitation is used as a control (Buck and Lieb, 2004). ChIP can also be coupled with next generation DNA-sequencing (NGS), in a technique known as ChIP-seq. ChIP-seq has several advantages over ChIP-chip, especially if microarrays are unavailable

A.
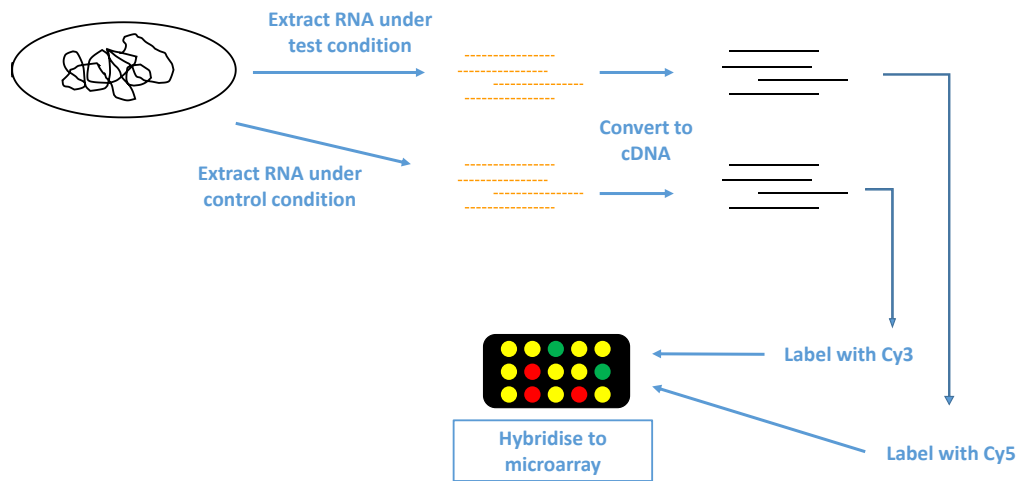
B.
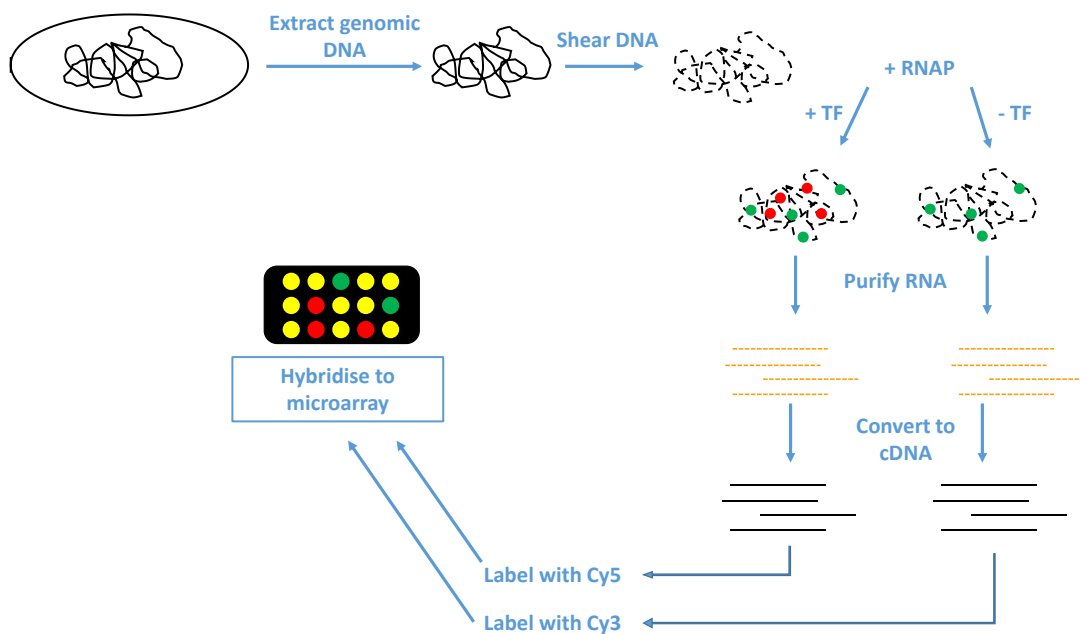
Figure 1.14 Global transcriptome analysis using *in vivo* profiling (A) and ROMA (B).

A. *In vivo* profiling: cells are grown under test and control conditions. RNA is extracted, converted to cDNA using reverse transcriptase. Test and control samples are labelled with Cy5 and Cy3 (respectively), and are hybridised to a microarray.

B. Run-off transcription analysis (ROMA): Genomic DNA (gDNA) is extracted from cells and is fragmented by either restriction enzymes or sonication. Purified RNA polymerase is added to purified gDNA *in vitro* in the presence or absence of the TF. RNA from *in vitro* transcription is purified, converted to cDNA. Samples are labelled with fluorescent dyes and hybridised to a microarray.

A

+ Formaldehyde
+Lysozyme

Sonication

De-crosslink and
label with Cy3

Labelling with Cy5

Hybridise to
microarray

Immunoprecipitation

Protein A

Elution and de-
crosslinking

Washing

Sepharose
bead

Antibody

B

Blunting
(T4 DNA polymerase/
Klenow fragment)

A' tail addition
(Klenow fragment exo⁻)

Ligation of barcoded
adaptors
(T4 DNA ligase)

Immunoprecipitation

Elution, de-crosslinking
size selection, purification
and sequencing

C

Figure 1.15 Identifying transcription factor binding sites *in vivo* using ChIP-chip (A)

and ChIP-seq (B), and *in vitro* using genomic SELEX (C)

A. ChIP-chip analysis. Following fixing, lysis and sonication, protein-DNA complexes are immunoprecipitated, eluted, and fluorescently labelled for hybridisation to a microarray. DNA is shown as black lines, coloured circles are DNA-binding proteins, red circles are DNA-binding proteins of interest. Based on Buck and Lieb (2004).
B. Illumina library preparation using ChIP-DNA, prepared as shown in (A). Black lines show DNA, red circles depict a DNA-binding protein.
C. Genomic SELEX, as developed by Shimada *et al.* (2005). gDNA is isolated from cells and is cloned into a cloning vector. Fragments are amplified through PCR, and are mixed with a purified, His-tagged TF. The mixture is passed through a Ni-NDTA column which binds the transcription factor. Following washing, DNA is eluted and is subject to further SELEX cycles, before sequencing or hybridisation to a DNA chip. Plasmid vectors are shown as red circles, black lines depict DNA, red circles are DNA-binding proteins. Based on Ishihama (2010).

for a particular organism. (Mardis, 2007, Park, 2009). Specifically, ChIP-seq can map DNA binding to a resolution of 20-30 bp depending on the size of sonicated DNA, whereas resolution from microarrays depends largely on microarray probe density (Mockler and Ecker, 2005).

The enhanced resolution afforded by NGS means that ChIP-seq is particularly useful for mapping protein binding sites in genomes with repetitive elements, which microarrays are unable to distinguish between (Park, 2009). The ChIP-seq workflow is initially similar to that of ChIP-chip. Cross-linked and sonicated DNA is immunoprecipiated. However, instead of labelling with fluorescent dyes, DNA fragments are used to prepare a library for NGS. For Illumina NGS platforms, this involves ligation of 'barcoded' DNA adapters to both ends of each DNA fragment (a workflow is shown in Figure 1.15B). One significant advantage of ChIP-seq is that multiple samples can be analysed together (multiplexing) if adapters with different 'barcodes' are chosen for each sample. (Mardis, 2007, Park, 2009).

In an early use of these approaches Grainger *et al.* (2005) exploited ChIP-chip to map CRP and RNAP binding across the *E. coli* K-12 MG1655 genome in the presence of IPTG, salicylic acid and rifampicin. This study identified 68 binding sites for CRP and 39 novel targets. Furthermore, this study provided evidence to support suggestions made by Robison *et al* (1998), which predicted thousands of weak sites for CRP binding across the K-12 genome. This hints at an ancillary ''NAP-like'' role for CRP, perhaps involving the DNA bend of 80° induced upon CRP binding to DNA (Schultz *et al*, 1991, Lawson *et al,* 2004).

1.6.4 Genomic SELEX

Genomic SELEX methodology generates comprehensive lists of TF binding sites (Shimada *et al.*, 2005, Ishihama, 2010). A simplified Genomic SELEX workflow is shown in Figure 1.15C. Genomic SELEX uses sonicated genomic DNA fragments, which are cloned into a plasmid vector, and amplified by PCR for each SELEX experiment. Once generated, this library is

repeatedly passed through a nickel-nitrilotriacetic acid (Ni-NDTA) column, which is pre-bound by a 6x His tagged TF. DNA fragments are then eluted from the column, and are either subject to another round of SELEX selection, or amplified by PCR, and cloned into a plasmid vector for SELEX-clos (cloning-sequencing) or SELEX-chip (microarray analysis) (Shimada *et al.*, 2005).

The study by Shimada *et al.* (2011), using both SELEX-clos and SELEX-chip, provided the most comprehensive analysis of genome wide CRP binding in *E. coli* K-12 and identified a total of 254 CRP targets.

1.7 Enterotoxigenic *E. coli* (ETEC)

The complex gene regulatory systems described above allow *E. coli* to compete with other microbes, and the host organism, for nutrients available in the mammalian gut. Whilst many *E. coli* strains live harmoniously with their host some *E. coli* strains are the cause of gastrointestinal disease. In countries with poor sanitation infrastructure Enterotoxigenic *E. coli* (ETEC) is a common causes of such disease. ETEC is an extracellular pathogen that mediates disease by secreting toxins and was isolated in 1971 from patients in Bangladesh thought to have Cholera (Sack *et al.,* 1971). ETEC is responsible for 200 million diarrhoeal episodes and 380,000 deaths, mostly in infants under 5 years of age, every year (Isidean *et al.,* 2011, Rivera *et al.,* 2013, Walker *et al.,* 2007).

1.7.1 An overview of ETEC virulence factors

ETEC that survive the low pH of the stomach express fimbrial adhesins and adhere to the brush border of the small intestine (Figure 1.16) (Porter *et al.*, 2011 Fleckenstein *et al.*, 2010). Once attached, ETEC secrete one or both of two toxins. The 84 kDa heat-labile toxin (LT) and the 2 kDa heat-stable toxin (ST). The LT toxin has homology to the cholera toxin (CTX) of *Vibrio*
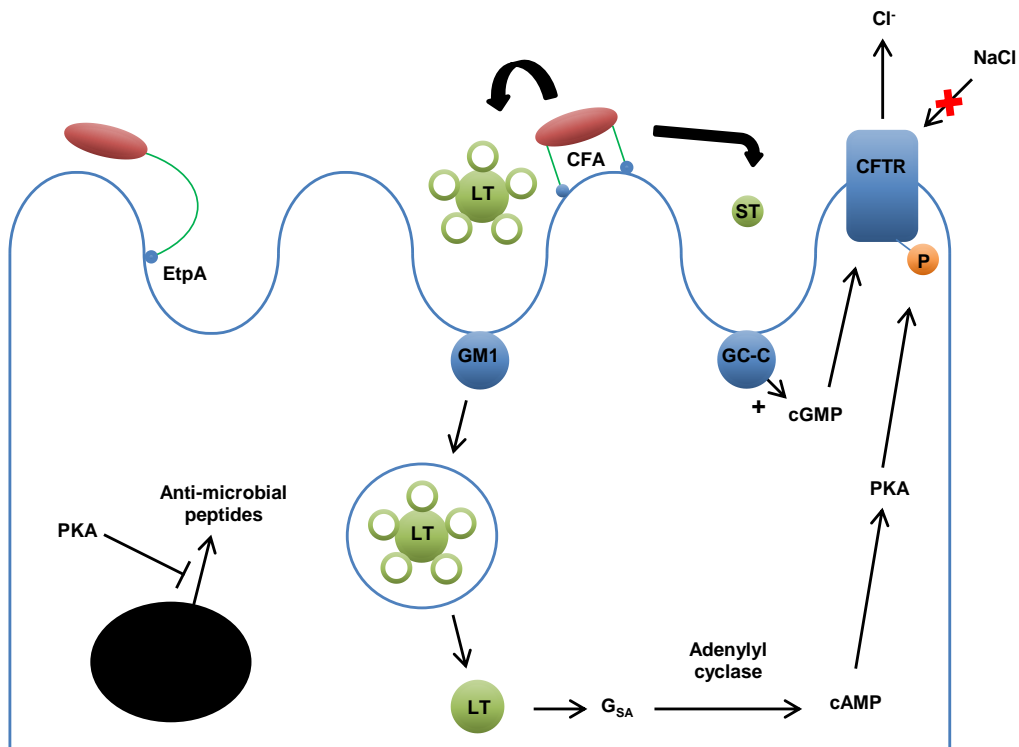
Figure 1.16 A model of ETEC infection

ETEC cells (depicted here in red) adhere to the brush border of the small intestine using a number of different adhesins- these include CFs (these vary widely), EtpA, and invasins such as TibA and Tia (not shown). Host proteins are shown in blue, ETEC proteins are shown in green. Either ST or LT (or both) are secreted and bind to their respective receptors (Guanylate cyclase C, and ganglioside GM1). PKA (Protein Kinase A) inhibits anti-microbial peptide production (here denoted as AMP). Based on Croxen & Finlay (2010).

cholerae (Gibert *et al.,* 1990). The ST toxin is a mimic of the human hormones uroguanylin and guanylin (Taxt *et al.,* 2010). Both LT and ST trigger the loss of sodium and chloride ions into the gut lumen, which causes watery diarrhoea, abdominal pain and vomiting. If untreated, these symptoms include coma and death (Nataro and Kaper, 1998).

1.7.2 Heat-labile toxins (LT)

LT is comprised of six subunits (five 'B' subunits and one 'A' subunit). The B subunits form a 'doughnut' structure around the A subunit. The A subunit consists of two regions, A1, which has ADP-ribosylating activity, and A2, which interacts with both the A1 and the B subunits (Sixma *et al.*, 1991). The A and B subunits are transported into the periplasm by the Sec pathway, and folded and assembled by DsbA, before being exported by the General Secretion Pathway (GSP) (Tauschek *et al.*, 2002, Yu *et al.*, 1992). Where present, a GTPase (encoded by the chromosomal *leoA* locus) may assist in secretion by interacting with GSP (Brown *et al*, 2007).

Once secreted, LT binds to the ganglioside GM1 receptor on host epithelial cells and is endocytosed. The A1 subunit of the toxin ADP-ribosylates host $G_{s\alpha,}$ (a trimeric G protein). This activates adenyl cyclase resulting in elevated cAMP levels, activation of host protein kinase A, and phosphorylation of the Cystic Fibrosis Transmembrane Regulator (CFTR) protein (Cheng *et al.,* 1991, Hamilton *et al.,* 1978, Moss and Richardson, 1978). The consequence of this cascade is loss of chloride ions into the intestinal lumen and inhibition of sodium and chloride ion absorption (Spangler, 1992). The LT toxin may also contribute to adherence and colonisation (Johnson *et al.,* 2009).

1.7.3 Heat-stable toxin

The heat-stable toxins (ST) are a group of non-immunogenic cysteine-rich peptides that remain functional after heating to 100 ºC (Evans *et al.*, 1973). They have significant homology to

human hormones uroguanylin, and guanylin (Taxt *et al.,* 2010). There are two variants of ST secreted by ETEC to induce diarrhoea; STa1 and STa2. These are closely related and function through identical mechanisms. The 18 amino acid STa1 peptide is distributed amongst ETEC isolates from animals and humans whilst the 19 residue STa2 variant is specific to humans. STa1 and STa2 share 14 amino acids and are both methanol soluble and protease resistant (Moseley *et al.,* 1983, Steinsland *et al.,* 2002, Taxt *et al.,* 2010). Expression of ST alone, without LT expression, is sufficient to cause diarrhoea (Levine *et al*, 1977).

Both STa1 and STa2 are expressed as ~70 amino acid pre-pro-peptides that are cleaved on export via the Sec system (Okamoto *et al*, 1995). Once in the periplasm, the disulphide oxidoreductase enzyme (DsbA) mediates disulphide bond formation between three pairs of cysteine residues (Okamoto *et al.*, 1995, Yamanaka *et al.*, 1994). The toxin is then secreted across the outer membrane through TolC (Yamanaka *et al.*, 1998). ST toxins bind mammalian membrane-bound guanyl cyclase C (GC-C), resulting in cGMP accumulation, and phosphorylation of CFTR by the cAMP-dependent protein kinase. Thus, both LT and ST drive release of chloride ions into the gut lumen by the CFTR, resulting in water loss and diarrhoea (Chao *et al.*, 1994, Newsome *et al.*, 1978). There is some evidence that STa forms 35Å hexameric rings to interact with GC-C (Sato and Shimonishi, 2003).

1.7.4 Colonisation Factors (CFs)

In order to attach to the intestinal brush border ETEC cells express colonisation factors (CFs). These interact with epithelial cell glyco-conjugates (Gaastra and Svennerholm, 1996). CFs are often plasmid-encoded, and the expression of CFs is associated with diarrhoeal disease (Evans *et al.,* 1975). There are at least 25 serotypically different CFs and many others are under investigation (Table 1.2) (Gaastra and Svennerholm, 1996, Qadri *et al.,* 2005). The CFs range in size from 15 to 25 kDa and most are fimbrial. In the prototypical H10407 strain, the major

ETEC Colonisation Factors	Nature of adhesin
CFA/I	Fimbrial
CS1	Fimbrial
CS2	Fimbrial
CS3	Fibrillae
CS4	Fimbrial
CS5	Helical
CS6	Non-Fimbrial
CS7	Helical
CS8	Fimbrial
CS10	Non-Fimbrial
CS11	Fibrillae
CS12	Fimbrial
CS13	Fibrillae
CS14	Fimbrial
CS15	Non-Fimbrial
CS17	Fimbrial
CS18	Fimbrial
CS19	Fimbrial
CS20	Fimbrial
CS21	Fimbrial
CS22	Fibrillae

Table 1.2 Colonisation Factors of ETEC

Based on Gaastra and Svennerholm, 1996, and Qadri *et al.*, 2005.

colonisation factor (CFA/I) was characterised in 1979 (Evans *et al.*, 1979). CFA/I forms a 7 nm fimbriae on the cell surface and is the archetypal member of class V pili, of which there are 8 expressed by different ETEC strains (Anantha *et al.*, 2004). Class V pili are assembled by the alternate chaperone pathway, which is similar but distinct from the chaperone-usher pathway involved in assembly of *pap* (P) and *fim* (Type I) pili. Assembly by the alternate chaperone pathway is simple and requires only the Sec system and the four genes of the fimbrial operon (Anantha *et al.*, 2004, Soto and Hultgren, 1999). In ETEC H10407 the CFA/I fimbrial apparatus is encoded by the *cfaBCE* operon. CfaA, B, C and E are the chaperone, major pilin subunit, outer membrane usher, and tip protein respectively (Jordi *et al.*, 1992). The transcription factor CfaD, controls expression of the operon (Savelkoul *et al.*, 1990).

1.7.5 Ancillary colonisation factors

In addition to the 'classical' virulence factors described above, a number of minor virulence factors have recently been reported. In particular, the discovery that ETEC can invade some epithelial cell lines has led to the characterisation of additional afimbrial adhesins (Elsinghorst and Weitz, 1994, Fleckenstein *et al.*, 1996).

1.7.5.1 The *tib* and *tia* loci

TibA is a 104 kDa outer membrane glycoprotein, expressed in 16 % of ETEC strains. The protein is encoded by the *tibABCD* operon (Elsinghorst and Weitz, 1994, Turner *et al.*, 2006). TibA is an autotransporter, which, when localised to the outer membrane, allows recombinant K-12 strains to adhere to and invade epithelial cells (Elsinghorst and Weitz, 1994, Lindenthal and Elsinghorst, 1999, Lindenthal and Elsinghorst, 2001). The *tia* locus was characterised at a similar time, and encodes Tia; a 25 kDa outer membrane invasin that is inserted into the bacterial membrane by 8 transmembrane domains (Fleckenstein *et al.*, 1996, Mammarappallil and Elsinghorst, 2000). The *tia* locus was found in 5 % of ETEC strains and also encodes an

invasin (Fleckenstein *et al.,* 1996, Turner *et al.,* 2006). Deleting *tia* leads to a 75 % reduction in invasion compared to H10407 strains expressing both Tia and TibA. Deleting *tib* reduces invasion by 85 % and deleting both invasins reduces invasion by 90 % (Elsinghorst and Weitz, 1994, Mammarappallil and Elsinghorst, 2000,). Despite this, the role and contribution of the *tib* and *tia* loci to invasion is unclear.

1.7.5.2 The Etp system

The Etp system (encoded by *etpBAC*) is another accessory virulence determinant in ETEC. EtpA is a large glycoprotein of 170 kDa. EtpB forms a β-barrel involved in transport and afimbrial adhesion. EtpC is a transglycosylase. Recombinant *E. coli* K-12 strains expressing *etpBAC* adhere to epithelial cells (Fleckenstein *et al.*, 2006). EtpA mediates this adhesion via flagella, acting as an adaptor (Roy *et al.*, 2009). EtpA has been shown to play an important role in efficient heat-labile toxin delivery (Roy *et al.*, 2012).

1.7.5.3 CexE and the Aat system

The CexE protein is well conserved between ETEC strains and may possess "dispersin-like" activity. Dispersin is a hydrophilic factor that coats the surface of some *E. coli* strains to enhance dispersal of cells by preventing self-aggregation mediated by hydrophobic fimbrial adhesins. In Enteroaggregative *E. coli* (EAEC) dispersin is secreted by the Aat type I secretion system that is also found in ETEC (Crossman *et al.*, 2010, Nishi *et al.*, 2003, Shiekh *et al.*, 2002). The *cexE* gene is genomically associated with the *aat* operon in H10407, hence, it seems likely that CexE is exported by the Aat system in ETEC (Crossman *et al.,* 2010). A model for the organisation of Aat system proposed by Nishi *et al.* is shown in Figure 1.17. In this model AatA forms an outer membrane channel that docks with the ATP-binding AatC subunit and the permease AatP. AatB and AatD may assist in this interaction and influence signal transduction.
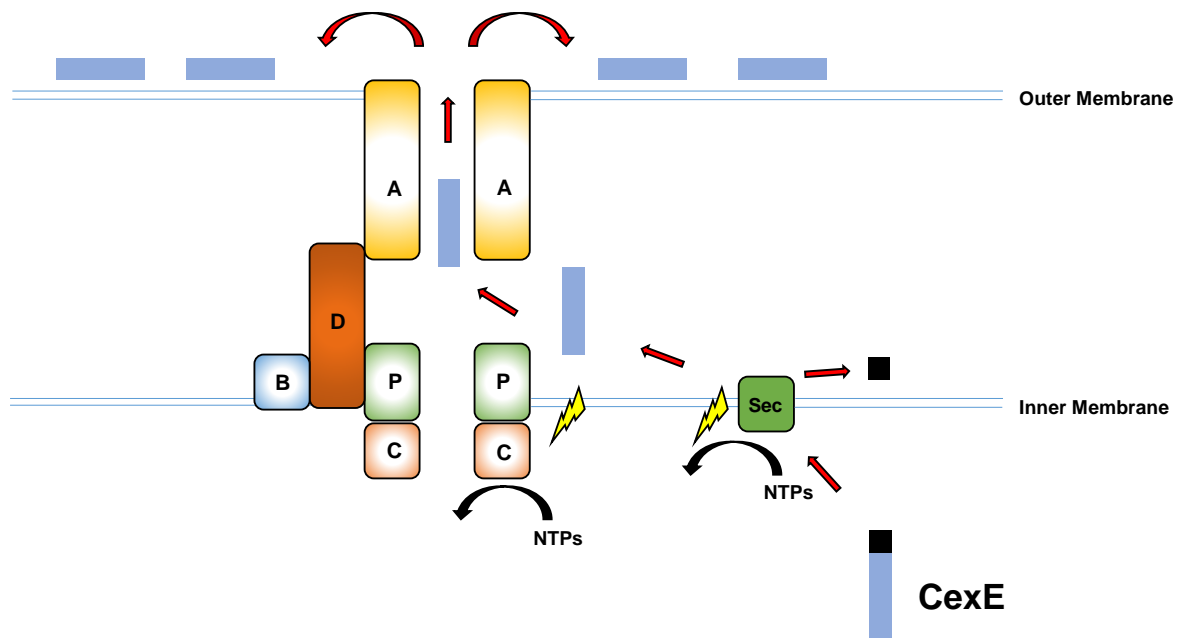
Figure 1.17 Predicted function of the ETEC Aat system based on Aat system of EAEC

Yellow bolts and black arrows highlight NTP hydrolysis. Red arrows show the proposed secretion mechanism of CexE. The black box represents a signal peptide on CexE. Based on Nishi *et al.* (2003).

Dispersin, (or CexE in this case) is transported across the inner membrane by the Sec machinery, and is exported across the outer membrane through AatA.

1.8 ETEC genomics and phylogeny

There are five phylogenetic groups of *E. coli* (A, B1, B2 D and E) (Chaudhuri and Henderson, 2012). Turner *et al.* (2006) identified ETEC strains belonging to all of these lineages and no correlation between phylogeny and preferred host. The best characterised ETEC is the prototype strain H10407, a member of the group A phylogeny. The ETEC H10407 genome contains a 5,153,435 bp chromosome, two large plasmids (p948 and p666) and two smaller plasmids (p58 and p52). Most virulence factors (including LT, ST, CFA/I and the Aat system (Figure 1.18)) are encoded by p948 and p666. The chromosome of H10407 contains an additional 755,359 bp of DNA compared to *E. coli* K-12, localised to 25 regions and 599 genes. Surprisingly, only 8 % of these genes, including the primary virulence factors, are found in other ETEC strains (Crossman *et al.,* 2010).

1.9 Transcriptional control of virulence genes in ETEC

Current evidence suggests that regulation of virulence factor expression in ETEC is mediated by a combination of ETEC specific regulators (such as Rns and CfaD) and TFs found in all *E. coli* strains (such as CRP and H-NS). Rns is an AraC family TF, 32 kDa in size, which is widespread in ETEC isolates (Caron *et al*., 1989, Caron *et al.,* 1990, Munson *et al.*, 2002). At the P*coo* promoter (which controls expression of CS1 fimbriae), Rns binds at positions -109.5 and -37.5 relative to the transcription start site and activates transcription. However, despite also regulating the transcription of other virulence factors the Rns regulon in poorly defined (Munson and Scott, 1999). CfaD activates transcription at the *cfa* operon (Jordi *et al.*, 1992) in addition to overcoming repression mediated by the nucleoid associated protein H-NS (Jordi *et al.,* 1994, Savelkoul *et al.*, 1990). Glucose represses expression of CFA/I fimbriae (Karjalainen
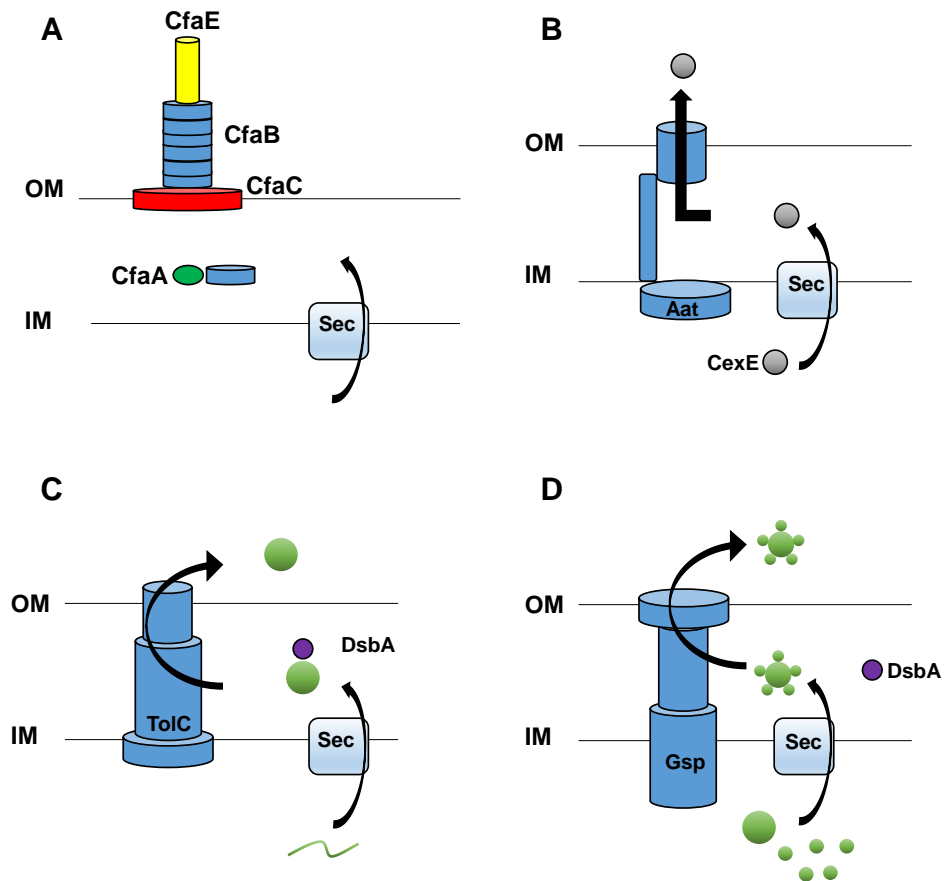
Figure 1.18 Assembly and secretion ETEC virulence factors

The four panels show the various assembly and secretion mechanisms used by different ETEC virulence factors. All four pathways require the Sec YEG complex located in the inner membrane for translocation of unfolded substrates. A) Assembly of CFA/I pili requires four genes encoded by the *cfaABCE* operon, and occurs through the process donor strand exchange. B) Secretion of CexE (a putative dispersin) requires export through the inner membrane by Sec, followed by export through the Aat system. Both the ST (C) and LT (D) toxins are assembled in the periplasm by DsbA. ST is secreted through TolC, and LT is secreted through the general secretory pathway type II secretion system. Based on Turner *et al.* (2006).

et al., 1991) and the ST toxin (Alderete and Robertson, 1977, Busque *et al.*, 1995) possibly implicating CRP in their regulation (Saier, 1998). Moreover, CRP is proposed to directly repress transcription of the *eltAB* gene, encoding the heat-labile toxin, by binding to three sites upstream of the promoter; a high affinity site centred at -31.5, and two lower affinity sites centred at -132 and -261 upstream of the promoter (Bodero and Munson, 2009).

1.10 Overview and aims of this study

The reports discussed above hint at a role for CRP in the regulation of ETEC virulence factors, particularly since CRP plays an important in implementing CCR, and hence glucose repression. Sugar metabolism is known to be an important determinant of intestinal colonisation by *E. coli* MG1655, and thus it is perhaps not surprising that CRP is co-opted as a virulence regulator in many enterobacteria (Chang *et al.,* 2004, Le Bouguénec and Schouler, 2011). It seems plausible that CRP is likely to play a role in the regulation of ETEC given the core genome shared by ETEC and laboratory K-12 strains (Crossman *et al.,* 2010). To investigate this, we choose to investigate the regulation of virulence factor expression in the prototype strain H10407, which is known to cause particularly severe disease (Porter *et al.,* 2011, Turner *et al.,* 2006).

In this work, we have elucidated the CRP, H-NS and σ^{70} regulons of the prototype ETEC strain H10407, using bioinformatics and ChIP-seq. In subsequent genetic and biochemical analysis of individual targets we show that i) regulation of ST toxin production in ETEC H10407 is dependent on CRP ii) the NAP H-NS represses expression of ST iii) repression of LT expression by CRP occurs via an indirect mechanism and iv) that CRP activates transcription of an open reading frame within the *aatC* coding sequence.

Chapter 2

Methods and Materials

2.1 Buffers and generic reagents

All solutions were made up in deionised and distilled water (ddH$_2$O) unless stated. Solutions were autoclaved at 121 ºC for 15 minutes at 5 psi before use. All reagents (e.g. enzymes etc.) were purchased from Sigma or Life Technologies unless stated otherwise.

Phenol/chloroform extraction:

- **Phenol/chloroform/isoamyl alcohol pH 8 (25:24:1).**

Ethanol precipitation:

- **100 % (*v/v*) Ethanol.**
- **70 % (*v/v*) Ethanol.**
- **3 M Sodium acetate pH.5.2**.
- **20 mg/ml Glycogen.**

Preparation of competent cells:

- **100 mM Calcium chloride.**
- **50 % (*v/v*) Glycerol.**

Polyacrylamide gel electrophoresis (PAGE) and denaturing sequencing gel reagents:

- **5x TBE**: 0.445 M Tris borate pH 8.3, 10 mM Na$_2$EDTA. Supplied by Fisher Scientific, and diluted to a 1x stock using ddH$_2$O before use.
- **Ammonium persulphate**. Supplied by Sigma.
- **TEMED (N,N,N′,N′-Tetramethylethylenediamine)**. Supplied by Sigma.
- **30 % (*w/v*) Acrylamide: 0.8 % (*w/v*) bisacrylamide mix**. Supplied by Geneflow.
- **Sequagel ureagel system.** Supplied by Geneflow.

Agarose gel electrophoresis:

- **6x Loading dye**: 10 mM Tris pH 7.5, 1 mM EDTA, 20 % (*v/v*) glycerol, 0.025 % (*v/v*) bromophenol blue, 0.025 % (*v/v*) xylene cyanol FF.

- **5x TBE**: 0.445 M Tris borate pH 8.3, 10 mM Na_2EDTA. Supplied by Fisher Scientific. The 5x solution is diluted to a 1x stock using ddH_2O before use.

- **Powdered agarose**: supplied by Bioline.

β-galactosidase assays:

- **13 mM (8 mg/ml) 2-Nitrophenyl β-D-galactopyranoside (ONPG)**: made up to 1 L in 'Z-buffer'.

- **'Z-buffer'**: 8.53 g Na_2HPO_4, 4.87 g $NaH_2PO_4.2H_2O$ 0.75 g KCl, 0.25 g $MgSO_4$ per L of ddH_2O, then autoclaved.

- **β-mercaptoethanol**: immediately before use, 271 μl β-mercaptoethanol was added per 100 ml of 'Z-buffer'.

- **1 % (*w/v*) Sodium deoxycholate**.

- **100 % (*v/v*) Toluene**.

- **1 M Sodium carbonate**.

CRP purification:

- **1 M Potassium phosphate (pH 7.5)**: 162.75 ml 1 M K_2HPO_4, 37.26 ml KH_2PO_4.

- **1 M Sodium phosphate (pH 6.8)**: 46.3 ml 1 M Na_2HPO_4, 53.7 ml KH_2PO_4.

- **Lysis buffer**: 50 mM potassium phosphate (pH 7.5), 2 mM EDTA, 0.2 M NaCl, 5 % (*w/v*) glycerol. DTT (2 mM) was added after autoclaving.

- **Dialysis buffer**: 50 mM potassium phosphate (pH 7.5), 2 mM EDTA, 5 % (*w/v*) glycerol. DTT (2 mM) was added after autoclaving.

- **Column wash buffer:** 0.5 M potassium phosphate (pH 7.5), 2 mM EDTA, 5 % (*w/v*) glycerol, DTT (2 mM) was added after autoclaving.

- **Column wash buffer with 1 M NaCl:** 0.5 M potassium phosphate (pH 7.5), 2 mM EDTA, 1 M NaCl, 5 % (*w/v*) glycerol, 2 mM DTT (2 mM) added after autoclaving.

- **CRP stock buffer:** 10 mM sodium phosphate (pH 6.8), 0.1 mM EDTA, 0.2 M NaCl, 50 % (*w/v*) glycerol.

- **Superbroth (per litre):** 20 g tryptone, 5 g yeast extract, 0.5 g NaCl, 0.19 g KCl.

- **2x SDS-PAGE loading dye.** Purchased from Life Technologies.

- **Pre-cast NuPage 4-12 % (*w/v*) Bis-Tris polyacrylamide protein gels.** Purchased from Life Technologies.

- **1x MES buffer:** 50 mM MES (2-[N-morpholino]ethanesulphonic acid), 50 mM Tris base, 0.1 % (*w/v*) SDS, 1 mM EDTA, pH 7.3. Diluted from 20x MES (purchased from Life Technologies) using ddH$_2$O.

In vitro transcription:

- **10x Transcription buffer (TNSC buffer)**: 400 mM Tris acetate pH 7.9, 10 mM MgCl$_2$, 1 M KCl, 10 mM DTT.

- **STOP solution**: 97.5 % (*v/v*) deionised formamide, 10 mM EDTA, 0.3 % (*v/v*) Bromophenol Blue/ Xylene Cyanol FF.

- **P^{32} labelled α-UTP (3000 Ci/mmol)**. Supplied by Perkin-Elmer.

Radiolabelling of DNA fragments:

- **G-50 sephadex beads**. Resuspended in a 12 % (*v/v*) slurry with TE.

- **T4 polynucleotide kinase**. Supplied by NEB, used with supplied 10x buffer.

- **P^{32} labelled γ-dATP (7000 Ci/mmol)**. Supplied by MP Biomedicals.

- **TE (Tris-EDTA) buffer**: 10 mM Tris-HCl, 1 mM EDTA (pH 8.0).

Electrophoretic mobility shift assays:

- **10x Transcription buffer (TNSC buffer)**: 400 mM Tris acetate pH 7.9, 10 mM MgCl$_2$, 1 M KCl, 10 mM DTT.

M13 sequencing reactions:

- **2 M Sodium hydroxide**

- **Annealing buffer**: 1 M Tris-HCl pH 7.5, 100 mM MgCl$_2$, 160 mM DTT.

- **'A' mix short:** 840 μM of dCTP, dGTP, and dTTP each, 14 μM ddATP, 93.5μM dATP, 40 mM Tris-HCl pH7.5, 50 mM NaCl.

- **T7 polymerase dilution buffer:** 25 mM Tris-HCl pH 7.5, 5 mM DTT, 100 μg/ml BSA, 5% (*v/v*) glycerol.

- **Label mix 'A':** 1.375 mM of dCTP, dGTP, and dTTP each, 333.5 mM NaCl.

- **STOP solution:** 0.025 % (*v/v*) bromophenol blue, 0.025 % (*v/v*) xylene cyanol FF, 10 mM EDTA pH 7.5, 97.5% (*v/v*) formamide.

- **P^{32} labelled α-dATP (3000 Ci/mmol)**. Supplied by Perkin-Elmer.

Primer extension:

All solutions DEPC treated and autoclaved (except 100 % and 70 % (*v/v*) ethanol).

- **1x Hybridisation buffer** (20 mM HEPES, 0.4 M NaCl, 80 % (*v/v*) formamide).

- **100 % (*v/v*) Ethanol.**

- **70% (*v/v*) Ethanol** (made up with DEPC-treated ddH$_2$O).

- **3 M Sodium acetate pH 5.2**

- **3 M Sodium acetate pH 7**

G+A ladder generation:

- **DNAse I blue:** 5 M urea, 20 mM NaOH, 1 mM EDTA, 0.025 % (*v/v*) bromophenol blue, 0.025 % (*v/v*) xylene cyanol FF.

- **10 M Piperidine.** Diluted with ddH$_2$O before use.

- **100% (v/v) Formic acid.**

DNAse I footprints:

- **Recombinant DNAse I**. Supplied by Roche.

- **DNAse I blue:** 5 M urea, 20 mM NaOH, 1 mM EDTA, 0.025 % (*v/v*) bromophenol blue, 0.025 % (*v/v*) xylene cyanol FF.

- **10x Transcription buffer (TNSC buffer)**: 400 mM Tris acetate pH 7.9, 10 mM MgCl$_2$, 1 M KCl, 10 mM DTT.

- **DNAse I STOP solution**: 0.3 M sodium acetate, 10 mM EDTA.

TCA precipitation:

- **100 % (*w/v*) Trichloroacetic acid.**

- **100 % (*v/v*) Methanol.**

- **2x SDS-PAGE loading dye.** Purchased from Life Technologies.

- **1 x PBS**: 137 mM NaCl, 10 mM Na$_2$HPO$_4$, 2.7 mM KCl, 1.8 mM KHPO$_4$. pH 7.3. Tablets purchased from Oxoid.

- **Pre-cast NuPage 4-12 % (*w/v*) Bis-Tris polyacrylamide protein gels.** Purchased from Life Technologies.

- **1x MES buffer:** 50 mM MES (2-[N-morpholino]ethanesulphonic acid), 50 mM Tis base, 0.1 % (*w/v*) SDS, 1 mM EDTA, pH 7.3. Diluted from 20x MES (purchased from Life Technologies) using ddH$_2$O.

Silver staining of protein gels:

- **SilverQuest kit.** Purchased from Life Technologies.

ChIP-seq reagents:

- **1x TBS** (20 mM Tris-HCl, pH 7.4, 0.9 % (*w/v*) NaCl).

- **FA lysis buffer "150mM NaCl"** (50 mM Hepes-KOH, pH 7, 150 mM NaCl, 1 mM EDTA, 1 % (*w/v*) Triton X-100, 0.1 % (*w/v*) Sodium Deoxycholate, 0.1 % (*w/v*) SDS).

- **FA lysis buffer "500mM NaCl"** (50 mM Hepes-KOH, pH 7, 500 mM NaCl, 1 mM EDTA, 1 % (*w/v*) Triton X-100, 0.1 % (*w/v*) Sodium Deoxycholate, 0.1 % (*w/v*) SDS).

- **ChIP wash buffer** (10 mM Tris-HCl, pH 8, 250 mM LiCl, 1 mM EDTA, 0.5 % (*w/v*) Nonidet-P40, 0.5 % Sodium Deoxycholate).

- **ChIP elution buffer** (50 mM Tris-HCl, pH 7.5, 10 mM EDTA, 1 % (*w/v*) SDS).

- **1x TE buffer** (10 mM Tris-HCl, pH 8.0, 1 mM EDTA).

- **NEXTflex ChIP-seq Barcodes**-24 (supplied by BioOscientific).

- **Agencourt AMPure magnetic beads** (supplied by Beckman Coulter) XP DNA clean-up kit.

- **Quick blunting and quick ligation kits** (supplied by NEB).

- **Klenow fragment** (3'→5' exo-) (supplied NEB).

- **Monoclonal anti-FLAG, anti-CRP, anti-β, and anti-σ^{70} antibodies** (supplied by Neoclone)**.**

- **Monoclonal anti-H-NS antibody** (kindly donated by Jay Hinton).

- **Qubit fluorometer/ assay kit** (supplied by Life Technologies).

- **10 mM Tris-HCl pH 7.5/8**.

- **Protein A sepharose beads** (supplied by GE Healthcare, GE17-0780-01). Washed with ddH$_2$O and stored in a 50 % (*v/v*) slurry with 1x TBS.

- **Spin-X** columns.

- **Acrylamide gel extraction buffer** (0.3 M NaCl, 10 mM Tris-HCl, pH 8, 1 mM EDTA).

2.2 Solid growth media

LB agar:

LB agar was purchased from Sigma. Thirty-five g of powdered agar was dissolved in 1 litre of ddH_2O, before autoclaving.

MacConkey lactose agar:

MacConkey agar containing lactose (10 g/l) was purchased from Oxoid. Fifty-two g of powdered agar was dissolved in 1 litre of ddH_2O, before autoclaving.

2.3 Liquid growth media

LB broth:

LB broth was purchased from Sigma. Twenty g of powdered LB broth was dissolved in 1 litre of ddH_2O before autoclaving.

10x M9 salts:

Sixty g Na_2HPO_4, 30 g KH_2PO_4, 5 g NaCl, 10 g NH_4Cl, 10 mg biotin and 10 mg thiamine were dissolved in 1 litre of ddH_2O, before the pH was adjusted to pH 7.4, followed by autoclaving.

M9 minimal media:

M9 minimal medium was made up on the day of use. To make 100 ml of this media, the following solutions were autoclaved individually and mixed in the following ratio; 10 ml 10x M9 salts, 200 μl 1 M $MgSO_4$, 100 μl 0.1 M $CaCl_2$, 1.5 ml 20 % (*w/v*) fructose, 0.5 ml 20 %

(*w/v*) casamino acids, made up to 100 ml with autoclaved ddH$_2$O. If required, 100 μg/ml thymine was added to the media.

2.4 Antibiotics

Antibiotic stocks were made up and used as follows:

- Ampicillin: 100 mg/ml (dissolved in water) stored at -20 °C. The final concentration used in agar and broth unless stated otherwise was 100 μg/ml.

- Kanamycin: 50 mg/ml (dissolved in water) stored at -20 °C. The final concentration used in agar and broth unless stated otherwise was 100 μg/ml.

- Tetracycline: 35 mg/ml (dissolved in methanol) stored at -20 °C. The final concentration used in agar and broth unless stated otherwise was 35 μg/ml.

- Trimethoprim: used at 20 μg/ml.

Antibiotics were added to media at the stated concentrations, once media had cooled to below 50 °C after autoclaving.

2.5 Preparation of competent cells

Overnight culture (1 ml) was used to inoculate 50 ml of fresh LB medium. This was incubated at 37 °C until the OD$_{650}$ of the culture was between 0.3-0.6 (measured using a Jenway 6300 spectrophotometer). The culture was then decanted into centrifuge tubes, chilled on ice for 10 minutes, and cells harvested by centrifuging at 1,600 xg for 5 minutes at 4 °C. Cells were then re-suspended in 25 ml 100 mM ice-cold sterile CaCl$_2$, incubated on ice for 20 minutes, re-harvested and suspended in 3.3 ml ice-cold 100 mM calcium chloride. The suspension was left overnight on ice to maximise competency before 1 ml was aliquoted into microfuge tubes with 333 μl of ice-cold 50 % (*v/v*) glycerol for storage at -80 °C. Cells were thawed on ice before use.

2.6 Transformation of bacterial cells

One to five µl of plasmid DNA (or a ligation reaction) was incubated on ice with 100 µl of competent cells for 1 hour. Cells were then heat shocked at 42 °C for 2 minutes, before being returned to ice briefly. Five hundred µl of LB medium was added to the cells which were incubated at 37 °C for at least 45 minutes to recover. After this period, cells were pelleted by centrifugation at 2,400 xg for 2 minutes. Cells were re-suspended in 100 µl of LB medium and plated out on an appropriate medium.

2.7 Phenol-chloroform extraction/ ethanol precipitation of DNA

An equal volume of the phenol-chloroform mix was added to the DNA sample to be extracted. This was mixed vigorously for 15 seconds before centrifugation at 2,400 xg for 3 minutes. The upper aqueous phase was transferred to a new microfuge tube, and 2.5-3 volumes of ice-cold 100 % ethanol, along with 0.1 volumes of 3 M sodium acetate pH 5.2 was added. If the size of the DNA was <100 bp 1 µl 20 mg/ml glycogen was also added. DNA was precipitated at -80 °C for 30 minutes and collected by centrifugation at 17,000 xg and the pellet was washed in ice-cold 70 % (*v/v*) ethanol, dried under vacuum, and resuspended in the required volume of dH$_2$O.

2.8 Agencourt AMPure XP magnetic bead clean up

For ChIP-seq experiments, DNA purification was performed using Agencourt AMPure XP magnetic beads, supplied by Beckman Coulter. All clean up steps were performed in accordance with the manufacturer's instructions. AMPure magnetic beads were also be used for size selection of DNA libraries containing DNA of variable sizes.

2.9 PCR reactions

PCR reactions were done in 50 µl volumes of reaction buffer containing ATP/GTP/CTP/TTP (0.5 mM each), 2 µM of each primer, and 50-200 ng template. Phusion DNA polymerase from NEB (New England Biosciences) and Velocity from Bioline were used ad hoc throughout the

work. The annealing temperature used in PCR varied according to the oligonucleotide and was typically 3 ºC lower than the lowest primer T_m used. We allowed 30 seconds of elongation for every 1000 bp of DNA amplified.

A typical PCR cycle consisted of the following:

94 ºC   1 minute   (initial melting of DNA)

94 ºC   30 seconds (DNA melting)

50 ºC   30 seconds (primer annealing)   Repeated for 35 cycles

72 ºC   35 seconds (elongation)

72 ºC   10 minutes (final elongation)

2.10 Megaprimer PCR reactions

Megaprimer PCR was used to generate point mutations in large DNA fragments that lacked appropriately positioned restriction sites (Figure 2.1). The first PCR reaction generates a "megaprimer containing the point mutation required. The megaprimer is then used in a second PCR reaction, along with a second flanking primer to generate the final product containing the point mutation.

2.11 Plasmid and RNA extraction using Qiagen maxiprep, miniprep, or RNeasy extraction kits

Nucleic acid extractions were done using Qiagen kits. Briefly, the DNA is bound to a silica column, washed with a high salt solution, and then eluted in water. DNA for *in vitro* transcription was extracted on a large scale using the Qiagen maxiprep kit. Small scale extractions for generating plasmid stocks and cloning vectors utilised the Qiagen miniprep kit. RNA extractions for primer extension assays were done using the RNeasy kit.
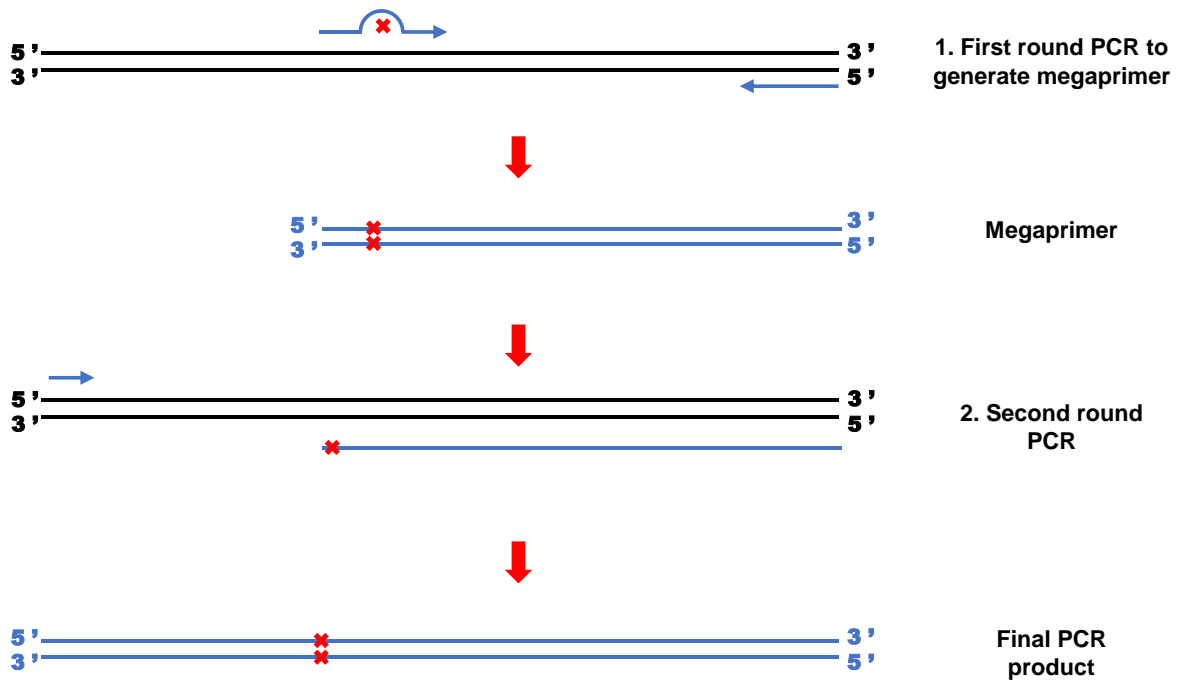
Figure 2.1 Mega-primer PCR

Black lines show template DNA, blue lines show PCR products and primers.

2.12 Ligation of PCR products into plasmid vectors

PCR products were purified using Qiagen PCR purification columns according to the manufacturer's instructions. The kit removes oligonucleotide primers under 40 bp. Following clean-up PCR products were digested with the appropriate restriction enzymes, re-purified, and then used in ligation reactions. Plasmid vectors used in ligation reactions were first digested with appropriate restriction enzymes and then treated with alkaline phosphatase to cleave 5' hydroxyl groups. Vectors smaller than 10 kb were purified using gel extraction as described above. Vectors larger than 10 kb were purified using phenol-chloroform extraction, followed by ethanol precipitation. Ligation reactions were done in 20 μl containing 3/1 ratio of plasmid vector to purified and restricted PCR product (about 50 ng and 100-200 ng respectively), 2 μl of supplied 10x buffer, and 2 μl of T4 DNA ligase.

2.13 Sanger sequencing of plasmid constructs

Sanger sequencing was done by the Functional Genomics and Proteomics Facility at the University of Birmingham. Sequencing reactions were submitted in a 10 μl total volume, with 1 μl 10 μM primer, and at least 200 ng of DNA.

2.15 Agarose gel electrophoresis

One % (*w/v*) agarose gels were prepared by dissolving 1 g of agarose to 100 ml of 1x TBE. The suspension was microwaved for 1 minute, on a high setting, to dissolve the agarose. Ethidium bromide (1 % *v/v*) was added before the gel was poured. Gels were run at 150 V for 20-30 minutes in 1x TBE running buffer, or as required.

2.16 Polyacrylamide gel electrophoresis (PAGE)

Acrylamide gels (7.5 % *w/v*) were prepared by mixing the appropriate reagents (5x TBE, 30 % (*w/v*) acrylamide stock (Geneflow), and distilled water) together to the required concentration, before polymerisation with 0.01 volumes of 10 % *w/v* APS (ammonium persulphate) and a

0.001 volumes of TEMED (N,N,N′,N′-Tetramethylethylenediamine). After running at 20 mA for 20 mins, gels were stained for 10 minutes in water containing 1 % (*v/v*) ethidium bromide (Fisher Scientific). Loading dye was used to track migration of samples in the gel. A UV transilluminator was used to visualise the DNA after electrophoresis.

2.17 Denaturing PAGE

Six % (*w/v*) denaturing PAGE was done using Sequagel reagents (Geneflow) or by adding urea to 6 % (*w/v*) acrylamide solutions to a final concentration of (36 % *w/v*).

2.18 Oligonucleotides used in this study

Oligonucleotides (Life Technologies or ALTA Biosciences) are shown in Table 2.1.

2.19 Bacterial strains used in this study

All strains are shown in Table 2.2.

2.20 Plasmid vectors used in this study

Plasmid are shown in Table 2.3.

Table 2.1 Oligonucleotides used in this study

Oligo name (F: Forward, R: Reverse)	Sequence (5' -> 3')
	Oligonucleotides used for generation of CRP target fragments located by bioinformatics screen on p948 & p666
H10407 CRP target 1F	GGCTGC**GAATTC**TTTTGTGATTAATTTCACAAAATAAGGTGTTATTCAGTGTGTGCTGCAATATTCAGGATGCATGA
H10407 CRP target 1R	GCCCG**AAGCTT**CATGTTATTCTCCATATCAGATAAATAGCACCTGTATTTCATGCATCCTGAATATTGCAGCACAC
H10407 CRP target 2F EXTENDED version F	GGCTGC**GAATTC**ATAAAGTGATAAAAATCACATAAAATTTTTATTAAAAGGATATAACCTTCATATCACTTGTAATTAAATTTGTGTCA
H10407 CRP target 2R EXTENDED version R	GCCCG**AAGCTT**CATGGATATACTTCTTAAGTATTATAAACAAGGTGGAACAGTTGTTATGGTAACCCATGACACAAATTTAATTACAA
H10407 CRP target 3F	GGCTGC**GAATTC**TAGTATGATACACATCACAAAAAAATAAAAAAGTTTGCGCAATCGTTCTGATTTTGAT
H10407 CRP target 3R	GCCCG**AAGCTT**CATATTACCTCCGAAACACGTCGTCCACGAATATTTAAATCAAAATCAGAACGATTGC
H10407 CRP target 4F	GGCTGC**GAATTC**TTTTGTGATTGATTTCATAAAACAAGACGTCGTTTAATATGCGCCTGAATATTCACGAGGGAAGTCTTTCTTTTTATATGGTG
H10407 CRP target 4R	GCCCG**AAGCTT**CATGTTTTTACCTTTTATTCATAACAGGGTGAGCAATGACAAAAAATATTTTATTTCCGGGAACACCATATAAAAGAAAGAC
H10407 CRP target 5F	GGCTGC**GAATTC**TTCGTACTGTCACCTTTTTAAATCATAAAAACTCTGATATTTGAACGAAGTCAAATTT
H10407 CRP target 5R	GCCCG**AAGCTT**GAACGATTTTTTCCAAAAAATGATGATGAATAAAAGAAATTTGACTTCGTTCAAATAT
H10407 CRP target 6F	GGCTGC**GAATTC**TGAATCATATGCGATTCACCAATGAATCCTTTTTATAATGTAAAATAAATTAAAATA
H10407 CRP target 6R	GCCCG**AAGCTT**CATTAACTTATATTTTAAATAATGTATTTTAATTTATTTTACATTA
H10407 CRP target 7F	GGCTGC**GAATTC**TAACATGATGCAACTCACAAAAAAAATAAAAAAATTGCAAAATCCGTTTAACTAATCT
H10407 CRP target 7R	GCCCG**AAGCTT**CATGTTACCTCCCGTCATGTTGTTTCACGGATATTTGAGATTAGTTAAACGGATTTTG
H10407 CRP target 8F	GGCTGC**GAATTC**ATATGTGATATTAATAGCACAAGTTCCTTACCAGATGAATTTGTAAAGAAC

69

H10407 CRP target 8R	GCCCG**AAGCTT**CATTCTTTGCCAGCAAATCCTGCACACGTGGGTTCTTTACAAATTCATCTG
H10407 CRP target 9F	GGCTGC**GAATTC**TTTTGTGGAGTGGGTTAAATTATTTACGGATAAAGTCACCAGAGGTGGAAAAATGAAA
H10407 CRP target 9R	GCCCG**AAGCTT**TTTCATTTTTCCACCTCTGGTGACTTTATCCGTAAATAATTTAACCCACTCCACAAAA
H10407 CRP target 10F	GGCTGC**GAATTC**TTTTCTCTGAAGCTCCCACTCCAGATGATGTAGTTCATTTTTTAACAACTTTATATTT
H10407 CRP target 10R	GCCCG**AAGCTT**CTTAAGAATAATATAGTAGAAAATCACAAAAAAAATATAAAGTTGTTAAAAAATAAAT

Oligonucleotides used for generating PestA2

P*estA2* F	GGCTGC**GAATTC**TAGTATGATACACATCACAAAAAAATAAAAAAGTTTGCGCAATCGTTCTGATTTTGAT
P*estA2* R	GCCCG**AAGCTT**CATATTACCTCCGAAACACGTCGTCCACGAATATTTAAATCAAAATCAGAACGATTGC

Oligonucleotides used for generating PestA1

P*estA1* F	GGCTGC**GAATTC**TAACATGATGCAACTCACAAAAAAAATAAAAAAATTGCAAAATCCGTTTAACTAATCT
P*estA1* R	GCCCG**AAGCTT**CATGTTACCTCCCGTCATGTTGTTTCACGGATATTTGAGATTAGTTAAACGGATTTTG

Oligonucleotides used for generation of PestA2/PestA1 hybrid promoters

P*estA2.1* F	GGCTGC**GAATTC**CATGATGCAACTCACAAAAAAATAAAAAAGTTTGCGCAATCGTTCTGATTTTGATTTAAAT
P*estA2.1* R	GCCCG**AAGCTT**CATATTACCTCCGAAACACGTCGTCCACGAATATTTAAATCAAAATCAGAACGATTGCGC
P*estA2.2* F	GGCTGC**GAATTC**TATGATACACATCACAAAAAAAATAAAAAAATTGCGCAATCGTTCTGATTTTGATTTAAAT
P*estA2.2* R	GCCCG**AAGCTT**CATATTACCTCCGAAACACGTCGTCCACGAATATTTAAATCAAAATCAGAACGATTGCGC
P*estA2.3* F	GGCTGC**GAATTC**TATGATACACATCACAAAAAAATAAAAAAGTTTGCAAAATCGTTCTGATTTTGATTTAAAT
P*estA2.3* R	GCCCG**AAGCTT**CATATTACCTCCGAAACACGTCGTCCACGAATATTTAAATCAAAATCAGAACGATT
P*estA2.4* F	GGCTGC**GAATTC**TATGATACACATCACAAAAAAATAAAAAAGTTTGCGCAATCCGTTTAACTAATCTTAAAT
P*estA2.4* R	GCCCG**AAGCTT**CATATTACCTCCGAAACACGTCGTCCACGAATATTTAAGATTAGTTAACGGATT
P*estA2.5* F	GGCTGC**GAATTC**TATGATACACATCACAAAAAAATAAAAAAGTTTGCAAAATCCGTTTAACTAATCTTAAAT

P*estA2.5* R	GCCCG**AAGCTT**CATATTACCTCCGAAACACGTCGTCCACGAATATTTAAGATTAGTTAAACGGATTTTGCAA
P*estA2.6* F	GGCTGC**GAATTC**TATGATACACATCACAAAAAAATAAAAAAGTTTGCGCAATCGTTCTGATTTTGATTCAAAT
P*estA2.6* R	GCCCG**AAGCTT**CATATTACCTCCGAAACACGTCGTCCACGAATATTTGAATCAAAATCAGAACGATTGCGC
P*estA2.7* F	GGCTGC**GAATTC**TATGATACACATCACAAAAAAATAAAAAAGTTTGCGCAATCGTTCTGATTTTGATTTAAAT
P*estA2.7* R	GCCCG**AAGCTT**CATGTTACCTCCCGTCATGTTGTTTCACGGATATTTAAATCAAAATCAGAACGATTGCGC

<div align="center">

Oligonucleotides for generating PeltAB derivatives

</div>

P*eltAB* 1.1 F	GGCTGC**GAATTC**GCATGGATGTTTTATAAAAAACATGATTGACATCATGTTGCATATAGG
P*eltAB* 1.2 F	GGCTGC**GAATTC**GCATGGATGTTTTATAAAAAACATGATTGACATCATGTTGCATATAGG
P*eltAB* 1.3 F	GGCTGC**GAATTC**GCATGGATGTTTTATAAAAAACAACATTGACATGTTGTTGCATATAGG
P*eltAB* R	GCCCG**AAGCTT**CATCGAGGATATATATCATACAAGAAGACAATCCGGAAAAAGAT

<div align="center">

Oligonucleotides used for inserting lacUV5 promoter into pJRH1-4

</div>

lacUV5 promoter F	GGCTGC**TGATCA**TTTACACTTTATGCTTCCGGCTCGTATAATGTGTGGAACATCCCCTTTATCTAGGAGG**CCATGG**CTGCAGACACTTGAATG

<div align="center">

Oligonucleotides used for sequencing and amplification of DNA fragments from pRW50 and pSR

</div>

pSR UP	CCATATATCAGGGTTATTGTCTC
pSR DOWN	CATCACCGAAACGCGCGAGG
pRW50 UP	GTTCTCGCAAGGACGAGAATTTC
pRW50 DOWN	AATCTTCACGCTTGAGATAC

<div align="center">

Oligonucleotides for used for primer extension

</div>

Universal Primer (for T7 sequencing reactions)	GTAAAACGACGGCCAGT
D49724	GGTTGGACGCCCGGCATAGTTTTTCAGCAGGTCGTTG

<div align="center">

Oligonucleotides for PaatS expression analysis (in addition to those listed above)

</div>

Target 2 ΔSD sequence reverse (used with Target 2 forward)	GCCCG**AAGCTT**CATGGATATTGAAGAAAAGTATTATAAACAAGGTGGAACAGTTGTTATGGTAACCCATGACACAAATTTAATTACAA

Oligonucleotides used for cloning aat operon subunits into BACTH plasmids

pKT25 aatC UP	GCATGC**CTGCAG**GGATGATTAAATTAAGTATTGA
pKT25 aatC DOWN	TACTTA**GGTACC**CGTTAAAAGGATATAACCTTCA
pUT18 aatS UP	GCATGC**CTGCAG**GATGATTTTTTTTTTGTTATT
pUT18 aatS DOWN:	TACTTA**GGTACC**CGTGTTGTTTTAAAAGAACTAG
pUT18 aatS short UP:	GCATGC**CTGCAG**GATGTCATCTGCTAGCAGTATTTT
pUT18C aatS DOWN:	TCATTA**GGTACC**CGTCATGTTGTTTTAAAAGAACTAG
pKT25 aatP F	GCATGC**CTGCAG**GGATGTCATTTGTTTTTATGATACTG
pKT25 aatP R	TACTTA**GGTACC**CGTGAAAATCTTTCTTTTATTAA
pKT25 aatA F	GCATGC**CTGCAG**GGATGAAGATTATATTTATATTATTA
pKT25 aatA R	TACTTA**GGTACC**CGCTAACCAATAAAAAACAAATGCAT
pKT25 aatB F	GCATGC**CTGCAG**GGATGAAATTGAATTATATAAAATGC
pKT25 aatB R	TACTTA**GGTACC**CGTTAATCATGTAGTGTTATCTCAAATGC
pKT25 aatD F	GCATGC**CTGCAG**GGATGATTGTTAAAATGGAGAATATAATT
pKT25 aatD R	TACTTA**GGTACC**CGTTATACATCAATGACAAAAAATGG
pUT18 aatP F	GCATGC**CTGCAG**GATGTCATTTGTTTTTATGATACTG
pUT18 aatP R	TACTTA**GGTACC**CGTGAAAATCTTTCTTTTATTAAGAT
pUT18 aatA F	GCATGC**CTGCAG**GATGAAGATTATATTTATATTATTAGCT
pUT18 aatA R	TACTTA**GGTACC**CGACCAATAAAAAACAAATGCAT
pUT18 aatB F	GCATGC**CTGCAG**GATGAAATTGAATTATATAAAATGC
pUT18 aatB R	TACTTA**GGTACC**CGATCATGTAGTGTTATCTCTGC
pUT18 aatC F	GCATGC**CTGCAG**GATGATTAAATTAAGTATTGAT
pUT18 aatC R	TACTTA**GGTACC**CGAAAGGATATAACCTTCA
pUT18 aatD F	GCATGC**CTGCAG**GATGATTGTTAAAATGGAGAATATAATT
pUT18 aatD R	TACTTA**GGTACC**CGTACATCAATGACAAAAAATGG

Oligonucleotides used for generating pJRH1-4

aat operon F	TGCGAAGCTG**CCATGG**CTGCAGACACTTGAATGTCATTTGTTTTTATGATACTG
aat operon *R*	TACTTA**GGATCC**TTATACATCAATGACAAAAAATGG
aatS inactive R	TACTTAAGAAGTATATCCATAATTTTTTTTTTTGTTATTATTATCTAAGCTGGCCGTTG G
CexE F	TGCGAAGCTG**CCATGG**CTGCAGACACTTGAATGAAAAAATATATATTAGGTGTTATTC TGG
CexE R	TACTTA**CCATGG**TTATTTATACCAATAAGGGGTGTCACCACCGGCAGA

72

Oligonucleotides used for sequencing pJRH1-4 constructs

CexE seq F 2	GTGTTATTCTGGCTATGGGGTCTCTCTC
aatP seq F 2	GATACTGAAAGAGGCGTTGC
aatA seq F 2	CAGTCAAACTTGTGATGCGG
aatB seq F 2	GTACTACAATGAAATTAACGCACT
aatC seq F 2	AATAACAGCGGTATGGAGG
aatD seq F 2	ACAAGGTGGAACAGTTGTTATGG
Pre-CexE R seq 2	AGAGAGACCCCATAGCC
CexE R seq 2	TACAATGTCGGGACTCAACC
Post-CexE R seq 2	TGACAGTTGCCCCTCCCT
aatP R seq 2	GAATAGTCATTGTCTCGACG
aatA R seq 2	ATAACAGAGTCCCCTCTCAG
aatB R seq 2	AGAAATAACAACGCGATCAG
aatC R seq 2	ACCCCTAATAACAATGTGC

Oligonucleotides used for FLAG₃ tagging of crp in ETEC H10407 using FRUIT

JW472	CCGACGCGCAGTTTA
JW473	CACGTTGTGTTTTCATGC
Upstream targeting *crp*FLAG₃	CTGGCTCTGGAGAAAGCTTATAACAGAGGATAACCGCGCATGGTGGGCGGTGGCGACTAC
Downstream targeting *crp*FLAG₃	AGACAAGAACCATTCGAGAGTCGGGTCTGTTTGCGGTTTGCCAAGCGCCGCCTTGTCATC
Upstream *crp*FLAG₃ check	CAGTTGATAGCCCCTTCC
Downstream *crp*FLAG₃ check	TGGTGAATAAGCGTGCTC

Oligonucleotides for ChIP-seq library amplification (for an Illumina platform)

NEXTflex ChIP Primer 1	AATGATACGGCGACCACCGAGATCTACAC
NEXTflex ChIP Primer 2	CAAGCAGAAGACGGCATACGAGAT

Table 2.1 Oligonucleotides used in this study

Bold text indicates restriction sites, bold and italicised text describes the application of different sets of oligonucleotides.

Table 2.2 Strains used in this study

Strain	Genotype	Source
M182	Δ(*lacIPOZY) X74 galK galU strA*	Casadaban and Cohen (1980)
M182 Δcrp	M182 $\Delta crp39$	Busby *et al.* (1983)
JCB387	$\Delta nirB \Delta lac$	Typas and Hengge (2006)
ETEC H10407	Wildtype ETEC strain, CFA/I LT$^+$ STh$^+$ STp$^+$ O78:H11	Crossman *et al.* (2010)
ETEC H10407 $\Delta thyA$	ETEC strain H10407 with *thyA* gene deleted.	Stringer *et al.* (2012)
ETEC *crp*FLAG$_3$	ETEC strain H10407 with the FLAG$_3$ inserted into the C-terminal end of the *crp* gene.	This study
BTH101	*F$^-$, cya-99, araD139, galE15, galK16, rpsL1 (Str^r), hsdR2, mcrA1, mcrB1.*	Karimova *et al.* 1998
ER2925	*ara-14, leuB6 fhuA31 lacY1 tsx78 glnV44 galK2 galT22 mcrA dcm-6 hisG4 rfbD1 R(zgb210::Tn10)TetS endA1 rpsL136 dam13::Tn9 xylA-5 mtl-1 thi-1 mcrB1 hsdR2*	(New England Biolabs)
M182 Δhns	Generated by P1 transduction of M182 with *hns*::kan from YN3144 (donated by Ding Jin)	This study

Table 2.3 Plasmids used in this study

Plasmid	Features	Source
pRW50	Large 16.9 kb plasmid featuring *Eco*R1-*Hin*dIII restriction sites sandwiching a cloning site, upstream of a *lacZ* fusion. Encodes Tet^R. RK2 origin.	Lodge e*t al.* (1992)
pRW225	15 kb pRW50-based derivative with a deleted ribosome binding site. Used for translational fusion, encodes Tet^R. RK2 origin.	Islam *et al.* (2012)
pSR	2.6 kb plasmid featuring *Eco*R1-*Hin*dIII restriction sites sandwiching a cloning site, upstream of a λ*oop* terminator site. A pBR322 derived plasmid encoding Amp^R. ColE1 origin.	Kolb *et al.* (1995)
pDCRP	*crp* gene preceding its native promoter (located on *Eco*RI-*Sal*I flanked fragment). pBR322 derived. Encodes Amp^R. ColE1 origin.	Bell *et al.* (1990)
pLG314	pLG339-based vector containing *melR* gene, controlled by a *gal*P2 promoter, subcloned on a *Eco*R1-*Bam*H1 fragment from pCM118-314 within the *Tet^R* gene. Encodes Kan$^R.$ pSC101 origin.	Wade *et al.* (2001)
pCexE	Pet26b vector with *aat* operon cloned into MCS using *Nde*I-*Xho*I restriction site. The *aatD* is in a non-native position downstream of *aatC*. Encodes Kan^R. f1 origin.	(dna2.0) Kindly donated by Dr Yanina Sevastsyanovich
pKD46	Carries λ recombinase system. Encodes Amp^R. R101 origin.	Datsanko and Wanner (2000)
pAMD135	pGEM-T- containing MG1655 *thyA* gene flanked by FLAG$_3$ epitopes. f1 origin.	Stringer *et al.* (2012)
pUT18	Encodes T18 domain of *Bordetella pertussis* adenyl cyclase, downstream of the *Pst*II-*Kpn*I cloning site. Both features are downstream of the *lac* promoter. Encodes Amp^R. ColE1 origin.	Karimova *et al.,* (1998), Karimova *et al.,* (2001)
pUT18C	Encodes T18 domain of *Bordetella pertussis* adenyl cyclase, upstream of the *Pst*II-*Kpn*I cloning site. Both features are downstream of the *lac* promoter. Encodes Amp^R. ColE1 origin.	Karimova *et al.,* (1998), Karimova *et al. al.* (2001)

pKT25	Encodes T25 domain of *Bordetella pertussis* adenyl cyclase, upstream of the *Pst*II-*Kpn*I cloning site. Both features are downstream of the *lac* promoter. Encodes Kan^R. p15a origin.	Karimova *et al.,* (1998), Karimova *et al.* (2001)
pJ204-aatS	Contains putative *aatS* gene under control of *lacUV5* promoter. Ordered from dna 2.0. Encodes Amp^R. Ori_pUC origin.	(dna 2.0.)
pJRH1	pLG314 derivative containing *aat* operon from ETEC strain H10407 (sub-cloned from pCexE). *aatD* is in a non-native position downstream of *aatC*. Under the control of a *lacUV5* promoter. Encodes Kan^R. pSC101 origin	This study
pJRH2	Derivative of pJRH1, but contains *cexE* cloned upstream of *aatP* downstream of the *lacUV5* promoter.	This study
pJRH3	Derivative of pJRH1, but contains an interrupted start codon in the putative *aatS* open reading frame.	This study
pJRH4	Derivative of pJRH2 but contains an interrupted start codon in the putative *aatS* open reading frame.	This study

2.21 β-galactosidase assays

β-galactosidase assays were done as in Miller *et al* (1972). Overnight cultures were set up from fresh colonies, the following day, 100 μl from each overnight incubation was used to sub-culture into fresh media. Cultures were grown to mid-log phase (OD_{650}= 0.3-0.6, measured using a Jenway 6300 spectrophotometer). Cells were lysed with 1 drop of (100 % *v/v*) toluene and 1 drop 1 % (*w/v*) sodium deoxycholate, before vigorous mixing. Lysates were aerated at 37 °C for 20 minutes to remove the toluene. Prior to doing the experiment, 2-Nitrophenyl β-D-galactopyranoside (ONPG) was made up to a concentration of 13 mM (8 mg/ml) in 'Z-buffer' and 271 μl of β-mercaptoethanol was added per 100 ml of ONPG.

Lysates (100 μl) were incubated with 2.5 ml 13 mM ONPG to start the reaction. The solution was allowed to turn yellow before the reaction was stopped with 1 ml of 1 M sodium carbonate. The OD_{420} of the reaction was then measured.

β-galactosidase activity was then calculated using the following formula:

$$\text{Activity (Miller units)} = \frac{1000 \times 2.5 \times A \times OD_{420}}{4.5 \times T \times V \times OD_{650}} = \frac{20,000 \times OD_{420}}{T \quad OD_{650}}$$

T= Time (mins)

A= Final assay volume (3.6 ml)

V= Volume of lysed culture used (0.1 ml).

2.5= Conversion factor to convert OD_{650} into dry protein mass (mg) – assumes that OD_{650} of 1 is equivalent to 0.4 mg/ml bacterial dry mass.

$1000/4.5 =$ Conversion factor to convert A_{420} into moles of O-nitrophenol- assumes that 1mole/ml O-nitrophenol has an absorption of 0.0045 at OD_{420}.

Data shown is from three independent replicates, the mean activity is shown and one standard deviation is shown either side of the mean.

2.22 Bacterial two-hybrid system (BACTH)

The bacterial two-hybrid system is used to detect interactions between proteins (Figure 2.2). Developed by Karimova *et al.* (1998), the system makes use of the *Bordetella pertussis* adenylyl cyclase, which consists of two domains; T25 and T18. Unless in close proximity the two domains generate cAMP. In the BACTH T25 and T18 are encoded by plasmids pKT25 and pUT18 respectively. Genes encoding proteins that may interact are cloned into these plasmids to create T25 and T18 protein fusions. The plasmids encoding the T18 and T25 fusion proteins are used to transform BTH101, a strain lacking adenyl cyclase. If the fusion proteins interact, the two domains of adenyl cyclase are brought together, and cAMP is synthesized. CRP then binds cAMP and switches on expression of the *lacZYA* operon. This is detected using β-galactosidase assays. In this study, all gene fusions were generated by cloning into pKT25 or pUT18 using the *Pst*II, and *Kpn*I restriction sites. In subsequent assays a positive control, which measures the interactions between two leucine zippers, is included. In addition a negative control, which measures the interaction between T25 and T18, was also included.
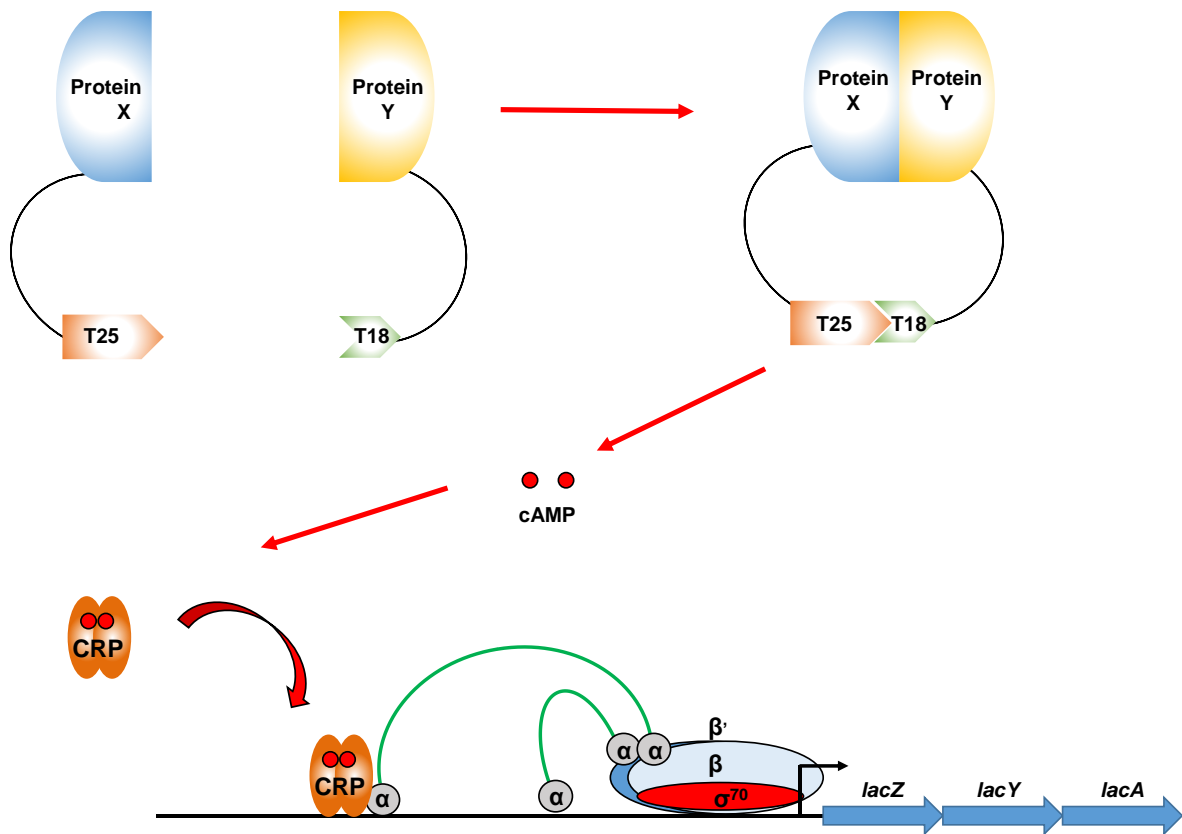
Figure 2.2 The bacterial two-hybrid system (BACTH)

Two proteins of interest are each genetically fused to domain of the *Bordatella pertussis* adenylyl cyclase (T25 and T18). If the two proteins interact, domains T25 and T18 are brought into close proximity and are able to interact, generating cAMP. This activates CRP which leads to *lacZYA* expression. Based on Karimova *et al.* (1998).

2.23 CRP purification

CRP was purified using the Ghosaini method.

A 5 ml overnight culture was inoculated with M182Δcrp cells harbouring the pDCRP plasmid. In preparation for the next day, 1.2 ml of ddH$_2$O was used to rehydrate 60 mg cAMP-agarose, which was mixed for 5 minutes and overnight at room temperature.

The following day, a Bio-rad poly prep column was prepared. This column has an internal reservoir, but is attached to an external reservoir, which is used for pre-loading washes.

The column, and rehydrated cAMP-agarose from the previous day, was moved to a 4 $^{\circ}$C cold room, and was half-filled with ddH$_2$O to equilibrate. cAMP-agarose was then gradually added to the column, and was allowed to settle for two hours (final column volume was 0.8 ml). Two washes were then carried out at high flow rate; first, 10 ml column wash buffer + 1 M NaCl, and then 5 ml dialysis buffer. Zero point two ml of dialysis buffer was kept in the column, on top of the cAMP-agarose; which was left overnight. One ml of the overnight culture was used to inoculate 600 ml of superbroth, supplemented with 100 μg/ml ampicillin, which was incubated with shaking overnight at 37 $^{\circ}$C.

The following day, the OD$_{650}$ of this culture was measured using a ten-fold dilution (in superbroth medium) in a Jenway 6300 spectrophotometer. A volume of this (calculated by using the formula V (μl) = 2000/OD$_{650}$), was saved at -20 $^{\circ}$C for the final SDS-PAGE analysis. A 1.5 ml aliquot of the culture was also saved for plasmid extraction.

The remainder of the 600 ml culture was harvested by first aliquoting the culture into two centrifuge pots, followed by centrifugation of at 1600 xg for 15 minutes at 4 $^{\circ}$C. One of the resulting pellets was stored at -20 $^{\circ}$C, the other was used for the purification. The culture pellet from one pot was resuspended in 6 ml lysis buffer (containing 50 μg/ml PMSF (phenylmethylsulphonyl fluoride in isopropanol) and sonicated. Complete lysis results in a ten-fold drop in OD$_{650}$, so the success of the sonication was assessed using a 1/500 dilution before,

and a 1/100 dilution after of the culture. The lysate was then transferred to an Oak Ridge tube for centrifuging at 10,000 xg for 30 mins at 4 °C. Two µl of the resulting supernatant was retained for SDS analysis (known as sample '2'), and the remainder was used for the remainder of the purifcation.

The culture supernatant was moved to the cold-room, where the cAMP-agarose column was equilibrated using 5 ml dialysis buffer on a high flow rate. Once the buffer volume in the external resevoir had depleted to 0.2 ml, the supernatant was loaded into internal reservoir (at a flow rate of 5 ml/hr). This wash was kept for SDS-PAGE analysis (known as sample '3'). The first 1 ml of lysate wash from the column, and the remainder of the eluate was also kept for SDS-PAGE analysis (known as sample '4'). When the column buffer volume had receded to 0.2 ml 5 ml dialysis buffer was added to the internal reservoir, with a flow rate of 5 ml/hr, an the wash was kept as SDS-PAGE sample '5'. When 0.2 ml buffer remained in the column, 1 ml column wash buffer containing 5 mM 5'+3' AMP was added. Once the buffer volume had receded to 0.2 ml, another 3 ml was added column wash buffer (containing 5 mM 5'+3' AMP) was added. The wash was kept as sample '6' for SDS-PAGE analysis. When the column buffer volume reduced to 0.2 ml, 1 ml column wash buffer was added. Once the column volume receded to 0.2 ml, another 3 ml of column wash buffer was added. The wash was saved as SDS-PAGE sample'7'.

Once the column buffer volume had receded to 0.2 ml, CRP bound to the column was eluted using 5 ml column wash buffer with 5 mM cAMP. Every 18 minutes, eluate fractions were collected in 0.5 ml aliquots using a slower 2 ml/h flow rate. In total, 8 fractions were collected and used in the SDS-PAGE analysis.

All 8 fractions and the other samples collected were analysed on 4-12 % (*w/v*) SDS-PAGE gels (supplied by Invitrogen). The final volumes of each sample were loaded after boiling for 5 minutes; 5 µl of sample 1 was loaded, 10 µl aliquots of samples 2-7 were mixed with 290 µl

SDS-PAGE sample buffer, 10 µl was then loaded. Five µl SDS-7 standards and 5 µl 0.2 µg/µl wild type CRP standard in sample buffer were run alongside the samples. Fractions were then combined according to the SDS-PAGE gels and were placed in dialysis tubing. The tubing was placed in 500 ml CRP stock buffer at 4 $^{\circ}$C and left overnight. The buffer was replenished the next day and left for a 3 hours. CRP preparations stored in microfuge tubes at -20 $^{\circ}$C.

CRP concentration was determined by Biorad assays. BSA standard curves were prepared using concentrations ranging from 0 to 10 mg/ml. CRP dilutions were made in 800 µl aliquot (up to 200-fold) using water, and 200 µl Biorad reagent was added to each. After incubated at room temperature for 5 minutes, OD_{595} measurements were taken. CRP concentration (g/L) was calculated using the BSA standard curve, and the Molar concentration was calculated (the molecular weight of CRP is 23.6 kDa).

CRP concentrations were compared by eye alongside BSA standards on 4-12 % (*w/v*) SDS-PAGE gel (supplied by Invitrogen). Samples were loaded after being boiled for 2 minutes. Samples loaded were: 0.5 µg, 1.0 µg, and 2.0 µg of CRP preparation in sample buffer (note that these amounts were guided by the Biorad analysis). Ten µl of previously purified CRP sample was also loaded. In addition, 5 µl of SDS-7 standards were loaded.

2.24 *In vitro* multi-round transcription assays

Supercoiled pSR plasmid DNA (containing the desired promoter DNA fragment) was purified using a Qiagen Maxiprep kit, and used as a template for subsequent *in vitro* transcription reactions as described by Kolb *et al.* (1995). Reactions were done in buffer containing 20 mM Tris pH 7.9, 5 mM $MgCl_2$, 500 µM DTT, 50 mM KCl, 100 µg ml^{-1} BSA, 200 µM ATP/GTP/CTP, and 10 M UTP and 5 µCi α-P^{32}- UTP, using 16 µg ml^{-1} template. If required DNA was incubated with purified CRP and 0.2 mM cAMP for 5 minutes (37 $^{\circ}$C), before 400 nM RNAP holoenzyme containing σ^{70} (Cambio) was added. The reaction was run for 10 minutes at 37 $^{\circ}$C, and stopped with 20 µl of 'STOP' solution. Four µl of each reaction was

analysed by denaturing PAGE (at 60 W for 1 hr) and exposure of the dried gel to a Biorad phosphorscreen. After exposure the gel image was captured using a Biorad FX® phosphoimager and Quantity One software.

2.25 Radiolabelling of DNA fragments

DNA fragments used in EMSA analysis were flanked by *Eco*RI and *Hin*dIII restriction sites to allow cloning into the plasmid pSR. This permitted *Aat*II-*Hin*dIII fragments to be excised, and ~100-200 ng of DNA was radiolabelled at both ends using T4 polynucleotide kinase according to the manufacturer's instructions. In other cases, 1 µl (~0.1 ng) of primer was used. One µl of P^{32} labelled γ-dATP (10 µCi/µl) was used to label DNA in a 20 µl reaction. Unincorporated nucleotides were removed using two 200 µl G-50 Sephadex columns.

2.26 Electrophoretic mobility shift assays

End-labelled DNA fragments were incubated at 37 °C with purified CRP protein (see results figure legends for more detail), and 0.2 mM cAMP in 1x Transcription buffer. Herring sperm DNA (HS-DNA) was added as a non-specific competitor at a final concentration of 12.5 µg ml^{-1}. After 20 minutes the sample was loaded on a 7.5 % (*w/v*) polyacrylamide gel. The gel was run at 250 V for 1-2 hours before being dried and exposed to a phosphorscreen.

2.27 M13 sequencing reaction

The T7 sequencing kit supplied by USB was used to generate sequencing reactions to calibrate primer extension reaction products.

Ten µg of single-stranded template M13mp18 phage DNA (supplied in kit) was diluted to a volume of 32 µl and incubated with 8 µl 2 M NaOH at room temperature for 10 minutes. Seven µl 3 M sodium acetate (pH 4.8), 4 µl dH$_2$O and 120 µl 100 % (*v/v*) ethanol was added, before incubation at -80 °C for 15 minutes. After centrifugation at 17,000 xg for 15 minutes at

4 ℃, the supernatant was discarded and the pellet washed with ice-cold 70 % (*v/v*) ethanol before centrifugation for 10 minutes at 17,000 xg at 4 ℃. The pellet was dried by vacuum desiccation, and was re-dissolved in 10 μl of dH$_2$O. Ten ρmol of the undiluted stock of the universal primer and 2 μl of the annealing buffer was added before vigorous mixing and centrifugation, before incubation at 65 ℃ for 5 minutes. The tube was transferred to a 37 ℃ heat block for 10 minutes, then to room temperature for 5 minutes, before brief centrifugation.

Two point five μl of each of 'A' Mix-short, 'C' Mix-short, 'G' Mix-short, and 'T' Mix-short was transferred into four tubes. One μl of T7 polymerase stock (8 units/μl) was diluted using 4 μl of dilution buffer in preparation for the labelling step. Fourteen to fifteen μl of the annealed template/primer mix was mixed with 3 μl labelling mix, 1 μl of α-P^{32} dATP (10 μCi/μl), and 2 μl of the diluted T7 polymerase mix, before incubation at room temperature for 5 minutes. During this period, the four tubes, 'A','G','C', and 'T' were warmed at 37 ℃ for over 1 minute. Termination reactions were carried out by transferring 4.5 μl of the reaction into each of the four tubes, with gentle mixing by pipetting. This was incubated at 37 ℃ for 5 minutes. The reactions were stopped using 5 μl of 'STOP' solution, before mixing and brief centrifugation. The tubes can be stored at -20 ℃ after this stage.

Before being loading on a 6 % (*w/v*) denaturing PAGE, 3 μl of each aliquot was heated at 75-80 ℃ for 2 minutes, before 1.5-2 μl of each sample was loaded.

2.28 Primer extension reactions

RNA was purified from cells harbouring a pRW50 plasmid containing the cloned promoter fragment of interest. The experiment was carried out over two days:

Day 1: Primer D49724 (anneals downstream of the *Hin*dIII site on pRW50) was radiolabelled using T4 polynucleotide kinase and γ- P^{32} labelled ATP, as described above. Forty μg of purified RNA was mixed with 1 μl of radiolabelled primer and precipitated using 1/10 volume

of 3 M sodium acetate, and 2.5 volumes of cold 100 % (*v/v*) ethanol. The sample was recovered by centrifugation at 17,000 xg at 4 $^{\circ}$C for 10 minutes. The pellet was washed with 1 ml 70 % (*v/v*) ethanol, before centrifugation at 17,000 xg at 4 $^{\circ}$C for 5 minutes, and drying in a speedvac at 45 $^{\circ}$C for 5-10 minutes. The pellet was re-suspended in 30 µl hybridisation buffer (20 mM HEPES, 0.4 M NaCl, 80 % formamide), before vigorous mixing and incubation for 5 minutes at 50 $^{\circ}$C. To anneal the primer, the RNA/primer solution was incubated at 75 $^{\circ}$C for 15 minutes, then at 50 $^{\circ}$C for 3 hours. Seventy-five µl cold 100 % (*v/v*) ethanol was added, the solution was mixed and then left overnight at -80°C.

Day 2: The annealed primer/RNA was centrifuged at 17,000 xg for 10 minutes at 4 $^{\circ}$C, before being washed with 1 ml of 70 % (*v/v*) ethanol, with a further spin at 17,000 xg for 5 minutes at 4 $^{\circ}$C. The supernatant was removed and the pellet was then dried by speedvac for 20 minutes. The pellet was re-suspended in 31 µL DEPC-treated dH$_2$O.

Primer extension reactions were carried out in 50 mM Tris-HCl pH 8.3 (at 25 $^{\circ}$C), 50 mM KCl, 10 mM MgCl$_2$, 0.5 mM spermidine (in the form of a 5x reverse transcriptase buffer, purchased from Promega), and 1 mM DTT, and 0.2 mM dNTPS (final concentrations) were added. The total reaction volume was made up to 50 µl, through the addition of 0.6 µl of RNasin (purchased from Promega, a RNase inhibitor), and 2.5 µl AMV reverse transcriptase (purchased from Promega). The sample was incubated at 37 $^{\circ}$C for 1 hour, before incubation at 72 $^{\circ}$C (to inactivate the reverse transcriptase), and was spun briefly. To degrade RNA, 1 µl of a 10 mg/ml RNase A solution and was incubated for 30 minutes at 37 $^{\circ}$C, before DNA was precipitated using 6.7 µl 3 M ammonium acetate (pH 4.8) and 125 µl 100 % (*v/v*) ethanol, with 30 minutes incubated at -80 $^{\circ}$C. The sample was centrifuged at 17,000 xg for 10 minutes 4 $^{\circ}$C, before washing with 1 ml cold 70 % (*v/v*) ethanol, and further centrifugation at 17,000 xg for 10 minutes at 4 $^{\circ}$C. The supernatant was removed and the pellet was dried by speedvac for 20 minutes. The pellet was resuspended in 4 µl STOP solution, and 1-2 µl of this was run on a

pre-run 6 % (*w/v*) denaturing gel alongside M13 sequencing reactions, which was run at 60 W

for 2 hours.

2.29 Generation of 'G+A' ladder

To generate 'G+A' ladders, the DNA sequence of interest was first cloned into pSR. From this,

an *Aat*II-*Hin*dIII fragment was created, and radio-labelled at the *Hin*dIII end. Three to four μl

(~ 100-200 ng) of this fragment was diluted in 8 μl of dH_2O and mixed with 50 μl of formic

acid. The mixture was incubated at room temperature for 2.5 minutes then stopped with 200 μl

0.3 M sodium acetate solution, 1 μl (20 mg/ml) glycogen, and 700 μl ice-cold 100 % (*v/v*)

ethanol. DNA was precipitated at -70 °C for 15 minutes, before centrifuging at 17,000 xg for

15 minutes at 4 °C. The DNA pellet was washed three times in 70 % (*v/v*) ethanol, each time,

centrifuging at 17,000 xg at 4 °C for 10 minutes. The pellet was then dried under vacuum,

before resuspension in 1 M piperidine (diluted with dH_2O from 10 M stock). The reaction was

incubated at 90 °C for 30 minutes and stopped by the addition of 1 μl glycogen, 10 μl of 3 M

sodium acetate, and 300 μl ice-cold 100 % (*v/v*) ethanol, before precipitation at -70 °C for

15 minutes. This DNA was collected by centrifugation at 17,000 xg for 15 minutes at 4 °C and

was washed twice in 70 % (*v/v*) ethanol as described previously. The DNA pellet was dried

under vacuum before being resuspended in 20 μl of DNAse I blue.

2.30 DNAse I footprints

Fragments for DNAse I footprinting were *Aat*II-*Hin*dIII fragments from a pSR plasmid

containing the promoter insert of interest. Plasmids were first digested using *Hin*dIII and

treated with alkaline phosphatase. Products were then purified, and digested using *Aat*II before

being purification by gel extraction. Fragments were then labelled at the *Hin*dIII-restricted end,

using γ-P^{32} ATP and T4 polynucleotide kinase. Footprints were done at 37 °C in buffer

containing 120 mM KCl, 100 μM EDTA, 20 mM Tris pH 7, and 10 mM $MgCl_2$. Ten to forty

ng of DNA was used per footprint with 12.5 µg ml^{-1} Herring Sperm DNA as a competitive inhibitor. Purified CRP was diluted in buffer containing 0.2 mM cAMP before being incubated with DNA for 10 minutes at 37 $^{\circ}$C (at a total reaction volume of 20 µl). Two µl diluted DNAse I (at a concentration where each DNA fragment is cleaved once) was added for 1 minute. This dilution was determined through a previous calibration experiment. Reactions were stopped using 200 µl of a 'STOP' solution containing, 0.3 M sodium acetate pH 7 and 10 mM EDTA. Reactions were then subjected to phenol-chloroform extraction and ethanol precipitation, and were resuspended in 4 µl of DNAse I blue. Prior to loading, all samples and the G +A ladder were heated to 90 $^{\circ}$C for 2 minutes. These were loaded onto a pre-run 6 % (*w/v*) denaturing gel run at 60 W.

2.31 TCA precipitations

TCA (Trichloroacetic acid) precipitation was used to extract secreted proteins in culture supernatants. One ml of overnight culture was used to inoculate 50 ml of LB containing the appropriate antibiotics. Cells were grown to mid-log phase (OD$_{600}$ = 0.5). Cells from 20 ml of each culture were centrifugation at 1600 xg for 20 minutes at 4 $^{\circ}$C. Supernatants were harvested, and pellets were saved (both supernatants and cell pellets can be stored at -20 $^{\circ}$C). The volumes of supernatants were adjusted to compensate for differences in OD$_{600}$ values, before being passed through 0.22 µM filters, into clean polypropylene tubes. Two ml of ice-cold 100 % (*w/v*) TCA was added to each tube, before thorough mixing and incubation for 45 minutes on ice. Precipitated proteins were collected by centrifugation at 10,000 xg for at least 90 minutes at 4 $^{\circ}$C. Following centrifugation, the supernatant was discarded and protein pellets were resuspended in 1 ml 100 % (*v/v*) cold methanol and transferred to clean microfuge tubes. After centrifugation for 45 minutes at 17,000 xg at 4 $^{\circ}$C, supernatants were discarded, and pellets were dried at 37 $^{\circ}$C for at least 30 minutes. Protein pellets were resuspended in 20 µl 2x SDS-PAGE loading dye. Cell pellets were re-suspended in 1 ml PBS. Volumes were adjusted

according to OD_{600} values. One μl of the cell suspension was mixed with 10 μl 2x SDS-PAGE loading dye.

TCA precipitation samples were run on 4-12 % (*w/v*) Bis-Tris polyacrylamide pre-cast protein gels purchased from Life Technologies. Protein gels were mounted in tanks containing 1x MES SDS running buffer. Prior to loading, 10 μl samples of both TCA precipitated protein samples and cell pellet samples (resuspended in 2x SDS-PAGE loading dye) were boiled for 10 minutes at 100 °C. All 10 μl of sample was loaded onto the gel. Gels were calibrated with 10 μl of PageRuler Plus prestained protein ladder (supplied by Thermo Scientific), which was boiled for 10 minutes at 100 °C before loading into a separate well. Gels were run at 100 V for 1 hour, before gels were fixed and silver stained.

2.32 Silver staining of protein gels

Protein gels were fixed and stained using a SilverQuest kit (purchased from Life Technologies).

2.33 FRUIT

For genome wide ChIP-seq experiment to measure CRP binding, it was necessary to tag CRP C-terminally, using a 3x FLAG tag (N-DYKDDDDK$_3$-C) (Einhauer and Jungbauer, 2001). This tag was fused to the *crp* gene on the chromosome of ETEC H10407 using a new recombineering technique, described in Figure 2.3, known as FRUIT (<u>F</u>lexible <u>R</u>ecombineering <u>U</u>sing <u>I</u>ntegration of *<u>thyA</u>*) (Stringer *et al*, 2012). The FRUIT technique for adding a FLAG$_3$ tag involves amplifying the *Escherichia coli* MG1655 *thyA* gene flanked by FLAG$_3$ from plasmid pAMD135 (Stringer *et al*, 2012). This requires a primer containing the 5' sequence consisting of 40 nucleotides of the N-terminal region of the ETEC *crp* gene
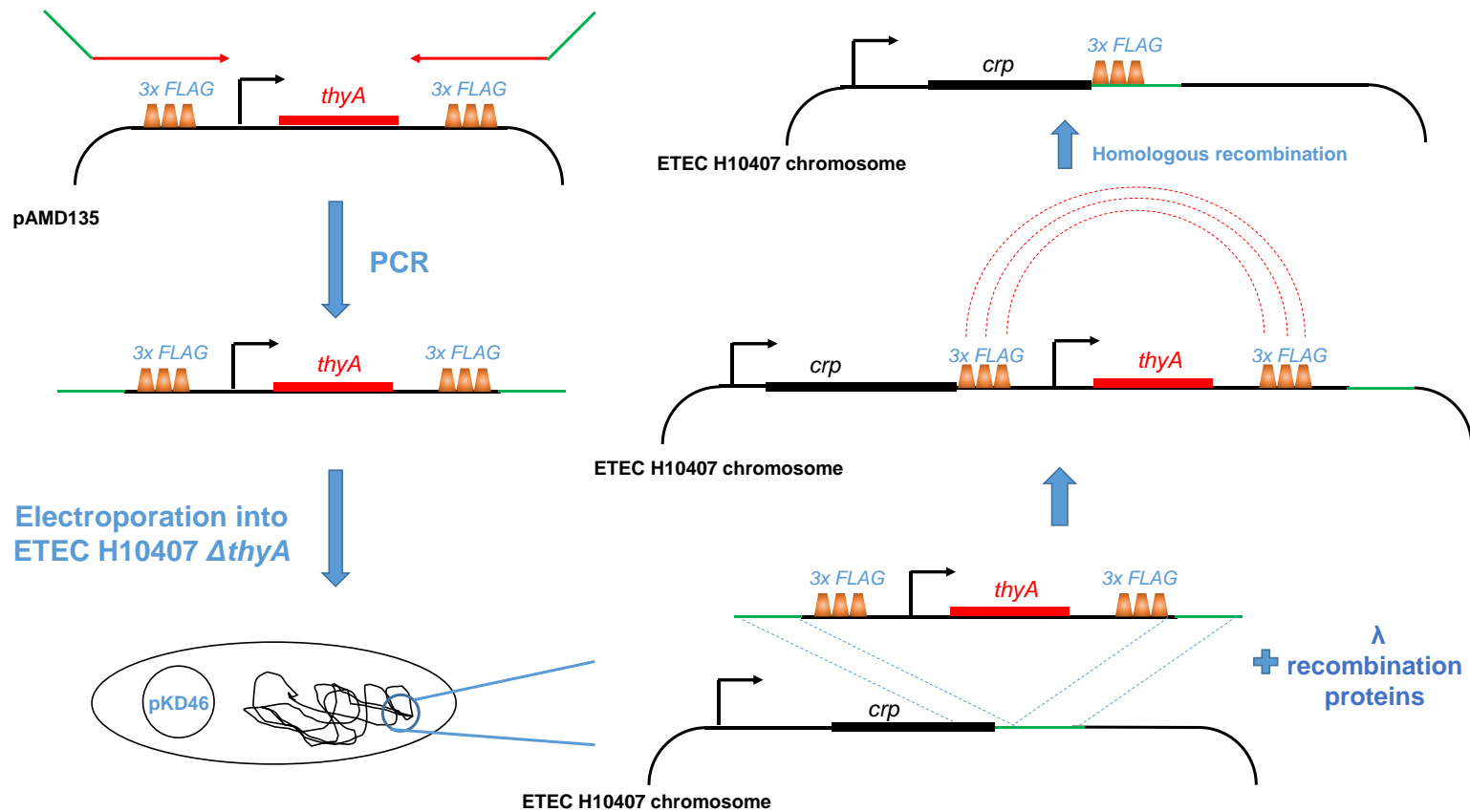
Figure 2.3 Flexible recombineering using integration of *thyA* (FRUIT)

The *thyA* gene, flanked by FLAG$_3$ was amplified from pAMD135 using primers 'upstream targeting *crp*FLAG$_3$' and 'downstream targeting *crp*FLAG$_3$', which have 5' regions (shown in green) which anneal to the 3' end of the *crp* gene in ETEC H10407. The resulting PCR product is electroporated into ETEC H10407 Δ*thyA*, together with pKD46 which encodes the λ Red recombination proteins. The PCR product is recombined into the 3' end of the *crp* gene on the ETEC H10407 genome, and is selected for using media without thymine. The *thyA* gene, and one set of FLAG$_3$ is removed by selection using media containing 20 μg/ml trimethoprim and 100 μg/ml thymine. Based on Stringer *et al.* (2012).

('upstream targeting *crp*FLAG₃' and 'downstream targeting *crp*FLAG₃'). This product (100-200 ng PCR product) is electroporated into ETEC cells lacking the *thyA* gene, and carrying pKD46, a plasmid which encodes the λ Red recombination system, (Datsenko and Wanner, 2000). Cells are grown in M9 minimal media containing 100 μg/ml thymine, and the Red system is then induced with 0.2 % (*w/v*) arabinose resulting in recombination of FLAG₃-*thyA*-FLAG₃ into the C-terminal region. Recombinant colonies are selected for by growth on M9 minimal media lacking thymine but containing ampicillin. Low-level homologous recombination occurs spontaneously between the two sets of FLAG tags, if cells are grown in minimal media containing 20 μg/ml trimethoprim and 100 μg/ml thymine. This removes the *thyA* gene, and one set of 3x FLAG tags. Strains were verified by sequencing colony PCR products using 'upstream *crp*FLAG₃ check' and 'downstream *crp*FLAG₃ primers'. Following tagging of the *crp* gene using FLAG₃, the resulting strains was streaked onto solid MacConkey maltose media alongside wildtype cells, and the Δ*crp* mutants. Colonies of the *crp*FLAG₃ strain appeared purple on the plates.

2.34 ChIP-seq in ETEC H10407

The method used for ChIP-seq and subsequent data analysis was as described by Singh *et al.* (2014). All experiments were performed in duplicate. ETEC cells were grown to exponential phase (sub-cultured from overnights) in 40 ml cultures (M9 minimal media supplemented with 1 % (*w/v*) fructose) and fixed using formaldehyde at a final concentration of 1 % (*v/v*). After incubation for 20 minutes, fixing was quenched using 10 ml of 2.5 M glycine. Cells were collected by centrifugation for 5 minutes at 1600 xg at 4 °C, and were re-suspended in 0.5-1x the original culture volume in 1x TBS. Cells were collected by centrifugation as before, before being resuspended in 1 ml 1x TBS, and this washing step was then repeated. After washing, cells were resuspended in 1 ml FA lysis buffer (150 mM NaCl unless stated) containing 4 mg/ml lysozyme, and incubated at 37 °C for 30 minutes. The lysate was then chilled on ice for

at least 5 minutes before sonication, using a Bioruptor sonicator, set to two 15-minute cycles (30 seconds on, 30 seconds off), at 4 °C. Cell lysates were cleared by centrifugation at 17,000 xg for 5 minutes, and supernatants saved. Lysates were diluted in FA lysis buffer in such a way that there was 1 ml of lysate for every 20 mls of original culture. Twenty µl of each diluted lysate was saved and used as an 'input' control for real-time PCR experiments.

For immunoprecipitations, cocktails were set containing 500 µl of diluted lysate, 300 µl FA lysis buffer, and 25 µl Protein A beads (GE Healthcare, GE17-0780-01), which were made up in a 50% (*v/v*) slurry with TBS. Note that all steps involving direct pipetting of Protein A beads used blunted tips to avoid shearing the beads. Two µl of anti-H-NS or anti-FLAG, or 1 µl of anti-σ^{70} was then added to the cocktail. The cocktail was mixed on a rotating wheel for 90 minutes at room temperature.

To create the library for Illumina NGS sequencing, the protein A beads were collected by centrifugation at 1500 xg for 1 minute, and the supernatant removed and discarded. Seven hundred µl of FA lysis buffer was used to resuspend the beads, which were then transferred to a Spin-X column. This column was rotated for 3 minutes at room temperature before the beads were collected by centrifugation at 1500 xg for 1 minute, and the flowthrough was discarded. All wash steps were done using 750 µl of the solution mentioned, with rotation for 3 minutes before the beads were collected by centrifugation at 1500 xg for 1 minute (with flow-through discarded). Following transfer of beads to the spin-X-column, beads were washed twice with FA lysis buffer and twice with 10 mM Tris-HCl, pH 7.5. DNA fragment blunting reactions were done by resuspending beads in buffer consisting of 10 µl 10x quick blunting buffer, 10 µl dNTP mix (supplied in kit), 80 µl dH$_2$O and 2 µl blunt enzyme mix. Column lids were sealed with parafilm and the columns were rotated (in such a way that tubes were not inverted) at room temperature for 30 minutes. The beads were then washed twice with FA lysis buffer, and twice with 10 mM Tris-HCl, pH 8. 'A tail' addition was done by resuspending beads in the

buffer consisting of 10 µl NEB buffer 2, 2 µl 100 mM dATP, 88 µl dH$_2$O and 2 µl Klenow fragment. Lids were sealed with parafilm and the columns were rotated (in such a way that tubes were not inverted) at 37 °C for 30 minutes. The beads were washed twice with FA lysis buffer, and twice with 10 mM Tris-HCl, pH 7.5. NEXTflex barcoded adaptors (BioOscientific) were ligated to DNA fragments by resuspending beads in buffer consisting of 100 µl 1x ligase buffer (50 µl 2x ligase buffer + 50 µl dH$_2$O), 1 µl NEXTflex adaptors (undiluted 0.6 µM stock) and 4 µl quick ligase. Tube lids were sealed with parafilm and the columns were rotated (in such a way that tubes were not inverted) at room temperature for 15 minutes. The beads were then washed twice with FA lysis buffer, once with high-salt FA lysis buffer (500 mM NaCl), once with ChIP wash buffer, once with 1x TE buffer, and another wash was carried out using 300 µl 1x TE, but without rotation. Spin-X-columns were then transferred to new dolphin-nosed tubes. The beads were resuspended in 100 µl elution buffer, and the columns were incubated 65 °C for 10 minutes. This was quickly transferred to a micro-centrifuge before the DNA was eluted by centrifugation at 1500 xg for 1 minute. Eluates and 'input' samples were de-crosslinked by boiling for 10 minutes. Phenol/chloroform extraction (using glycogen as a carrier molecule) or AMPure magnetic beads were used to clean up samples, which were eluted in 11/12 µl of dH$_2$O. Before library amplification, a real-time PCR was run using oligos 'NEXTflex primer 1 & 2' and 0.1-0.4 ng of library, with manual critical threshold set to 0.1. PCR was then used to amplify libraries to the equal concentrations, using a different number of cycles for each sample.

Following PCR, libraries were purified using AMPure magnetic beads and loaded onto an 8 % (*w/v*) acrylamide gel, which was run at 1 hour for 100 V. After running, the gel was stained in 1 % (*v/v*) ethidium bromide for 5 minutes. Using a UV transilluminator to visualise the DNA, a 200-600 bp "smear" was cut out from the gel with care taken to avoid the "adaptor-adaptor" band at 150 bp. The gel slice was shredded as follows; the bottom of a 0.7 ml tube was pierced

using a hot needle, the tube was then placed inside a 1.5 ml tube. The gel slice was placed inside the 0.7 ml tube, and both tubes spun at 17,000 xg for 5 minutes. The 0.7 ml tube was removed and discarded, and 0.4 ml acrylamide extraction buffer was added to the shredded gel. Tube lids were sealed with parafilm, and the tubes rotated overnight at 4 ℃. The next day, the entire mixture (liquid and gel) was transferred to a Spin-X-column, which was spun for 5 minutes at 17,000 xg. DNA was ethanol precipitated. Individual libraries were quantified using a Qubit 2.0 fluorometer (Life Technologies). Values obtained were used to create an equimolar mix of libraries in a single tube (~ 4 nM final concentration), which was sent for sequencing on an Illumina HiSeq platform at the University of Buffalo Next Generation Sequencing Facility).

2.35 Bioinformatics

Raw sequence reads were aligned to the ETEC H10407 genome using the CLC Genomics workbench, and peak calling was carried out using custom python scripts, as described by Singh *et al.,* (2014). Initially, the analysis was performed separately on each replicate. The maximum read density across the genome is first calculated *(M)*, and another parameter, known as variable *(V)*, was assigned a value of 1. In the process of peak calling, *V* was incrementally increased by a value of 1 until it equalled the value of *M*, and at every increase, the genomic coordinates scoring above *V* were registered. Following this, a total value for the number of times each coordinate registered was calculated. A threshold value *(T)*, previously assigned (based on visual observation of the data) was applied, and coordinates with cumulative scores of *V* greater than *T* were called as 'strand peaks'. Separate values of *T* were applied to different data sets to accommodate for differences in total read numbers. This process is performed for both strands. Note that peaks generated by ChIP-seq on the positive and minus strands are off-set due to the nature of alignment, which is affected by experimental parameters. As a result,

pairs of peaks on the positive and minus strand which closely coincide are recorded. The optimum distance between these pairs is defined separately for each replicate.

Next, the peaks on the positive and negative strand are moved towards each other, in 1 bp steps, from a range of 1-200 bp. At each stage, the correlation of the two peaks is calculated using a Pearson correlation. The distance yielding the best correlation (highest R value) was used to calculate the 'optimal strand shift'. This was used to further define pairs of positive and minus strand peaks; peaks on both strands which co-localised within two 'optimal strand shift' distances of each other were recorded. Distances between pairs of positive and minus strand peaks were averaged to calculate peak centres. Finally, datasets from individual replicates were combined. Peaks were called from both replicates only if peaks occurred in both replicates, and peak centres were within 40 bp of each other. Again, distances between peak centres from two replicates are averaged to create final peaks. This analysis works well at low stringency peak calling, and for calling peaks which occur in close succession to one another, as pairs of peaks must conform to the expected shape and peak width, and must be reproducible. The thresholds used in the earlier analysis are adjusted for total read numbers and are applied to the mock replicates, to look for artefactual enrichment. The DNA sequence 50 bp upstream to 50 bp downstream of peaks were submitted to MEME (Bailey *et al.,* 2009) to identify CRP binding sites, and to generate sequence motifs. Sequences around peaks were submitted to BLAST searches to search for homologues in *Escherichia coli* K-12 MG1655. Distances from transcription start sites were calculated (using mapped transcription start sites from Kim *et al.,* (2012), and Cho *et al.,* (2014)). To identify the most frequent distances between CRP peaks and transcription start sites, distances from +200 bp upstream and -100 bp downstream of transcription start sites were selected, and organised into groups separated by 5 bp. The data are available on the EBI database (EBI submission: E-MTAB-2917).

Circular plots were generated to show genome-wide σ^{70}, H-NS and CRP binding using DNAPlotter software (Carver *et al.,* 2009). Linear ChIP-seq plots were generated using SignalMap software (Nimblegen).

To investigate the overlap between σ^{70}, H-NS and CRP peaks, peak locations were aligned manually in an excel spreadsheet. Note that for H-NS, this is not straightforward, since binding is filamentous and peak calling preferentially identifies the nucleation site as opposed to the whole binding tract. Hence, some comparisons involving H-NS have been made over greater distances since H-NS filaments have been found to extend, on average, over 1 kb in *E.coli* (Kahramanoglou *et al.,* 2011).

In silico prediction of CRP sites in both intergenic and intragenic regions was done using PREDetector software (Hiard *et al.*, 2007), which generates a PWM from a list of given sites. In this analysis, we used a list of 68 sites previously identified by Grainger *et al.,* 2005. Predicted sites are scored based on their similarity to the consensus; higher scoring sites better resemble the consensus. Genbank files were imported into the program for plasmids p948 and p666. A cut-off of 7 was used for all screens, sites scoring below this were not used.

Chapter 3

Identification of CRP binding sites across the H10407

genome

3.1 ChIP-seq analysis of CRP, H-NS and σ^{70} binding across the ETEC H10407 genome

In this chapter, chromatin immunoprecipation coupled with next generation sequencing (ChIP-seq) was used to map CRP, H-NS and RNAP σ^{70} subunit binding across the genome of ETEC H10407. For immunoprecipitations we used a i) *crp*FLAG$_3$ H10407 derivative and mouse mono-clonal anti-FLAG, ii) rabbit polyclonal anti-H-NS or iii) mouse mono-clonal anti-σ^{70} antibody.

The H10407 derivative carrying the *crp*FLAG$_3$ allele was generated using a recombineering method known as FRUIT (Flexible Recombineering Using Integration of *thyA*) (Stringer *et al.,* 2012). Strains encoding either an N- or C- terminally tagged CRP were generated. However, the strain harbouring the N-terminal FLAG$_3$ fusion was defective for growth and was not studied further. Conversely, the strain encoding the C-terminally tagged CRP had a CRP^{+} phenotype (appeared purple of MacConkey maltose agar (Figure 3.1)) and was used for subsequent ChIP-seq experiments.

ETEC H10407 or the *crp*FLAG$_3$ derivative were grown to mid-log phase at 37 $^{\circ}$C in M9 minimal media supplemented with 1 % (*w/v*) fructose before being fixed with formaldehyde and harvested. Immunoprecipitations were done with anti-FLAG, anti-H-NS or anti-σ^{70} in duplicate alongside mock immunoprecipitations (i.e. with no antibody) as a control. After immunoprecipitation, the resulting DNA sample was used for Illumina HiSeq library generation. After sequencing reads were aligned to the H10407 genome and binding peaks were called automatically using custom python scripts (Singh *et al.,* 2014). The CRP binding motifs within peaks were found using MEME (Bailey *et al.,* 2009).

3.1.1 CRP binding sites across the ETEC H10407 chromosome

We identified 111, 364 and 1089 peaks for binding of CRP, H-NS and σ^{70} respectively. The binding profiles of CRP (orange), H-NS (green), and σ^{70} (blue) are plotted against the features of the ETEC chromosome (Figure 3.2A) or plasmids (Figure 3.2B).

The 111 CRP binding peaks, and associated DNA motifs, are shown in Table 3.1. Of the 111 CRP sites, 19 were within 100 bp of a peak for σ^{70} binding and 5 overlapped with loci at which H-NS was bound. Similarly, 30 of the 364 H-NS bound loci were also bound by σ^{70}. A Venn diagram illustrating the overlap in CRP, H-NS and σ^{70} binding is shown in Figure 3.3. Some examples of overlap are shown in Figure 3.4.
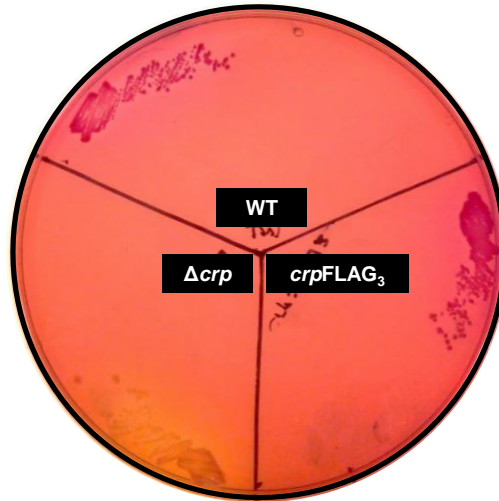
Figure 3.1 Image of *E. coli crp*FLAG₃ growing alongside wildtype cells and Δ*crp* cells on solid MacConkey maltose media
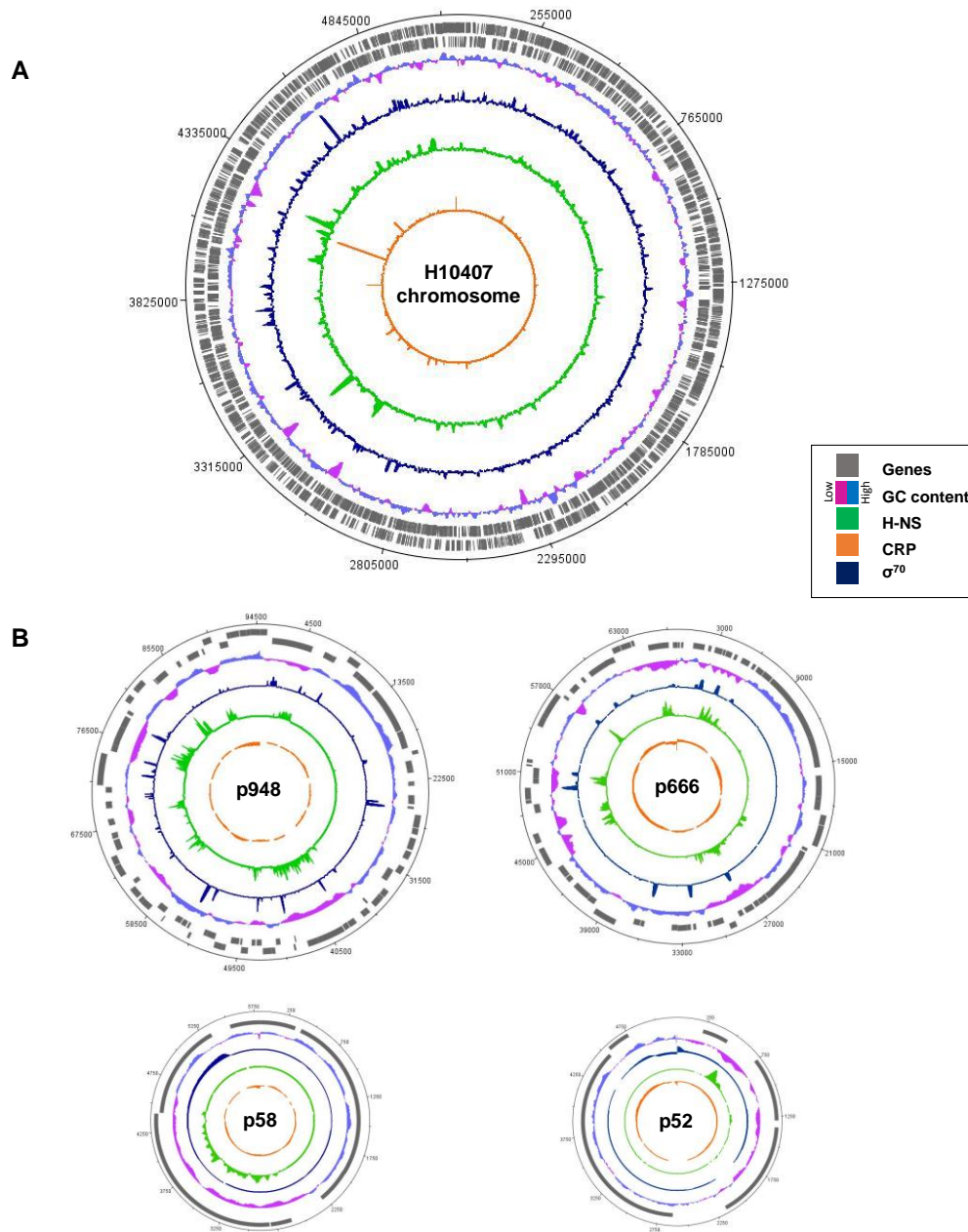
Figure 3.2 CRP, H-NS and σ⁷⁰ binding across the chromosome (A) and virulence

plasmids (B) of ETEC H10407 as determined by ChIP-seq analysis

In both sets of plots (A and B), genes are indicated in by grey blocks in the outermost two tracks. The GC content is depicted by the third track, where regions highlighted light blue contain a GC content which is higher than average, and purple highlighted regions indicate regions containing a lower than average GC content. The green tracks shows binding by H-NS, the orange tracks shows binding by CRP, and dark blue tracks show binding by σ⁷⁰ as determined by ChIP-seq analysis. Plots were made using DNAPlotter software (Carver *et al.,* 2009). Note that for σ⁷⁰ plots on p948 and p666, reads have been capped at 4500 reads and 10000 reads respectively, to allow visualisation of smaller peaks.

Table 3.1 CRP targets identified by ChIP-seq analysis

Peak Centre [a]	Binding Site(s) [b]	Gene(s) [d]	K-12 Homologues [e]
45284	**TGTGA**TTGGTA**TCACA**	ETEC_0040	caiT
92014	A**GTGA**TGGATG**TCAC**G	(ETEC_0078)	(cra)
176905	AG**C**GTTCCACG**TCACA**	(ETEC_0150)	(hemL)
408885	**TGTGA**TCTCTC**TCGCA**	ETEC_0385/ETEC_0386	yahN/yahO
468009	AG**G**GA**C**TCGCG**TCACA**	ETEC_0443	aroM
492973	ATC**GA**TTGCGT**TCAC**G	ETEC_0464	**tsx**
540805	**TGTGA**TCTTTA**TCACA**	ETEC_0511	maa
574230	GA**TGA**CGACGA**TCACA**	(ETEC_0538)	(ybaT)
683187	A**GTGA**TCGAGT**TAACA**	ETEC_0628	**cstA**
697540	A**GTGA**TTTGCG**TCACA**	ETEC_0639	rnk
739223	C**GTTA**CCCTTG**TCGCA**	ETEC_0680	rihA
941002	**TGTGA**TGAGTA**TCAC**G	ETEC_0869	ybiJ
958866	**TGTGT**ACGAAAT**CACA**	ETEC_0886/ETEC_0887	**ybiS/ybiT**
1128472	n.d.	(ETEC_1030)	(yccS)
1205350	A**GTGA**TGTAGA**TCACA** TG**AGA**TCGAGCA**CACA**	ETEC_1101	ycgZ
1263558	T**TTGA**CGGCTA**TCAC**G	ETEC_1166	**ptsG**
1274886	**TGTGA**TCTGGA**TCACA**	ETEC_1176/ETEC_1177	**ycfQ/bhsA**
1301786	GA**TGA**TCCGCA**TCACA**	(ETEC_1206)/ETEC_1207	ETEC-specific/ETEC-specific
1348166	AT**TGA**ACAGGA**TCACA**	(ETEC_1259)/ETEC_1260	(rluE)/icd
1376374	GG**TGA**GCTGGC**TCACA**	ETEC_1292/ETEC_1293	**ycgB/dadA**
1388620	A**GTGA**GCCAGT**TAACA**	(ETEC_1303)	(dhaI)
1541732	C**GTGA**ACCGGG**TCACA**	ETEC_1443/ETEC_1444	ycjZ/mppA
1567885	G**TTAA**GTAAAA**TCACA**	ETEC_1462/ETEC_1463	**paaZ/paaA**
1701402	**TGTGA**TGGATG**TCACT**	ETEC_1568	ydeN
1767726	**TGTGA**TTAACAG**CACA**	ETEC_1628	**mlc**
1777143	**TGTGA**TCTAGCG**CCAA**	ETEC_1637	pntA
1811426	C**GTGA**TCAAGA**TCAC**G	(ETEC_1668A)	(ETEC specific)
1859265	AT**TGA**GCGGGA**TCACA**	(ETEC_1713)	(sufS)
1887513	A**GTGA**TGCGCA**TCAC**G TG**CGA**GGTGTG**TCACA**	ETEC_1737	aroH
2126754	**TGTGG**CGTGCA**TCACA**	n.a.	n.a.
2201816	GG**TGA**CGCGCG**TCACA**	ETEC_2057	yedP
2210222	C**GTGA**TCTCGC**TCACA**	ETEC_2065/ETEC_2066	yedR/ETEC-specific
2458348	**TGTGA**TCTGAAT**CTCA** TG**CGA**TGCGTCG**CGCA**	ETEC_2278	**cdd**
2492757	AT**TGA**TCGCCC**TCACA**	ETEC_2309	yeiQ
2555083	C**GTGA**CCAAAG**TCTCA**	(ETEC_2360)	(yfaQ)
2729713	T**TTGA**AGCTTG**TCACA**	ETEC_2510/ETEC_2511	mntH/nupC
2735124	A**GTTA**TTCATG**TCAC**G	ETEC_2514	yfeC
2795423	**TGTGA**GCCATGA**CACA**	(ETEC_2572)/ETEC_2573	**(aegA)/narQ**
2810983	C**GTGA**TCAAGA**TCACA**	ETEC_2586	**hyfA**
2887131	T**TTGA**TCTCGC**TCACA**	(ETEC_2666)/ETEC_2665	**(xseA)/guaB**
3012645	**TGTGA**TCCCCACA**ACA**	(ETEC_2793)	(**ung**)
3048307	T**TTGA**CGAGCA**TCACC**	(ETEC_2822)	(emrB)
3132920	GG**TGA**CCGGTT**TCACA**	ETEC_2905/ETEC_2906	ascG/ascF
3161660	**TGTGA**CCGTGG**TCGCA**	(ETEC_2933)	(**nlpD**)
3184337	C**GTGA**TGCGTG**TAACA**	(ETEC_2956)/ETEC_2955	(cysI)/cysH
3196088	**TGTGA**TTACGA**TCACA**	ETEC_2966/ETEC_2967	ygcW/yqcE
3223792	A**GTGA**TCTTGA**TCTCA** A**GTTA**TGTATC**TATCA**	ETEC_2986	sdaC
3234980	TG**CGA**TCGTTA**TCACA**	(ETEC_2994)/ETEC_2995	(fucU)/fucR
3265047	**TGTGA**CCTGGG**TCAC**G	ETEC_3017	rppH
3324543	**TGTG**GGCTACG**TAACA**	(ETEC_3075)	(ydhD)
3361162	n.d.	ETEC_3105	**serA**
3368992	T**TTGA**TGCACCG**CACA**	(ETEC_3113)	(ygfI)
3382158	**TGTGA**TCTACAA**CAC**G	ETEC_3126	cmtB
3390811	**TGTGA**TTTGCTT**CACA**	ETEC_3133	**galP**
3408173	**TGTGA**TGTGGAT**AACA**	ETEC_3154	**nupG**
3442697	**TGTGA**TGATTG**TCGCA**	ETEC_3186	ETEC-specific
3558573	A**GTGA**TTTGGC**TCACA**	ETEC_3291	ygiS
3580767	A**GTGA**CTTGCA**TCAC**G	(ETEC_3318)	(yqiH)
3635301	AT**TGA**TCTAAC**TCAC**G	ETEC_3362	**uxaC**
3642302	CT**TGA**AGTGGG**TCACA**	(ETEC_3372)	(yqjG)
3665634	**TGTGA**TCAATG**TCAAT** **TGTG**CTTTAGCG**CGCA**	ETEC_3393/ETEC_3394	garP/garD
3721308	GG**TGA**TTGATG**TCACC**	(ETEC_3446)	(greA)
3785700	C**GTG**GGTCGCA**TCACA**	(ETEC_3510)	(mreC)
3878729	GG**TGA**TTTTGA**TCAC**G	ETEC_3614/ETEC_3615	ppiA/tsgA
3908574	GG**TGA**TCGCGC**TCACA**	(ETEC_3645)	(**hofM**)
3918861	**TGTGA**GTGGAA**TCGCA**	ETEC_3652/ETEC_3653	yhgE/pck

101

3986400	CGTGATTTTATCCACA	ETEC_3707	***rpoH***
4105040	AGTAAGGCAAGTCCCT	n.a.	n.a.
4111116	TGTGACGGGGCTAACA	(ETEC_3806)	(wecH)
4153055	TGTGATCTGAATCACA TGTGATCTACAGCATG	ETEC_3840	***yibI***
4153191	TGTGATTGATATCACA TGTGATGAACGTCACG	ETEC_3841	_mtlA_
4158433	n.d.	ETEC_3846	lldP
4196869	TGCAATCGATATCACA	ETEC_3886	dinD
<u>4251326</u>	CTTACTCCTGCTCACA	ETEC_3938	ETEC specific
4266125	GGTGATGGCATCCGCG	(ETEC_3956)	(nepI)
4290730	GGTGAGCAAAACCACG	(ETEC_3979)	(yidR)
4322430	ATTGACCTGAGTCACA	(ETEC_4010)	(yieL)
4340544	CTTGACCACGGTCAGA	(ETEC_4025)/ETEC_4024	(atpA)/atpG
4344649	TGTGATCTGAAGCACG	ETEC_4030	_atpI_
4373517	TGTAATGCTGGTAACA	(ETEC_4051)	(ilvG)
4402013	CGTGCTGCATATCACG	(ETEC_4077)	(rffM)
4412999	CGTGATCAATTTAACA	ETEC_4085/ETEC_4085	_hemC/cyaA_
4438352	GGTGATGAGTATCACG TGTGATTTGAATCACT	ETEC_4107/ETEC_4108	_ysgA/udp_
4508745	TGTGATATTTGTCACA	(ETEC_4165)/ETEC_4164	(fdhD)/_fdoG_
4517442	CGTGATCGCTGTCCCA	(ETEC_4173)	(rhaA)
4564670	TGCGATCCGCCTCATA	ETEC_4216/ETEC_4217	ptsA/frwC
4668870	TGTAACAGAGATCACA	ETEC_4289/ETEC_4290	***malE/malK***
4725047	TGTGCGGATGATCACA	n.a.	n.a.
4731402	TGTGATCTTGCGCATA	(ETEC_4365)	(aphA)
4761367	CGTGATGGCTGTCACG	ETEC_4389	fdhF
<u>4846352</u>	n.d.	ETEC_4464	ETEC-specific
<u>4848117</u>	CGTGAGTTCTGTCACA	n.a.	n.a.
4863253	TTTGATCAACATCGCA	(ETEC_4478)	(ETEC-specific)
4873926	GGTGATCTATTTCACA	ETEC_4486/ETEC_4487	aspA/fxsA
4930149	TGTGATGAACTTCAAA	ETEC_4545/ETEC_4546	yjfY/rpsF
<u>4940903</u>	TGTGATCACTATCGCA	ETEC_4557/ETEC_4558	ETEC-specific/ytfA
4993073	TGTGACTGGTATCTCG	(ETEC_4604)	(valS)
5002854	TGTAACCTTTGTCACA	ETEC_4610/tRNA-Leu	yjgB/tRNA-Leu
5030724	TGCGATGAATGTCACA	ETEC_4633/ETEC_4634	_gntP/uxuA_
5129400	CGTACCGTCGGTCACA	(ETEC_4736)	(yjjI)
5129944	TGTGATGTATATCGAA	ETEC_4736/ETEC_4737	yjjI/deoC

^a Peak centres are annotated according to the chromosome of H10407. Underlined co-ordinates share no homology to *E. coli* K-12.

^b Binding sites found within an identified peak, using MEME software. Bases conforming to the consensus CRP binding site are shown in bold. Binding sites unidentified by MEME as listed as 'not determined' (n.d.).

^c Values represent the fold of enrichment above the background ('Fold above threshold').

^d Genes located within 300 bp downstream of an identified CRP site. Brackets indicate where the CRP site is intragenic. CRP sites not associated with a gene (further than 300 bp upstream) are listed a 'not associated' (n.a.).

^e K-12 homologues to the ETEC gene indicated. ETEC genes without a listed homologue are listed as 'not associated' (n.a.). Genes underlined match CRP targets previously by Grainger *et al*., 2005, and genes in bold are listed as CRP-dependent in the Ecocyc database. Where genes are listed in brackets, the CRP site is intragenic.
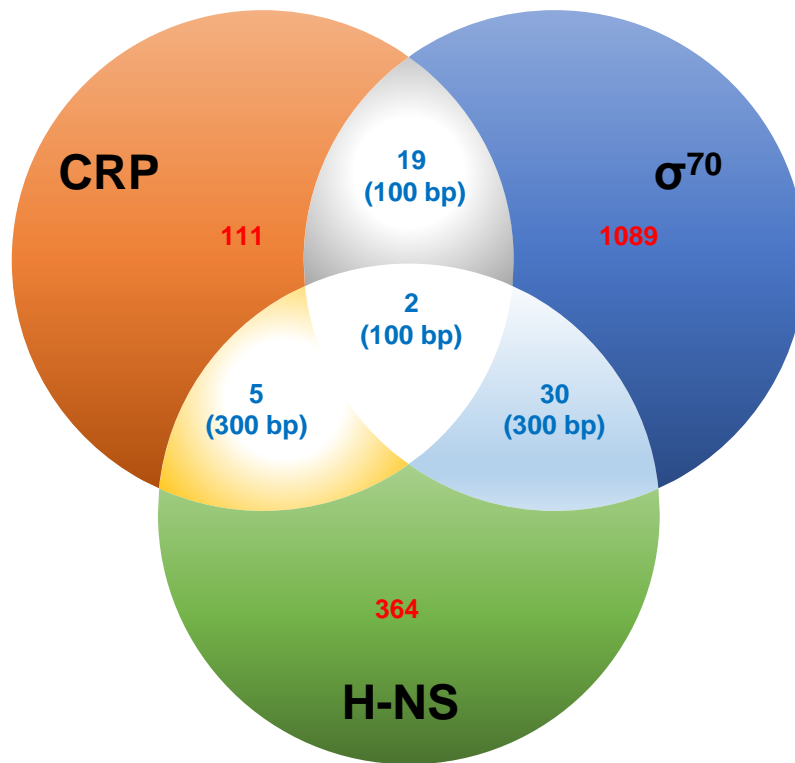
Figure 3.3 Overlap of CRP, H-NS and σ^{70} binding sites identified in *E. coli* H10407

The total number of peaks for each protein is shown in red numbers. Overlap between peaks is shown in blue numbers, and the distance within which these peaks occur is displayed below in parenthesis.
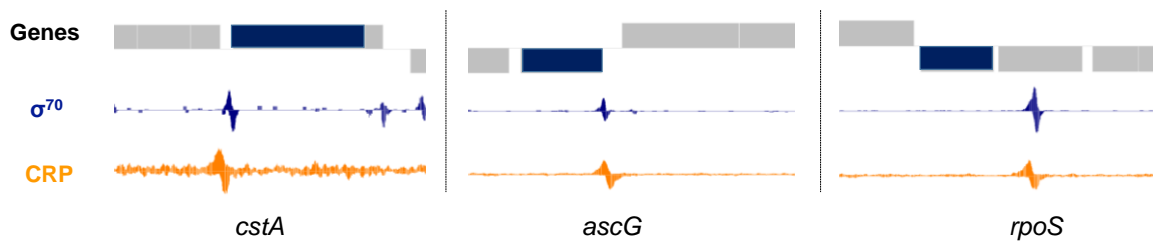
Figure 3.4A Overlap of σ⁷⁰ and CRP binding sites at selected chromosomal loci

Genes are shown in grey, the gene shown in blue is defined below each panel. σ^{70} binding is shown in dark blue, and CRP binding is shown in orange below. Reads per peak are as follows; *cstA*: σ^{70}: 527, CRP: 715. *ascG*: σ^{70}: 767, CRP: 2321. *rpoS*: σ^{70}: 7438, CRP: 2498.
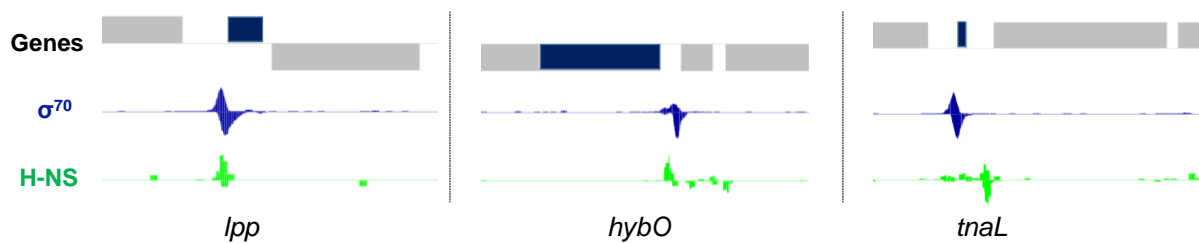
Figure 3.4B Overlap of σ⁷⁰ and H-NS binding sites at selected chromosomal loci

Genes are shown in grey, the gene shown in blue is defined below each panel. σ^{70} binding is shown in dark blue, and H-NS binding is shown in green below. Reads per peak are as follows; *lpp*: σ^{70}: 2644, H-NS: 366. *hybO*: σ^{70}: 3868, H-NS: 275. *tnaL*: σ^{70}: 3799, H-NS: 428.
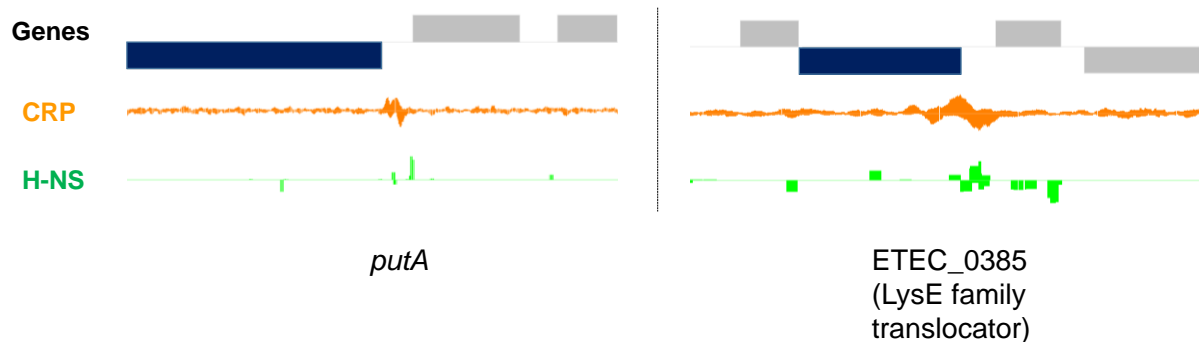
Figure 3.4C Overlap of CRP and H-NS binding sites at selected chromosomal loci

Genes are shown in grey, the gene shown in blue is defined below each panel. CRP binding is shown in orange, and H-NS binding is shown in green below. Reads per peak are as follows; *putA*: CRP: 859, H-NS: 171. ETEC_0385: CRP: 700, H-NS: 175.
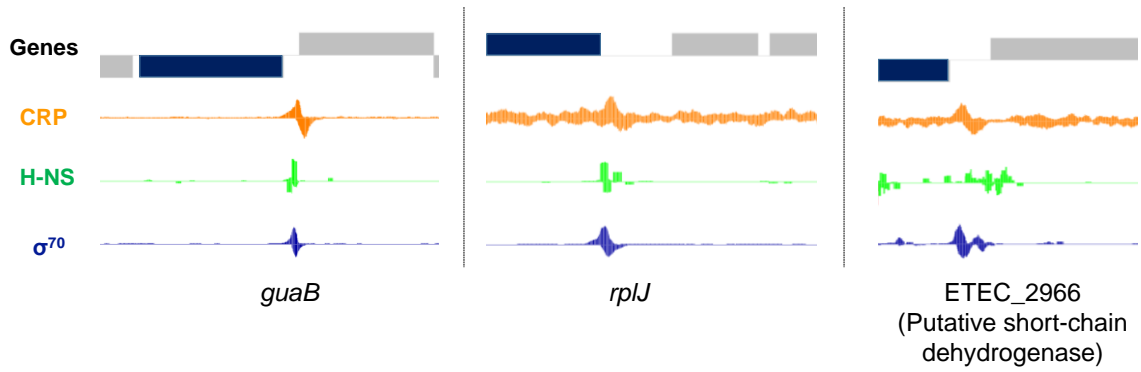
Figure 3.4D Overlap of CRP, H-NS, and σ⁷⁰ binding sites at selected chromosomal loci

Genes are shown in grey, the gene shown in blue is defined below each panel. CRP is shown in orange, H-NS binding is shown in green, and σ^{70} binding is shown in dark blue, below. Reads per peak are as follows; *guaB*: CRP: 3918, H-NS: 106, σ^{70}: 21413. *rplJ*: CRP: 741, H-NS: 253, σ^{70}: 29108. ETEC_2966: σ^{70}: 613, H-NS: 213, CRP: 678.

3.1.2 Properties of DNA loci bound by CRP

We next analysed the DNA sequence and genomic context of peaks for CRP binding. The CRP ChIP-seq signal correlated weakly (correlation coefficient= 0.139) with the number of bases in the CRP site that matched the known consensus (Figure 3.5A). Ninety-three percent of the 111 CRP binding peaks fell within or adjacent to genes that are common to both *E. coli* K-12 MG1655 and H10407 (Figure 3.5B). The CRP motifs associated with binding peaks were most frequently found 92.5 and 40.5 bp upstream of known transcription start sites, in agreement with CRP regulatory paradigms (Busby and Ebright, 1999). The majority of CRP sites were in intergenic regions (66 %) with the remaining 34 % being located within genes. Twenty two of the 111 CRP binding sites identified are nearby genes listed as CRP regulated in *E. coli* K-12 on the Ecocyc and Regulon databases (Keseler *et al.,* 2005, Huerta *et al.,* 1998). Additionally, 12 of the *E. coli* K-12 CRP sites previously identified using ChIP-chip were identified in H10407, and in total, 46 out of the 111 targets identified in H10407 have been identified in K-12 by previous studies (Figure 3.5B) (Grainger *et al*., 2005, Shimada *et al.,* 2011). However, 8 CRP sites were located upstream of, or within, ETEC-specific genes. The function of genes targeted by CRP in ETEC H10407 and *E. coli* K-12 is similar (Figure 3.5D). Amongst the 111 chromosomal peaks for CRP binding were many well studied CRP regulated promoters (Aiba *et al.,* 1985, Busby *et al.*, 1983, Kimata *et al.,* 1997, Shimada *et al.,* 2011). A selection of these targets is shown in Figure 3.6. For example, CRP is known to bind to five sites upstream of the mannitol operon (*mtlADR*) (Ramseier and Saier Jr, 1995).
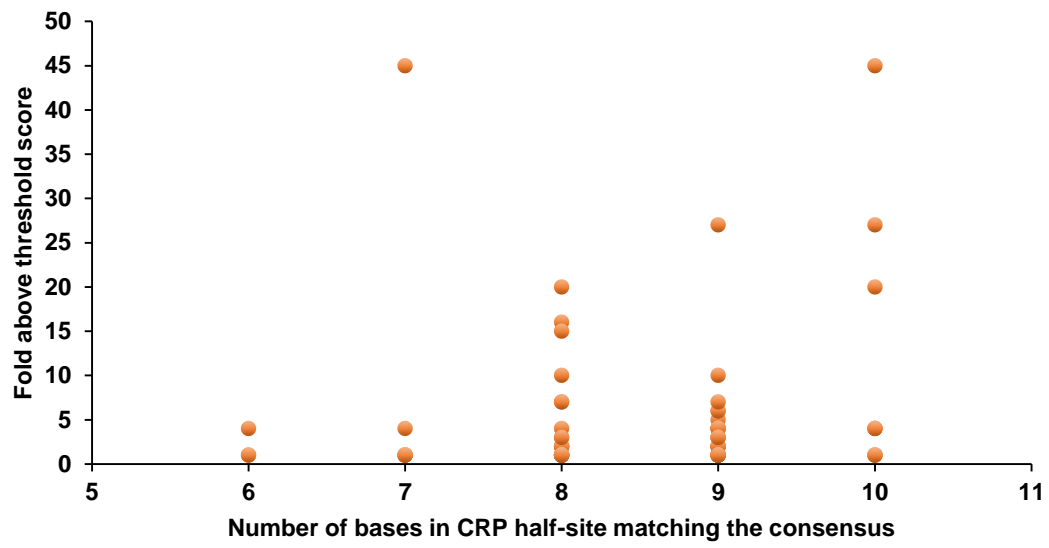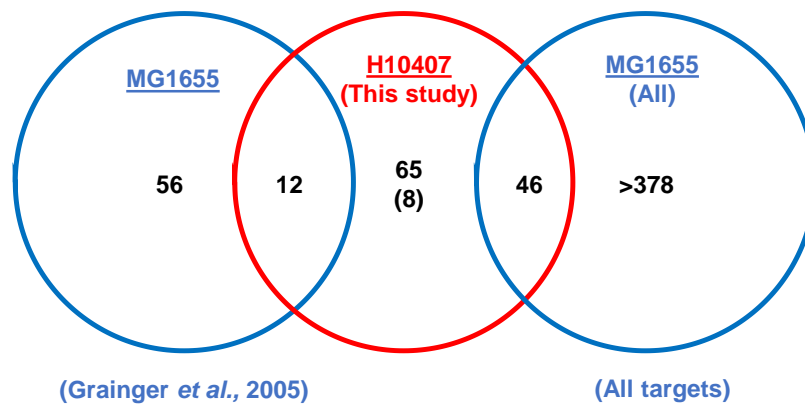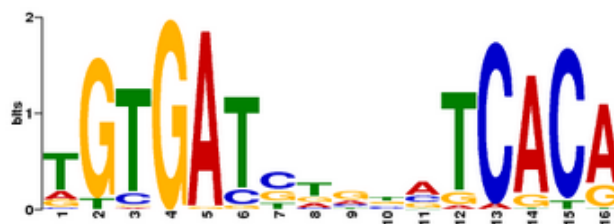
D
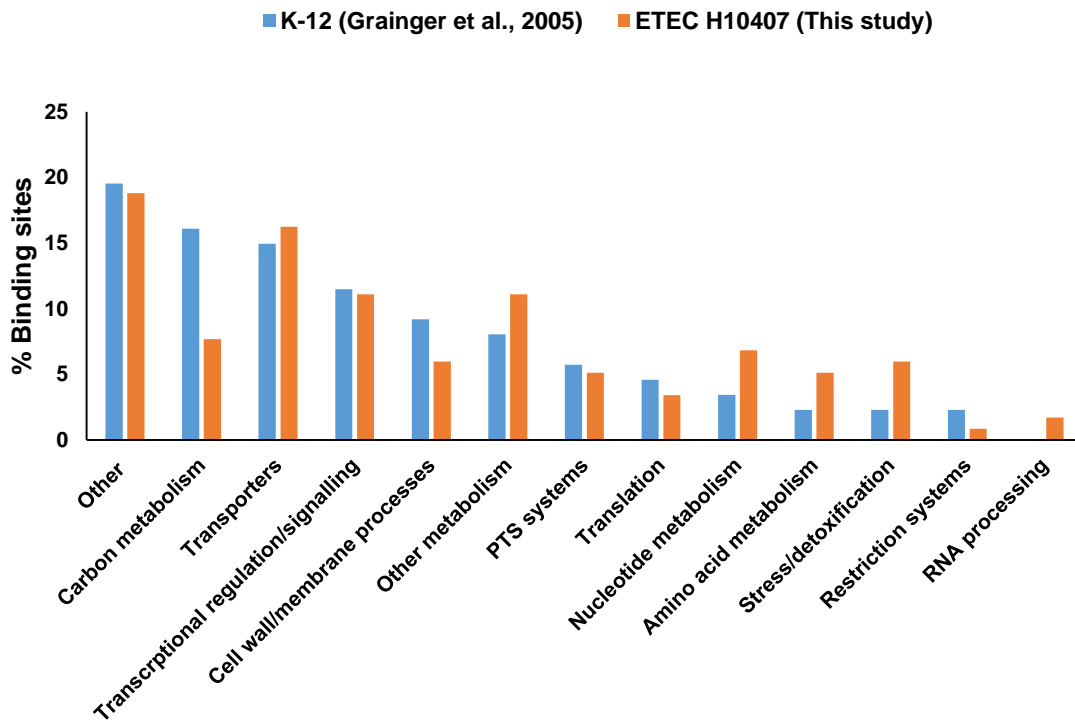
Figure 3.5 Analysis of CRP targets found by ChIP-seq on the ETEC strain H10407

chromosome

A) Correlation of FAT (Fold Above Threshold) score with half-site match to the consensus (correlation coefficient= 0.139).
B) Overlap of CRP targets found in H10407 with those of K-12, found by Grainger *et al.,* (2005) by ChIP-chip, and by all studies. The total number of CRP targets was obtained from Shimada *et al.,* (2010), and uses data from RegulonBD, Ecocyc, Grainger *et al.,* (2005), and Shimada *et al.,* (2010). Numbers in parenthesis indicate targets in or upstream of ETEC-specific genes. All other targets are upstream or within K-12 genes or homologs.
C) Consensus motif for CRP binding associated with 96 % of identified CRP targets, shown as a sequence logo, generated using MEME software (Bailey *et al.*, 2009).
D) Gene categories in which CRP targets fall as found by ChIP-chip or ChIP-seq on the chromosome of *E. coli* K-12 (Grainger *et al.,* 2005) and ETEC H10407 respectively. Genes have are broadly sorted based on general function. Genes shown contain a CRP binding site within the coding sequence or in a nearby intergenic region. Genes characterised as "other" include hypothetical and predicted proteins of unknown function.
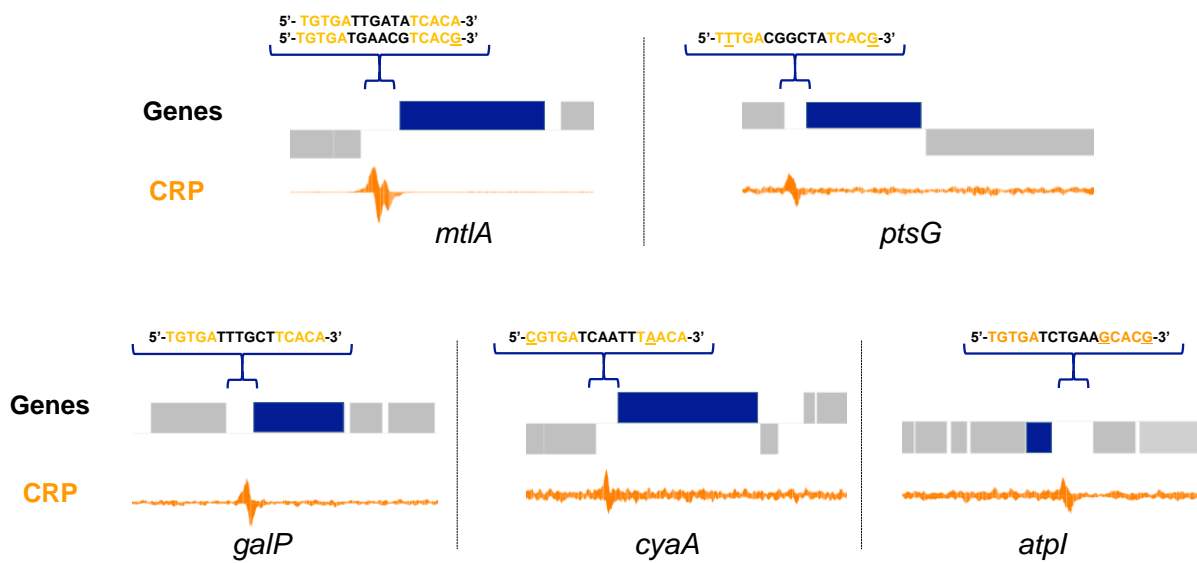
Figure 3.6 *In vivo* binding of CRP at selected chromosomal loci as determined by ChIP-seq analysis

Genes are depicted by gray blocks, the gene previously shown to be regulated by CRP are depicted by a blue block. CRP binding is shown by orange peaks. CRP binding motifs identified by MEME are shown above the tracks, CRP half-sites are shown in orange, and are mismatches to the consensus are underlined. Reads per peak are as follows; *mtlA*: 13681, *ptsG*: 822, *galP*: 2007, *cyaA*: 1223, *atpI*: 1137.

3.1.3 Properties of chromosomal DNA loci bound by H-NS

Of the 364 H-NS peaks identified, 28 % of these were within genes. H-NS binding was focused to AT-rich regions of the genome as has been previously observed (Navarre *et al.,* 2007). We observed binding at previously identified H-NS targets. Some examples of H-NS binding are shown in Figure 3.7 (Dole *et al.,* 2004, Lucht *et al.,* 1994, Olsén *et al.,* 1993, Pul *et al.,* 2010, Shin *et al.,* 2005).

3.1.4 Properties of chromosomal DNA loci bound by σ^{70}

1089 σ^{70} peaks were identified, 85 % of which localised to intergenic regions. Figure 3.8 shows some examples of σ^{70} binding. We observed σ^{70} peaks upstream of genes important for rapid growth including *rRNA* and *tRNA* genes.
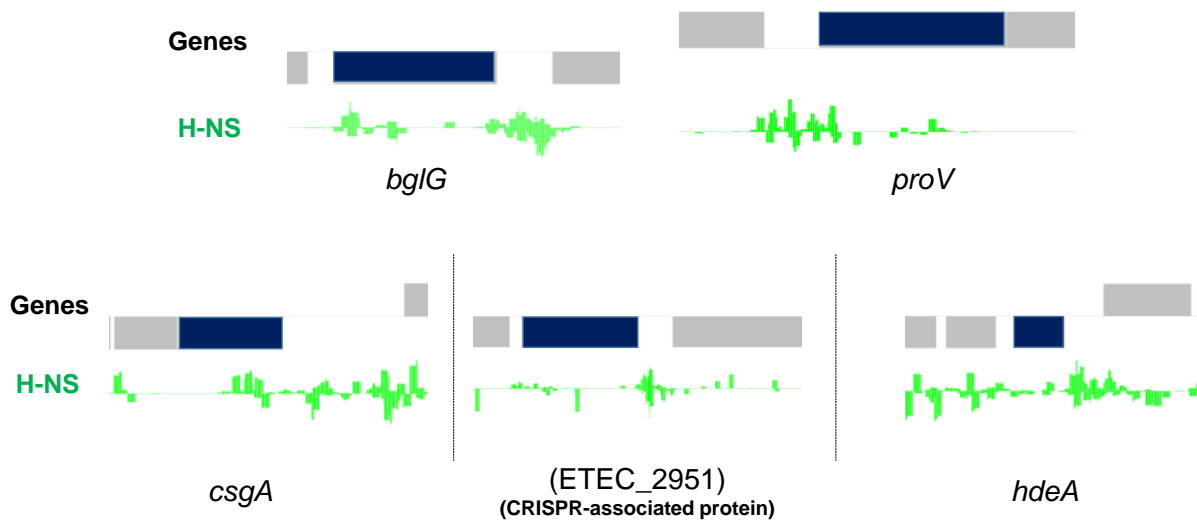
Figure 3.7 *In vivo* binding of H-NS at selected chromosomal loci as determined by ChIP-

seq analysis

Genes are shown in grey, the gene shown in blue is defined below each panel. H-NS binding is shown in green below. Reads per peak are as follows; *bglG*: 336, *proV*: 223, *csgA*: 169, ETEC_2951: 250, *hdeA*: 383.
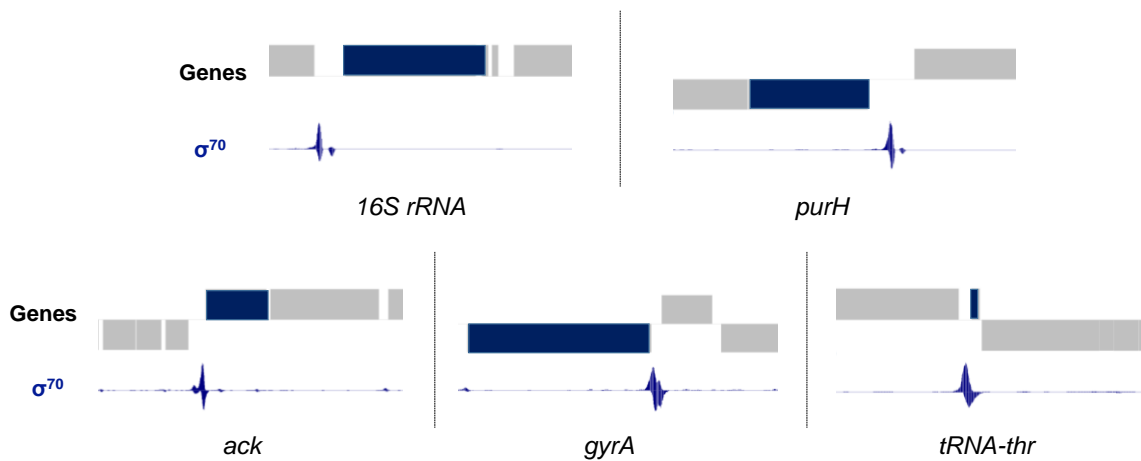
Figure 3.8 *In vivo* binding of σ⁷⁰ at selected chromosomal loci as determined by ChIP-seq analysis

Genes are shown in grey, the gene shown in blue is defined below each panel. σ⁷⁰ binding is shown in dark blue below. Reads per peak are as follows; *16S rRNA*: 4919, *purH*: 6600, *ack*: 5375, *gyrA*: 2868, *tRNA-thr*: 7490.

3.2 CRP binding across plasmids p948 and p666 of H10407

Unexpectedly, CRP binding was restricted to the chromosome of ETEC H10407, and no CRP binding was detected on any of the four plasmids (Figure 3.2B). This was surprising since the heat-labile toxin promoter is thought to be bound by three CRP dimers (Bodero and Munson, 2009). Conversely, H-NS binding was observed across all four plasmids. To further investigate CRP binding across the plasmids p948 and p666, which carry the 'classic' virulence factors, we used a regulon prediction program (PREDetector) that generates a PWM from a list of known transcription factor targets and then searches for matches to the PWM (Hiard *et al.*, 2007). Each match is given a "score" and the user sets an arbitrary cut-off below which predicted sites are discarded. Sites with high scores are a better match to the PWM. Table 3.2 is a list of all sites identified from the bioinformatics screening. Binding sites were grouped into bins with other similarly scored sites.

3.2.1 *In vitro* characterisation of predicted sites on H10407 virulence plasmids

Following the bioinformatic screen the affinity of CRP for a selection of sites was investigated using Electrophoretic Mobility Shift Assays (EMSAs). The 7 sites with the highest scores, and the highest scoring site from each of the three lowest scoring bins, were selected. All selected sites are shown in red in Table 3.2. For simplicity, the sites were arbitrarily named 1-10. Figure 3.9A is a cartoon showing the position of each predicted CRP binding site relative to genes located on plasmids p948 and p666. For the EMSA assays ~120 bp DNA fragments, corresponding to each target, were generated. The sequences are shown in Figure 3.9B.

Table 3.2 CRP targets predicted by PREDetector software

Bin[a]	Predicted CRP site[b]	Score[c]	Arbitrary name	Target[d]	Bound *in vitro* (n.d. = not done)[e]
13 to 14	ttt**tgtga**aattaa**tcaca**aaa	13.78	(1)	ETEC_p666_0090	Yes
	taaa**gtga**taaaaa**tcaca**taa	13.77	(2)	ETEC_p948_0870c, *aatC*	Yes
12 to 13	ttt**tgtga**tgtgta**tca**t**a**cta	12.57	(3)	ETEC_p948_0360c, *estA2*	Yes
11 to 12	ttt**ta**t**ga**aatcaa**tcaca**aaa	11.82	(4)	ETEC_p666_0110	No
	ata**tt**t**ga**acgaag**tca**a**a**ttt	11.1	(5)	ETEC_p666_0360, *traJ*	Yes
10 to 11	taa**tgt**a**a**ataaa**tt**a**a**aata	10.5	(6)	ETEC_p948_0510	No
	ttt**tgtga**gttgca**tca**tgtta	10.41	(7)	ETEC_p666_0750, *estA1*	Yes
9 to 10	ata**tgtga**tattaa**ta**g**ca**caa	9.86	(8)	ETEC_p948_0410, *cfaC*	No
	tta**tt**t**ga**agcaaa**tca**actt**t**	9.85		ETEC_p948_0490, *traJ*	n.d.
	taa**tgtga**ttttga**ta**a**t**gaaa	9.62		ETEC_p948_0230c, *repA2*	n.d.
	aac**tt**a**a**taaacac**caca**tta	9.56		ETEC_p948_0830	n.d.
	att**tt**t**ga**tatgtc**tga**catta	9.53		ETEC_p948_0700	n.d.
	att**tttt**atatata**tt**g**catta**	9.5		ETEC_p948_0700	n.d.
	aat**ta**t**ga**tgttgta**cata**taa	9.36		ETEC_p58_0005	n.d.
	aag**tg**c**ga**atctat**ta**a**ca**ata	9.08		ETEC_p948_0800	n.d.
	aaa**ta**a**ga**tacaca**ta**a**a**aaaa	9		ETEC_p52_0001	n.d.
	att**tgtga**atcact**tcac**gacc	9		ETEC_p52_0006, *rop*	n.d.
8 to 9	ttt**tgtg**gagtggg**tt**a**a**atta	8.89	(9)	ETEC_p948_0510	No
	aaac**gt**t**a**ctacgt**tcac**gttt	8.73		ETEC_p666_0880, *relE*	n.d.
	aaaat**tg**taatagata**a**a**a**aaa	8.63		ETEC_p948_0910, *cexE*	n.d.
	atag**gt**a**a**atctgt**tca**aaaaa	8.62		ETEC_p948_0790	n.d.
	tgca**gt**a**a**ttaacg**tcaca**ttt	8.59		ETEC_p948_0020, *eatA*	n.d.
	aaa**tg**c**ga**tgaagc**ta**a**ta**aat	8.46		ETEC_p666_0560, *stbB*	n.d.
	aaac**ggga**agctaa**tc**g**ca**taa	8.43		ETEC_p948_0530	n.d.
	aatct**tga**aaaata**ta**a**cc**aaa	8.28		ETEC_p948_0870	n.d.
	tac**tgtga**attaat**ta**a**c**gttg	8.28		ETEC_p948_0120, *etpB*	n.d.
	aat**tgtg**gaagaaa**tt**a**ca**atg	8.25		40522 -> 40543	n.d.
	ttt**tact**atcaata**tca**t**a**tta	8.24		ETEC_p666_0750, *sta1*	n.d.
	ataat**tga**tataca**ta**a**a**ctta	8.16		ETEC_p948_0990	n.d.
	ttt**tc**t**ga**atgaga**tc**g**c**cttt	8.14		ETEC_p58_0001	n.d.
	caa**tgt**t**a**tttata**tca**ttaaa	8.12		ETEC_p948_0990	n.d.
	taa**tat**c**a**aataaag**caca**act	8.08		ETEC_p948_0120, *etpB*	n.d.
	aaa**tgt**attccgtc**tcaca**tgt	8.04		ETEC_p948_0700	n.d.
	ttt**tgtga**ttttcta**c**tatatt	8.04		ETEC_p948_0890c	n.d.
7 to 8	att**tttt**a**a**caact**tt**a**t**attt	7.98	(10)	ETEC_p948_0890c	No

aaa**ttt**t**a**tctaaaga**aaa**taa	7.9		42550 -> 42571	n.d.
aaagaca**a**cataaa**ta**a**ca**ttt	7.9		ETEC_p948_0020c, *eatA*	n.d.
aatc**gtga**cgcaag**ttac**gaaa	7.85		ETEC_p948_0500c, *traM*	n.d.
ttacact**a**taaaca**tcaca**gtt	7.83		42327 -> 42348	n.d.
taa**ta**tttagaaca**tca**t**a**ata	7.82		44227 -> 44248	n.d.
caaat**tga**accaga**tca**a**a**atc	7.77		ETEC_p948_0490c, *traJ*	n.d.
att**tctga**tacattaa**ac**gtat	7.65		43228 -> 43249	n.d.
aaa**tg**aattttcaa**tca**a**a**tta	7.65		ETEC_p666_0870c	n.d.
ttt**ta**t**t**attccata**caca**taa	7.64		ETEC_p666_0650c, *eltB*	n.d.
aag**tgcga**atctat**ta**a**ca**ata	7.62		ETEC_p948_0800c	n.d.
attca**tg**cagcaaa**tcaca**tca	7.61		ETEC_p948_0500c, *traM*	n.d.
aaa**tgt**t**a**tatcttc**ctc**ttta	7.58		ETEC_p666_0350c, *traY*	n.d.
tta**tgt**t**a**actcaa**ttt**c**a**ata	7.52		37232 -> 37253	n.d.
aaactca**a**attgag**tcaca**aca	7.51		ETEC_p948_0680, *stbA*	n.d.
aat**ta**t**g**gtgataa**ta**a**ta**ttt	7.48		41344 -> 41365	n.d.
aataa**ta**a**t**aaaaag**ca**a**a**aaa	7.47		ETEC_p948_1070c	n.d.
aaa**tgt**a**a**ttgataat**a**a**a**aaa	7.45		38188 -> 38209	n.d.
aat**tgtg**gaagaaa**t**t**aca**atg	7.44		40522 -> 40543	n.d.
atc**ttcga**ccatat**tcg**c**a**tat	7.4		47921 -> 47942	n.d.
cat**tgtc**atattt**a**t**ca**g**a**aaa	7.39		ETEC_p948_0990c	n.d.
caa**tgt**t**a**tttatat**ca**ttaaa	7.39		ETEC_p948_0990c	n.d.
att**tctta**tagaaa**tta**c**t**ttt	7.38		ETEC_p948_0800c	n.d.
aatg**gcga**tgctaa**ta**a**a**ataa	7.37		ETEC_p666_0660c, *eltA*	n.d.
ttt**tgt**t**a**ttatta**tc**t**a**a**g**ct	7.31		ETEC_p948_0870c	n.d.
gaaca**tga**gcagca**tca**t**a**aaa	7.31		ETEC_p666_0360c, *traJ*	n.d.
tac**tgc**ggcgcagt**tcac**gatt	7.3		ETEC_p948_0050c	n.d.
aat**tt**t**gg**tctcgg**tca**g**a**tat	7.3		ETEC_p666_0650c, *eltB*	n.d.
aatc**gtga**actgcgc**c**g**ca**gta	7.3		ETEC_p948_0240	n.d.
ttt**ttt**t**a**attgcg**t**tg**c**atat	7.27		ETEC_p948_0900c	n.d.
ataa**gtga**tagtct**ta**a**tac**ta	7.26		ETEC_p948_0470	n.d.
taa**tgtg**ttaggca**tt**a**a**catt	7.25		61025 -> 61046	n.d.
tga**tgtg**gtatgtt**tt**a**t**cttt	7.21		ETEC_p666_0110c	n.d.
caa**ta**t**ga**tttagt**ta**a**c**ggta	7.2		ETEC_p666_0550, *stbA*	n.d.
aaac**g**ca**a**tgtatt**tca**ttatt	7.2		60523 -> 60544	n.d.
tcag**gtga**tgcact**tca**a**a**aag	7.19		29938 -> 29959	n.d.
atc**tt**t**ga**taattt**tc**t**ca**atg	7.19		25210 -> 25231	n.d.
ttactca**a**tcctct**tcaca**aca	7.18		ETEC_p666_0560, *stbB*	n.d.
att**tgtg**tcatggg**tt**a**c**cata	7.16		75209 -> 75230	n.d.
ata**tt**t**ga**tatctg**t**g**a**t**a**tct	7.14		43102 -> 43123	n.d.
aga**tt**aa**a**aaaacac**caca**aaa	7.14		90310 -> 90331	n.d.
aatga**tga**atatca**tca**attat	7.13		39580 -> 39601	n.d.

	ttt**ttcc**a**t**ctgca**tca**aaatt	7.12		67889 -> 67910	n.d.
	ttt**ta**t**ga**tggata**t**agt**a**cta	7.12		42475 -> 42496	n.d.
	aat**tgt**t**a**ttggtgaa**a**ta**a**tt	7.12		5162 -> 5183	n.d.
	ttg**tttt**a**t**caaaa**tca**tgttt	7.11		4348 -> 4369	n.d.
	tct**tgtga**gggaga**t**tgt**a**ttt	7.11		77103 -> 77124	n.d.
	ata**tt**a**ga**acgata**t**tg**ca**taa	7.11		74898 -> 74919	n.d.
	tac**tgtga**attaat**taac**gttg	7.09		ETEC_p948_0120c, *etpB*	n.d.
	aac**tt**a**a**taaacac**cac**atta	7.09		ETEC_p948_0830c	n.d.
	tgt**tttt**tacaaca**tcaca**ctt	7.08		ETEC_p666_0750c, *sta1*	n.d.
	tgt**tttt**tacaaca**tcaca**ctt	7.08		ETEC_p666_0750c, *sta1*	n.d.
	atag**gt**a**a**atctgt**tca**aaaaa	7.08		ETEC_p948_0790c	n.d.
	taac**gtg**gcattaac**cac**gtaa	7.07		ETEC_p948_0490c, *traJ*	n.d.
	tga**tt**t**ga**ctgctc**tt**aaattt	7.06		ETEC_p666_0870c	n.d.
	aaac**g**g**ga**acgggt**tcaca**aac	7.03		10961 -> 10982	n.d.
	aaa**tg**ca**a**actttat**g**at**a**tat	7.02		37876 -> 37897	n.d.
	aaa**tgt**attccgtc**tcaca**tgt	7		61211 -> 61232	n.d.

Sites investigated using EMSA analysis are indicated in red.

[a] Each bin contains sites with a similar affinity for CRP as predicted by PREDetector.

[b] Denotes the sequence of the CRP site predicted by PREDetector. Nucleotides in bold conform to the consensus CRP binding site.

[c] Denotes the score assigned to the target by PREDetector depending on similarity to the consensus. Sites identified below a cut-off score of 7 were omitted from the analysis.

[d] Denotes the position of the target site. Where targets are associated with genes, these are given in italics.
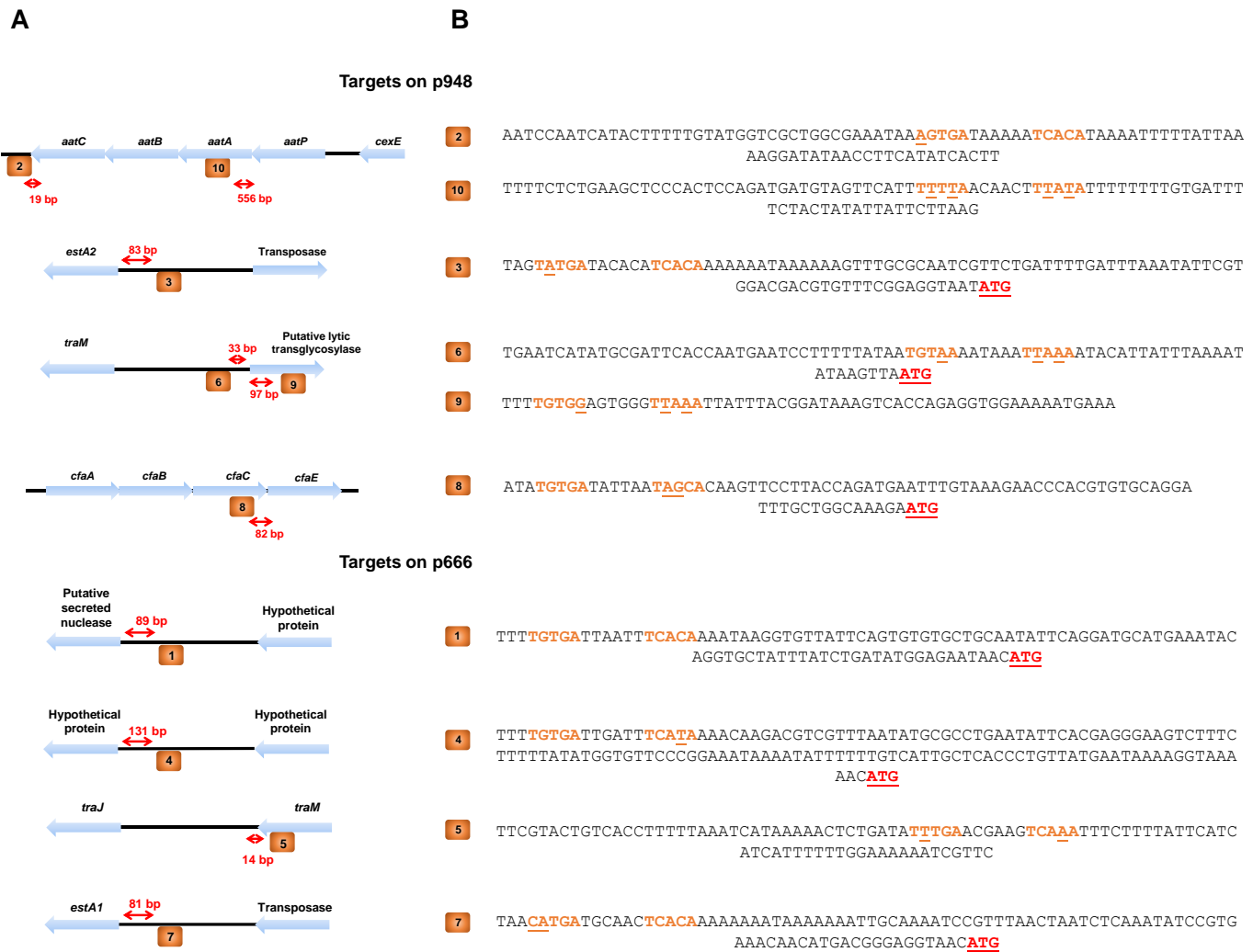
Figure 3.9 Genomic context and sequence of high-affinity CRP targets predicted by

bioinformatics analysis on plasmids p948 and p666

The genomic context of each predicted CRP target is shown (A), where the CRP sites are shown as orange boxes, the site number is shown within each box. Blue arrows indicate nearby genes, black lines represent intergenic DNA. Red text and arrows give the distance from the centre of the CRP site to the nearest gene boundary. The corresponding sequence of each predicted CRP sites are shown (B). CRP half-sites are shown in orange, with mis-matches to the consensus underscored. The start codons of nearby genes are shown in red.

The raw EMSA data are shown in Figure 3.10 and a quantification of the data for one representative site of each bin (i.e sites from Figure 3.10A) is shown in Figure 3.11. Targets 4, 6, 8, 9 and 10 did not bind CRP whilst targets 1, 2, 3, 5 and 7 did. Hence, binding of CRP broadly correlated with predicted site score. Interestingly, these five high-affinity sites (1, 2, 3, 5 and 7) were all bound by H-NS *in vivo* (Figure 3.12). Thus, we speculate that H-NS might prevent CRP binding to these sites *in vivo*. Importantly, targets 3 and 7 were upstream of the genes encoding STa2 and STa1 on plasmids p948 and p666 respectively.
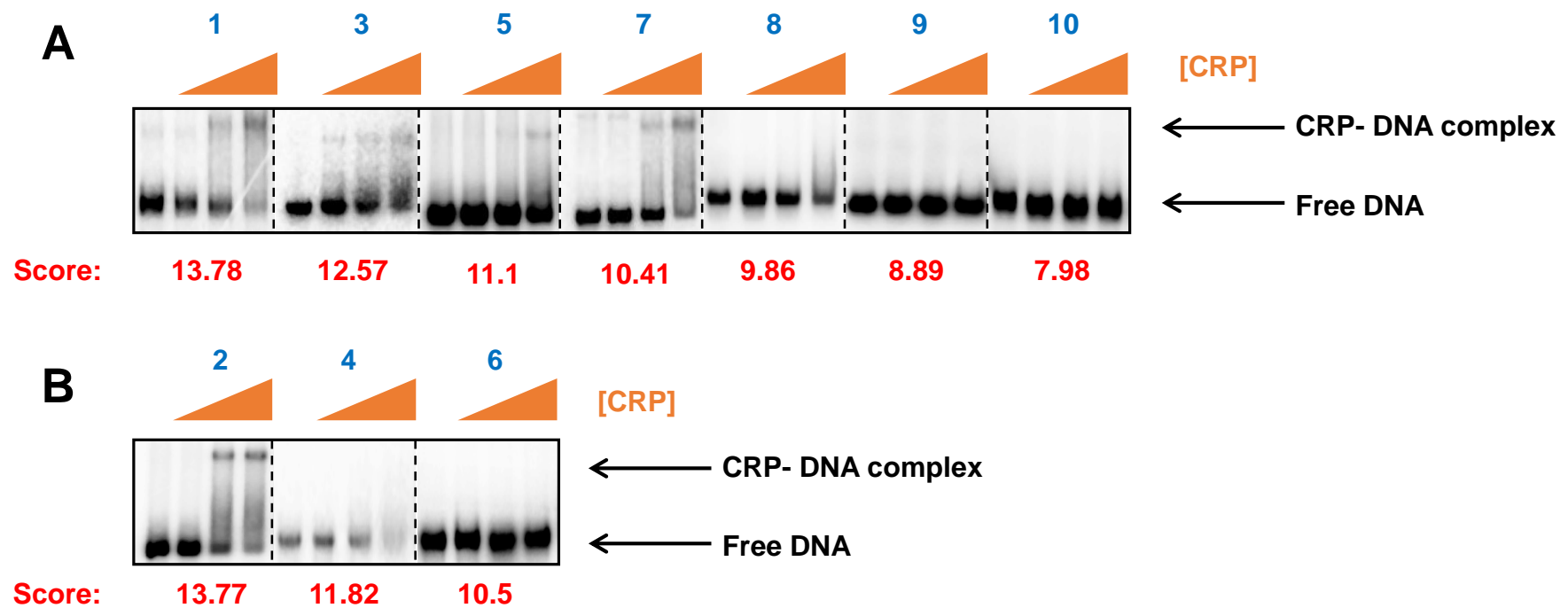
Figure 3.10 *In vitro* analysis of CRP binding to selected targets predicted *in silico* by EMSA

Radiolabelled *Aat*II-*Hin*dIII fragments were incubated with purified CRP for 20 minutes at 37 °C before being run on a 7.5 % (*w/v*) polyacrylamide gel. Gels were run at 200 V for 2 hours, before being dried and exposed to a phosphorscreen. Each window shows the results from an EMSA experiment using a different CRP target. CRP concentrations from left to right in each window are as follows; 0, 175, 350 and 700 nM. All reactions contained 0.2 mM cAMP. Free DNA and CRP-DNA complexes are indicated by arrows. Scores assigned by the bioinformatics screen are given in red.
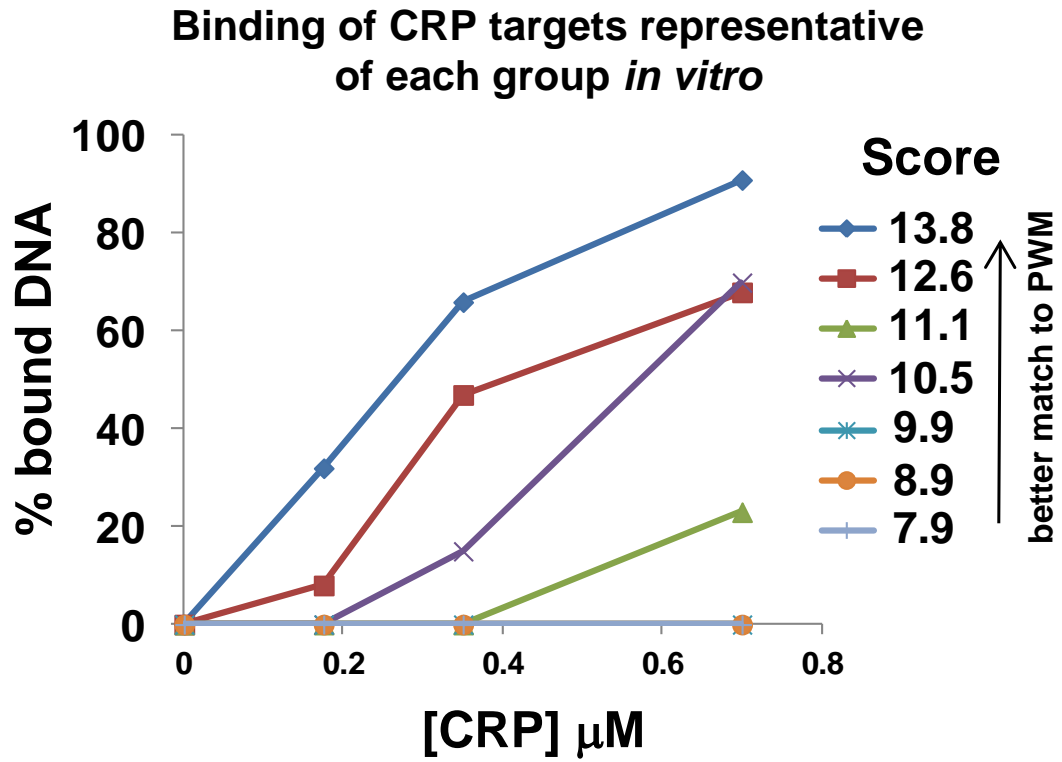
Figure 3.11 Graphical representation of *in vitro* analysis of CRP binding to selected

targets predicted *in silico*

One representative of each bin (Table 3.2) is shown. The figure graphically shows data from Figure 3.10A.
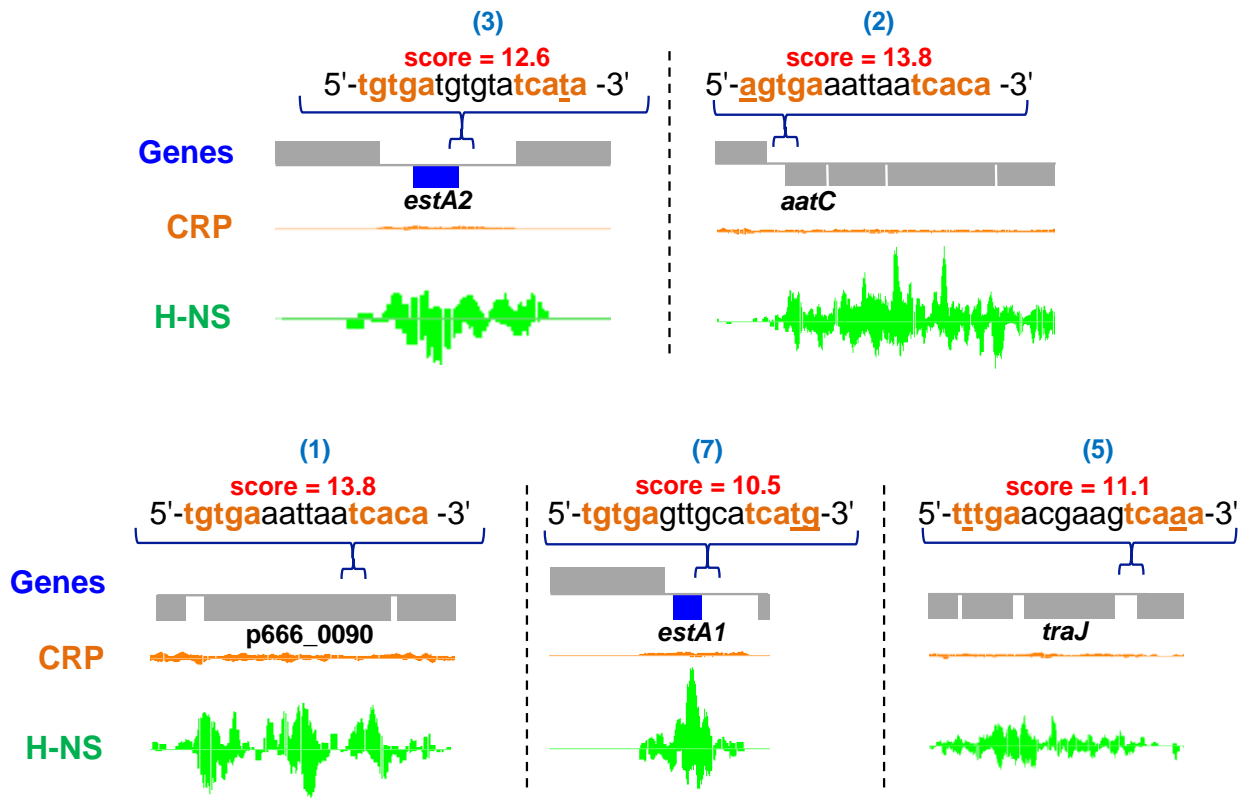
Figure 3.12 *In vivo* **binding of CRP and H-NS at high-affinity CRP target identified by**

bioinformatics on virulence plasmids p948 and p666 as determined by ChIP-seq

Each panel shows CRP and H-NS binding across a different locus. Genes are depicted in grey blocks. CRP targets identified by bioinformatics are shown above genes, half sites are shown in orange. Deviations from the consensus are indicated by underlined bases. Binding scores are given in red. CRP binding across each region as determined by ChIP-seq analysis is shown in orange. H-NS binding is shown in green. The blue indicates the relative position of the CRP site. Blue blocks indicate the heat-stable toxin genes *estA1* and *estA2*. The maximum number of reads mapped per locus is as follows: (1): 1023 reads, (2): 1800, (3): 606, (5): 906, (7): 892.

3.3 Discussion: a role for CRP in ETEC virulence

The aim of the work presented in this chapter was to investigate the role of CRP, H-NS and σ^{70} in regulating expression of ETEC virulence factors. Using ChIP-seq 111 high-confidence CRP binding sites were identified on the chromosome. Additionally, 364 peaks for H-NS, and 1089 peaks for σ^{70} were found. The majority of CRP binding sites were in intergenic regions but very few (7 %) of these sites were located upstream of ETEC-specific genes. In summary, binding targets of CRP in ETEC H10407 appears to be very similar to that of K-12 (Table 3.1 and Figure 3.5). Hence, CRP binding was abundant in the vicinity of genes encoding transporters, phosphotransferase systems, metabolic enzymes, and transcription factors.

We observed small differences in the broad categories of chromosomal genes targeted by CRP in K-12 and ETEC H10407, as shown in Figure 3.5D. The precise functioning of these sites might be a subject for further study. For example, in ETEC H10407, CRP has more binding targets associated with genes involved in amino acid metabolism and nucleotide metabolism compared to K-12. It is possible that this might reflect small changes in metabolic regulation in ETEC, which could confer an advantage in colonisation, when in competition with intestinal commensals. For example, Fabich *et al.* (2008) have shown that *E. coli* O157:H7 and MG1655 utilise different sugars when colonising the mouse intestine. In addition, strain O157:H7 switches to using gluconeogenic substrates in streptomycin-treated mice co-colonised with MG1655.

Surprisingly, CRP binding was observed only on the H10407 chromosome, and no sites could be observed on any of the four virulence plasmids. CRP binding to these plasmids is of particular interest since they encode the pathogenicity factors. Despite a lack of CRP binding as determined by ChIP-seq we were able to identify a number of high affinity CRP targets bioinformatically. Two such targets were found upstream of genes encoding STa1 and STa2 on p666 and p948 respectively. These sites were bound by H-NS, rather than CRP *in vivo*. Thus

it is likely that CRP binding is occluded by H-NS in these regions. H-NS occlusion of transcription factors has been previously described by Myers *et al.* (2013), who described a significant increase in FNR binding across the K-12 MG1655 chromosome in a Δ*hns/*Δ*stpA* strain. Interestingly, like our study, Myers *et al.* found that transcription factor occupancy of a binding site correlated poorly with its resemblance to the consensus. Hence, the binding of conventional transcription factors is sensitive to other factors, including NAPs like H-NS.

Transcriptional regulation of ETEC virulence gene expression by H-NS will require further study, although there are numerous reports of H-NS involvement in ETEC pathogenicity. For example, repression of the *eltAB* operon by H-NS has been reported by Yang *et al.* (2005). The heat-labile toxin is secreted through association with outer membrane vesicles (OMVs). Mutating the *hns* gene results in increased production of OMVs (Horstman and Kuehn, 2002). Jordi *et al.* (1994), have also reported H-NS-mediated repression of the *cfaABCE* operon. It will be interesting to investigate whether H-NS plays a conserved role in other ETEC strains.

Chapter 4

CRP-dependent regulation of ETEC enterotoxin expression

4.1 Introduction

The analysis presented in Chapter 3 identified CRP sites upstream of the *estA1* and *estA2* genes that encode heat-stable enterotoxin derivatives. However, these targets were bound by H-NS rather than CRP *in vivo* under the conditions tested (Figure 4.1A and Figure 4.1B). Similarly, *eltAB* has been proposed to bind three CRP dimers, but the promoter region was bound by H-NS not CRP in our ChIP-seq analysis (Figure 4.1C). A possible explanation for these observations is occlusion of CRP binding sites by H-NS. Binding of H-NS to promoters usually involves co-operative interactions between H-NS monomers bound at multiple adjacent sites. If promoters are removed from their genomic context this co-operativity is often lost and H-NS binding is negated (Singh *et al.,* 2014). Thus, in this chapter, we have examined the effect of removing H-NS bound CRP targets from their position within H-NS filaments.

4.2 Characterisation of P*estA2*

4.2.1 Identification of P*estA2* transcription start site

To characterise P*estA2*, a 93 bp DNA fragment was generated that included the predicted CRP site and translation start codon of *estA2*. Importantly, this fragment included a potential regulatory region (shown in Figure 4.2A) and lacked the H-NS bound *estA2* coding sequence. The fragment was cloned into pRW50 to create a promoter::*lacZ* fusion, and used to transform *E. coli* K-12 strain M182. As a first step towards understanding P*estA2* regulation we isolated RNA from the transformants and determined the P*estA2* transcription start site by primer extension using an oligonucleotide primer that bound to the 5' end of the *lacZ* mRNA of pRW50. The result of the primer extension assay is shown in Figure 4.2B. A 109 nucleotide primer extension product was observed that maps to the transcription start site labelled '+1' in Figure 4.2A. From the mapped transcription start site, we were able to identify a -35 hexamer (sequence 5'-TTGCGC-3') and a -10 hexamer (sequence 5'-TTAAAT-3'). The predicted
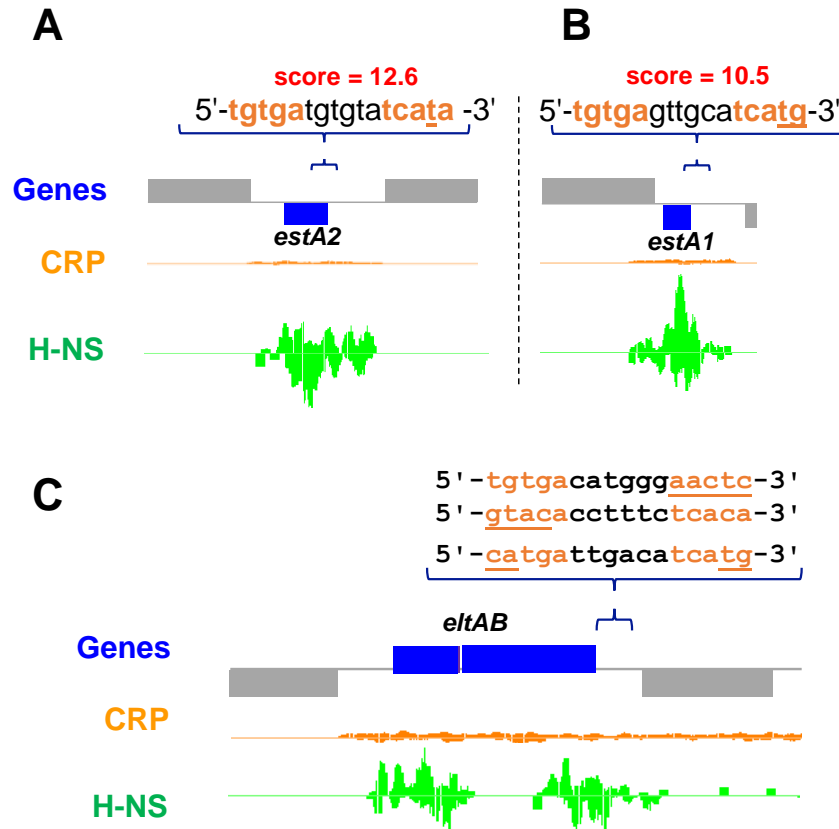
Figure 4.1 Distribution of CRP and H-NS across ETEC H10407 enterotoxin gene loci as determined by ChIP-seq

Genes are shown as grey and blue blocks, CRP binding is shown in orange, and H-NS binding is shown in green. Above the panel, the sequences of the predicted CRP sites (A and B), or previously characterised CRP sites (C) are shown. CRP half-sites are shown in orange, deviations from the consensus are underlined. Where appropriate, the predicted score assigned by PREDetector is shown in red.

A) CRP and H-NS binding across the *estA2* locus on p948. Maximum number of H-NS reads: 606.
B) CRP and H-NS binding across the *estA1* locus p666. Maximum number of H-NS reads: 892.
C) CRP and H-NS binding across the *eltAB* locus on p666. CRP sites shown above the panel are those proposed by Bodero and Munson (2009). Maximum number of H-NS reads: 820.

126

A

```
        -60       -50       -40       -30       -20       -10       +1
5'-TATGATACACATCACAAAAAAATAAAAAAGTTTGCGCAATCGTTCTGATTTTGATTTAAATATTCGTGG-3'
         CRP        UP element     -35     spacer      -10  discriminator
```

B
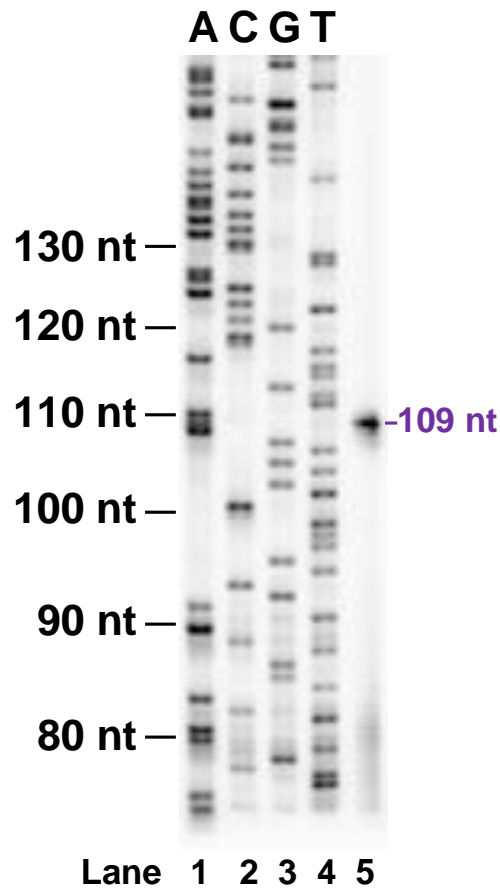
Figure 4.2 Mapping of the transcriptional start site of P*estA2*

A) Sequence of the P*estA2* regulatory region. The transcription start site (determined by primer extension) is labelled as '+1', and is underlined in red. The -35 and -10 hexamers are shown in purple, the UP element is shown in blue, and the CRP half sites are shown in orange.

B) Primer extension analysis of P*estA2*. Lane 5 shows the primer extension product, which labelled in purple. Lanes 1-4 are M13 T7 sequencing reactions ('A', 'C', 'G', and 'T') used to calibrate the gel. The experimentally derived transcription start site is marked as "+1" in (A).

CRP binding site was centred 59.5 bp upstream of the transcription start site.

4.2.2 Binding of CRP to the predicted P*estA2* CRP binding site

DNAse I footprinting was used to confirm CRP binding to the predicted site. To this end, the P*est*A2 fragment was cloned into plasmid pSR and a 171 bp DNA fragment, containing the 93 bp P*estA2* sequence, was generated by digestion with *Aat*II and *Hin*dIII. The fragment was end-labelled and used in DNAse I footprinting assays. Figure 4.3 shows the footprint. The gel was calibrated with a Maxim/Gilbert G+A ladder, and numbered according to the P*estA2* transcription start site (+1). Lane 1 shows the pattern of DNase I cleavage in the absence of CRP. Addition of CRP results in changes in the cleavage pattern between promoter positions -54 and -70 (Figure 4.3, lanes 2-8).

4.2.3 CRP activates transcription from P*estA2* *in vitro* and *in vivo*

The configuration of P*estA2* suggested that it was likely a class I CRP activated promoter (Zhou *et al.*, 1993, Busby and Ebright, 1999). To test this prediction the pRW50 construct, containing P*estA2*, was used to transform M182 and the Δ*crp* derivative. LacZ activity levels were then determined in the transformants (Figure 4.4A). A four-fold increase in LacZ activity was observed in wildtype cells compared to cells lacking CRP. To confirm that the effect of CRP was direct, the pSR construct containing P*estA2* was used as a template for *in vitro* transcription with purified RNA polymerase, and CRP. Transcription from P*estA2* generates a 112 nucleotide RNA separable from the RNAI control (which is not dependent on CRP) using high resolution denaturing PAGE. Production of the 112 nuclotide transcript is CRP dependent (Figure 4.4B).
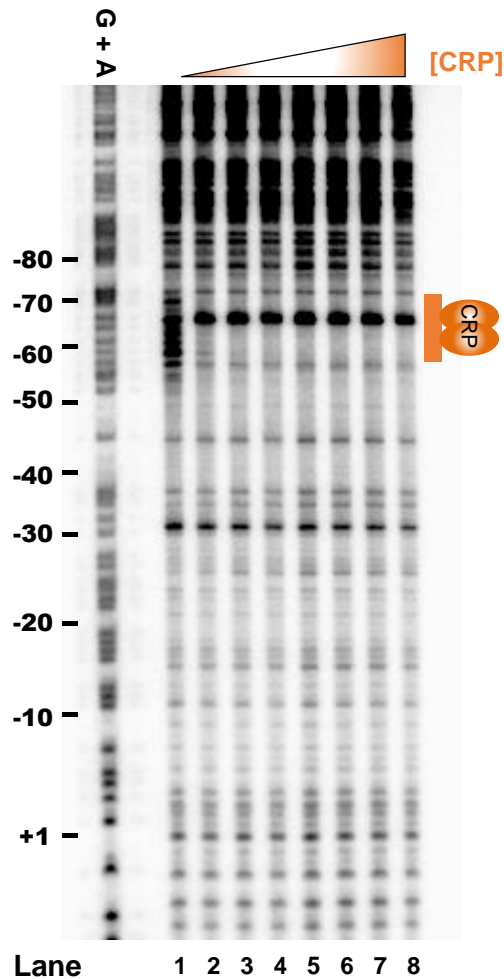
Figure 4.3 Confirmation of the position of the CRP site at P*estA2* using DNAse I footprinting

G + A is a Maxim/Gilbert G + A ladder, which shows the position of pyrimidine bases in the sequence. Numbering refers to distance from the transcription start site, as determined by primer extension. Each band represents a site at which DNA has been cleaved. Lane 1 shows the cleavage pattern of P*estA2* when exposed to DNAse I, in the absence of CRP. Lanes 2 to 8 show cleavage patterns of P*estA2* when exposed to increasing concentrations of CRP. Final CRP concentrations used in lanes 2-8 respectively are; 0.35 μM, 0.7 μM, 1.05 μM, 1.4 μM, 1.75 μM, 2.1 μM and 2.45 μM. All reactions contain 0.2 mM cAMP. All reactions contained 12.5 μg/ml Herring Sperm DNA as a non-specific competitor. 0.2 units of DNAse I were used per reaction. Reactions were run on a 6 % (*w/v*) denaturing gel.

A
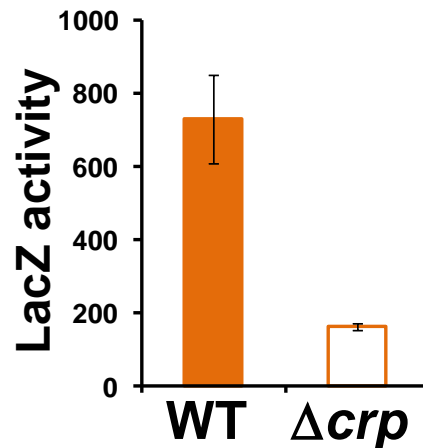
B
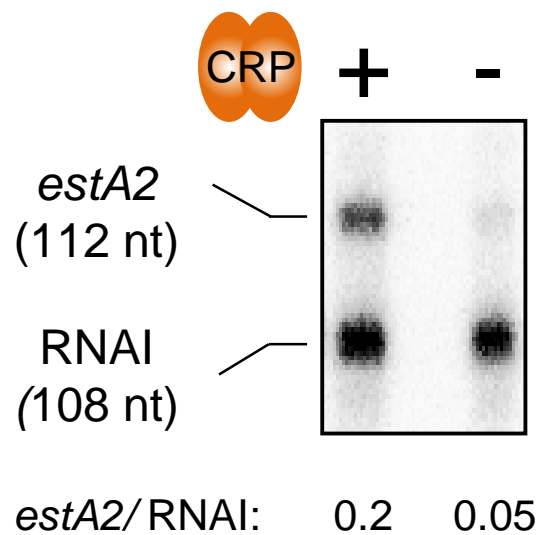
Figure 4.4 P*estA2* **is activated by CRP** *in vivo* **and** *in vitro*

A) β-galactosidase activity of P*estA2* cloned into pRW50, in M182 wildtype and M182 Δ*crp* background. Cells were grown to mid-log phase (OD$_{650}$=0.3-0.6). Standard deviation is shown for three biological replicates. Activity shown is in Miller units.

B) *In vitro* transcription products of P*estA2* cloned upstream of a *λoop* terminator in pSR. RNAP containing σ^{70} was used at a concentration of 400 nM, CRP was used at a concentration of 350 nM. Reactions were run for 10 minutes at 37 °C before being stopped. Transcription products were run on a 6 % (*w/v*) denaturing gel. The P*estA2* transcript and 108 nt RNAI control transcript are labelled.

4.3 Characterisation of P*estA1*

4.3.1 Identification of the P*estA1* transcription start site

Having established that P*estA2* is a class I dependent CRP activated promoter we next characterised regulation of e*stA1*. Examination of the sequence upstream of *estA1* revealed DNA elements similar to those at *PestA2*. Hence, a predicted CRP site and promoter-like sequences were present (Figure 4.5A). To determine if P*estA2* and P*estA1* share an equivalent transcription start site a 92 bp fragment was generated, equivalent to the P*estA2* DNA fragment described above (Figure 4.5A). This DNA fragment was cloned into pRW50, and the resulting construct was used to transform M182 and the Δ*crp* derivative. RNA was extracted, and primer extension was used to determine the transcription start site of P*estA1*. Figure 4.5B shows the primer extension product, alongside calibrating sequencing reactions. Just as for P*estA2*, a 109 nucleotide product is observed that maps to the equivalent transcription start site. However, the amount of primer extension product increased in cells lacking CRP. This was unexpected, since P*estA1* superficially resembles P*estA2*.

4.3.2 P*estA1* is repressed by CRP

To confirm that P*estA1* was repressed by CRP, we measured LacZ activity in M182 and corresponding Δ*crp* cells transformed with the pRW50 construct containing P*estA1* fused to *lacZ*. The data show that P*estA1* is repressed by CRP and that, compared to P*estA2*, P*estA1* is poorly active (compare Figure 4.4A and Figure 4.5C). To check that the CRP effect was direct, the 92 bp P*estA1* fragment was cloned into pSR and used as a template for *in vitro* transcription assays. Figure 4.5D shows that production of the transcript from P*estA1* is repressed by CRP.

A

```
                -60           -50           -40          -30           -20           -10           +1
                 •             •             •            •             •             •             •
PestA2  TATGATACACATCACAAAAAAATAAAAAAGTTTGCGCAATCGTTCTGATTTTGATTTAAATATTCGTGG
                |||||  || ||  |||||||||||||| |||| |||  |||| ||||  | | |  | |||||| |||
PestA1  CATGATGCAACTCACAAAAAAATAAAAAAATTGCAAaaATCCGTTTAACTAATC-TCAAATATCCGTGA
```

CRP **UP element** **-35** spacer **-10** discriminator

B

C

D

132
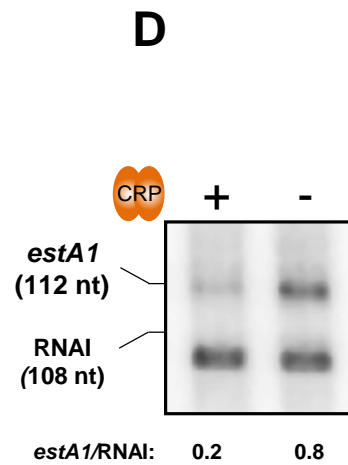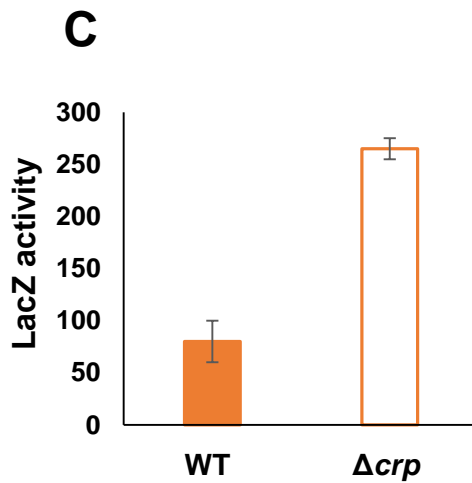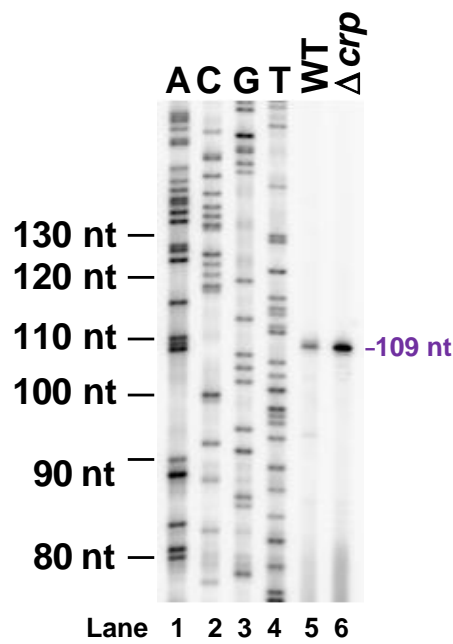
Figure 4.5 P*estA1* is a CRP-repressed promoter

A) Alignment of P*estA2* and P*estA1* promoters. Numbering refers to P*estA2* transcriptional start site. CRP half-sites are shown in orange, UP elements are shown in blue, -35 and -10 hexamers are shown in purple.

B) Primer extension assay to map the transcriptional start site of P*estA1*. Lanes 1-4 are M13 T7 sequencing reactions ('A', 'C', 'G', and 'T') used to calibrate the gel. RNA was extracted from M182 cells and M182 Δ*crp* cells harbouring a pRW50 plasmid containing the P*estA1* promoter fragment. Lanes 5 and 6 show primer extension products generated from RNA from M182 wildtype and Δ*crp* cells respectively which are labelled in purple. The mapped transcription start site is shown in Figure 4.5A.

C) β-galactosidase activity of P*estA1* cloned into pRW50, in M182 wildtype and M182 Δ*crp* background. Cells were grown to mid-log phase (OD_{650}=0.3-0.6). Standard deviation is shown for three biological replicates. Activity shown is in Miller units.

D) *In vitro* transcription products of P*estA1* cloned upstream of a *λoop* terminator in pSR. RNA polymerase was used at a concentration of 400 nM, CRP was used at a concentration of 350 nM. Reactions were run for 10 minutes at 37 °C before being stopped. Transcription products were run on a 6 % (*w/v*) denaturing gel. The P*estA1* transcript and 108 nt *RNAI* control transcript are labelled. A quantification of the transcripts is shown below each lane.

4.4 Construction of hybrid P*estA2*:P*estA1* promoters

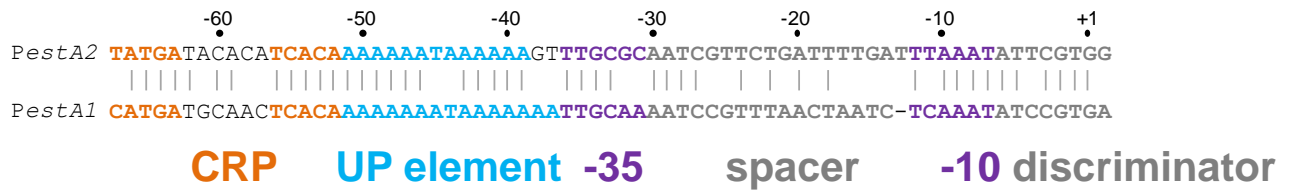To understand why P*estA1* and P*estA2* behaved differently, a series of hybrid promoters was generated. The various hybrids are each based on P*estA2* but contain a discrete promoter element from P*estA1*. The constructs are shown in Figure 4.6A, where underlining represents replacement of P*estA2* sequence with the equivalent P*estA1* element. The hybrid constructs were cloned into pRW50 and used to transform M182 and M182Δ*crp*. LacZ activities determined for the transformants are shown in Figure 4.6B. The table beneath the graph also describes the composition of each promoter construct.

Each change has an effect on promoter activity. Introducing the P*estA1* CRP site (P*estA2.1*), or P*estA1* spacer region (P*estA2.4*) renders the hybrid promoter insensitive to CRP. Introducing the P*estA1* UP element (P*estA2.2*) or -10 element (P*estA2.6*) reduces promoter activity, but not CRP-dependent activation. Crucially, replacement of the P*estA1* -35 element of P*estA2* with that of *Pest1* (P*estA2.3* and P*estA2.5*) results in repression by CRP. Only the discriminator sequence of P*estA1* had no effect on promoter activity or regulation. We conclude that the -35 element is responsible for CRP-dependent repression and that the other changes in the P*estA1* sequence (relative to P*estA2*) result in the overall lower activity of P*estA1*.

4.5 CRP-dependent regulation at *PeltAB*

Having established that CRP activates Pe*stA2* and represses P*estA1* we turned our attention to regulation of the *eltAB* operon encoding the heat-labile toxin gene. Recall that, P*eltAB* was previously suggested to bind CRP at three sites upstream of the transcription start site (Bodero and Munson, 2009). Binding of CRP to these sites is proposed to repress expression of *eltAB*. However, our bioinformatic screen (Chapter 3) did not predict any CRP sites at P*eltAB*. Moreover, the CRP sites predicted by Bodero and Munson are all a poor match to the consensus for CRP binding. Thus, we next sought to determine whether CRP was able to bind P*eltAB*.

A

```
PestA2.1  CATGATGCAACTCACAAAAAAATAAAAAAGTTTGCGCAATCGTTCTGATTTTGATTTAAATATTCGTGG
PestA2.2  TATGATACACATCACAAAAAAAATAAAAAAATTGCGCAATCGTTCTGATTTTGATTTAAATATTCGTGG
PestA2.3  TATGATACACATCACAAAAAATAAAAAAGTTTGCAAAATCGTTCTGATTTTGATTTAAATATTCGTGG
PestA2.4  TATGATACACATCACAAAAAATAAAAAAGTTTGCGCAATCCGTTTAACTAATC-TTAAATATTCGTGG
PestA2.5  TATGATACACATCACAAAAAATAAAAAAGTTTGCAAAATCCGTTTAACTAATC-TTAAATATTCGTGG
PestA2.6  TATGATACACATCACAAAAAATAAAAAAGTTTGCGCAATCGTTCTGATTTTGATTCAAATATTCGTGG
PestA2.7  TATGATACACATCACAAAAAATAAAAAAGTTTGCGCAATCGTTCTGATTTTGATTTAAATATCCGTGA
```

CRP   UP element   -35   spacer   -10 discriminator

B

			2.1	2.2	2.3	2.4	2.5	2.6	2.7	
CRP site	N/A	PestA2	*PestA1*	PestA2	PestA2	PestA2	PestA2	PestA2	PestA2	*PestA1*
UP element	N/A	PestA2	PestA2	*PestA1*	PestA2	PestA2	PestA2	PestA2	PestA2	*PestA1*
-35 element	N/A	PestA2	PestA2	PestA2	*PestA1*	PestA2	*PestA1*	PestA2	PestA2	*PestA1*
spacer	N/A	PestA2	PestA2	PestA2	PestA2	*PestA1*	*PestA1*	PestA2	PestA2	*PestA1*
-10 element	N/A	PestA2	PestA2	PestA2	PestA2	PestA2	PestA2	*PestA1*	PestA2	*PestA1*
descriminator	N/A	PestA2	PestA2	PestA2	PestA2	PestA2	PestA2	PestA2	*PestA1*	*PestA1*

pRW50 only   PestA2   different PestA hybrid promoters   PestA1

135

Figure 4.6 P*estA2*:P*estA1* hybrid promoters

A) Hybrid promoters generated for P*estA2*:P*estA1* trajectory mutagenesis. All sequences are based on P*estA2*, but contain one or more promoter elements from *PestA1*, these are underlined. CRP half-sites are shown in orange, the UP element in shown in blue, -35 and -10 hexamers are shown in purple.

B) β-galactosidase activity of P*estA2*/P*estA1* hybrid promoters cloned into pRW50, in M182 wildtype and M182 Δ*crp* backgrounds. Cells were grown to mid-log phase (OD$_{650}$=0.3-0.6). Standard deviation is shown for three biological replicates. Activity shown is in Miller units. The schematic below the graph indicates the composition of the promoter. Grey boxes indicate elements from P*estA2*, yellow boxes indicate elements from P*estA1*. Empty pRW50 vector control activities are included.

4.5.1 CRP does not bind the P*eltAB* region *in vitro*

As a starting point a 359 bp *Eco*RI- *Hin*dIII fragment containing P*eltAB* was generated. This is a precise match to the P*eltAB* fragment used in the study of Bodero and Munson (Figure 4.7A) (2009). The DNA fragment was radiolabelled and used for *in vitro* EMSA experiments to test CRP binding. We also ran and EMSA experiments with P*estA2* as a positive control. The data are shown in Figure 4.7B. The P*estA2* fragment bound CRP tightly and formed a discrete complex with CRP (lanes 3-6). Further increases in CRP concentration led to non-specific binding of CRP to P*estA2*, which manifested itself as a smear in lane 7 of Figure 4.7B. Conversely, no specific binding of CRP to P*eltAB* was observed (lanes 8-13) but, at high CRP concentrations, non-specific binding was apparent (lane 14).

4.5.2 Repression of P*eltAB* by CRP is indirect

Our EMSA data suggest that CRP does not bind specifically to P*eltAB*. Consequently, we reasoned that the repressive effect of CRP on P*eltAB* reported by Bodero and Munson must be indirect. To test this prediction, a series of P*eltAB* derivatives was generated where the CRP sites were sequentially removed (Figure 4.8A). P*eltAB* 1.1 is the 359 bp P*eltAB* fragment, identical to that used by Bodero and Munson (2009). P*eltAB* 1.2 is a shorter 118 bp fragment, which contains only the promoter proximal CRP site. P*eltAB* 1.3 is a derivative of P*eltAB* 1.2 in which the remaining CRP site has two point mutations in each half site. The various P*eltAB* fragments were cloned into pRW50 and used to transform M182 cells or the Δ*crp* derivative. Results of β-galactosidase assays are shown in Figure 4.8B. Removal of the two upstream CRP sites and mutation of the remaining downstream CRP site did not change the response of the promoter to CRP; in all P*eltAB* derivatives promoter activity increased 2.5 fold in Δ*crp* cells. Taken together, the lack of CRP binding to P*eltAB*, and the insensitivity of the promoter to CRP binding site mutation, suggests that CRP repression of P*eltAB* is indirect.

A

```
     -300          -290          -280          -270          -260          -250          -240
      .            .             .             .             .             .             .
TTCTGGTGTGGACTTTCTGGTGCTCCAGGTTGTGTGACATGGGAACTCATTCTGGATGGTTA

     -230          -220          -210          -200          -190          -180
      .            .             .             .             .             .
CTCTGAAAGCTCATATTCTGCCACCCCCCGATTTGCAGCCGCCAAGCTGCCGTGGTTCAAGT

     -170          -160          -150          -140          -130          -120
      .            .             .             .             .             .
CGCGACTAATAAAAATAATCAGGTTGCCATGATTCAATGTACACCTTTCTCACATTCGTCTC

     -110          -100          -90           -80           -70           -60
      .            .             .             .             .             .
CGGCATGAAAACGATGCACTCTTTCTTTATCGCTTTCACTACACATTTTATCCTCGCATGGA

     -50           -40           -30           -20           -10           +1
      .            .             .             .             .             .
TGTTTTATAAAAAACATGATTGACATCATGTTGCATATAGGTTAAACAAAACAAGTGGCGTT

     +10           +20           +30           +40           +50
      .            .             .             .             .
ATCTTTTTCCGGATTGTCTTCTTGTATGATATATAAGTTTTCCTCGATG
```

B
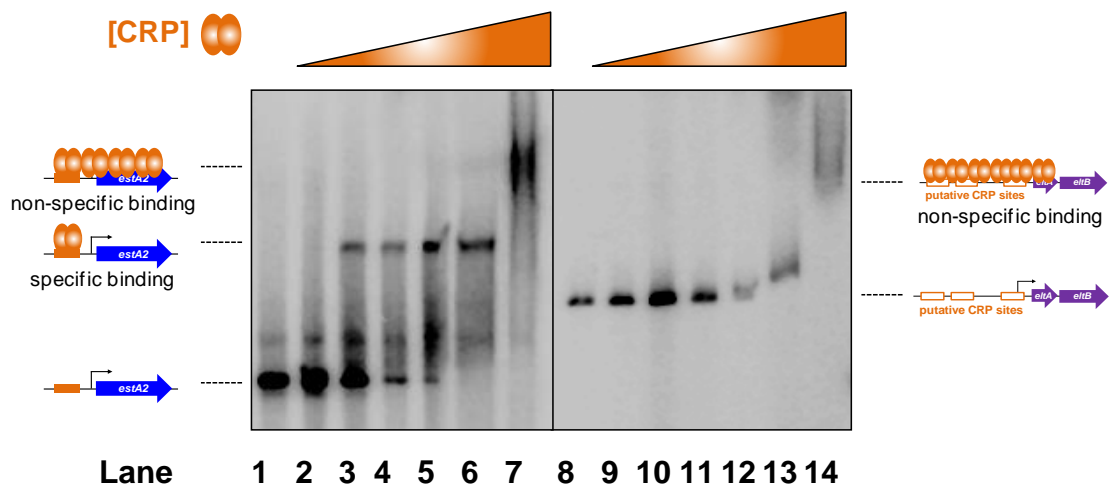
Figure 4.7 CRP does not bind to the P*eltAB* promoter in *in vitro* EMSAs

A) Sequence of the P*eltAB* region. CRP half-sites defined by Bodero and Munson (2009) are shown in orange and are underlined. The -35 and -10 hexamers are shown in purple. Annotation is relative to the transcription start site (indicated by '+1').

B) *In vitro* EMSA of the P*eltAB* (lanes 8-14) and P*estA2* promoter (lanes 1-7). End- labelled *EcoRI-HindIII* flanked fragments were incubated with various concentrations of CRP at 37 °C for 20 minutes before being loaded onto a 7.5 % (*w/v*) polyacrylamide gel. CRP concentrations for each panel are as follows; 0 μM, 0.175 μM, 0.35 μM, 0.7 μM, 1.4 μM, 2.8 μM, and 7 μM. All reactions contain 12.5 μg/ml Herring Sperm DNA as a non-specific competitor, and 0.2 mM cAMP.
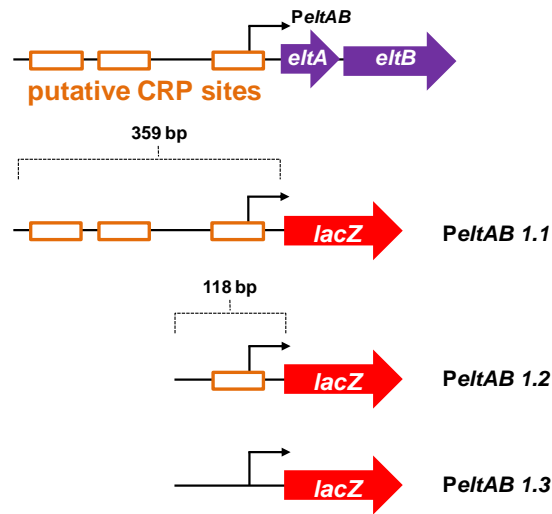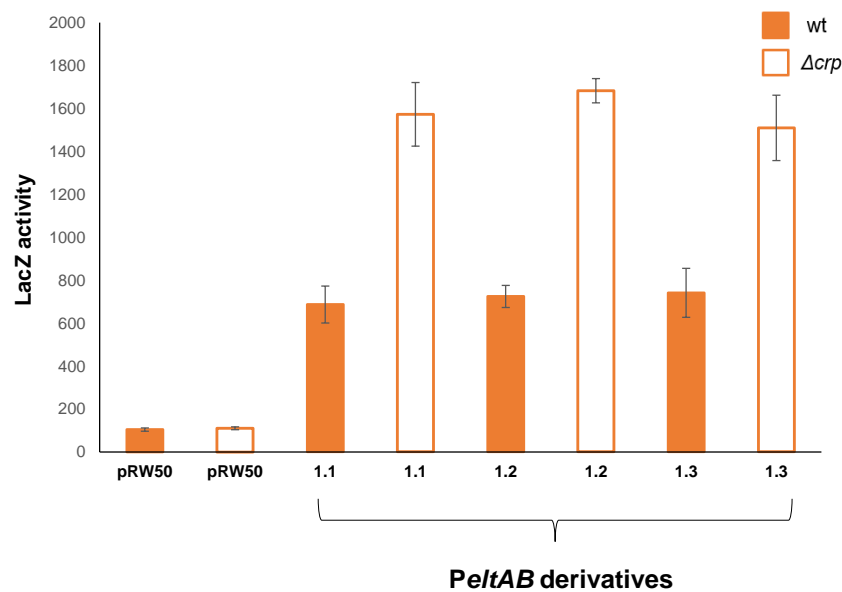
Figure 4.8 Truncation/mutation of the CRP sites at PeltAB does not alter CRP-dependent regulation

A) Variants of the P*eltAB* promoter cloned into pRW50 and subsequently used for truncation/mutation analysis. P*eltAB* 1.1 is the full length 359 bp P*eltAB* promoter containing all three CRP sites defined by Bodero and Munson (2009). P*eltAB* 1.2 is a 118 bp truncated variant, containing only the downstream CRP site. P*eltAB* 1.3 is a derivative of P*eltAB* 1.2 with a mutated CRP site.

B) β-galactosidase activity of P*eltAB* variants cloned into pRW50, in M182 wildtype and M182 Δ*crp* backgrounds. Cells were grown to mid-log phase (OD$_{650}$=0.3-0.6). Standard deviation is shown for three biological replicates. Activity shown is in Miller units.

4.6 Discussion:

4.6.1 Regulation of ST expression in ETEC H10407

Here we demonstrate that the two heat-stable toxin promoters of H10407 are directly regulated by CRP. The P*estA2* promoter of plasmid p948 is activated by CRP from a class I position. The architecture of the P*estA1* promoter on plasmid p666 appears to be similar to P*estA2*, however, there are differences that result in P*estA1* being repressed by CRP and having an overall lower activity than P*estA2*. The trajectory mutagenesis showed the difference in P*estA1* activity results from a combination of differences but the P*estA1* -35 element is key for repression by CRP.

Repression mediated by a single CRP bound at position -58.5 relative to the transcription start site has not previously been reported. However, repression by transcription factors bound with incorrect helical phasing (as defined by Gaston *et al.,* (1990)) has been reported. At the *Bacillus subtilis* φ29 phage early A2c promoter, the phage-encoded regulator p4 normally binds 39 bases upstream of the transcription start site. Binding of RNA polymerase re-positions p4 to a site centred 71 bases upstream of the transcription start site, which represses transcription activation (Monsalve *et al.,* 1998). The α-CTD-p4 interaction is critical for repression, so presumably p4 prevents RNAP from escaping the promoter. At class I promoters, where the CRP site is positioned at -59.5, an α-CTD protomer sits between CRP and σ4 bound at the promoter -35 element (Ross *et al.,* 2003). Thus, one possible explanation might be that replacing the -35 element of P*estA2,* with that from P*estA1*, results in subtle changes in the positioning of the RNAP holoenzyme. This could constrain the positioning of α-CTD so that i) CRP binding to its site displaces α-CTD and prevents RNAP recruitment or ii) unproductive contacts are made by α-CTD.

4.6.2 CRP indirectly regulates the heat-labile toxin

Bodero and Munson (2009) proposed three CRP sites upstream of P*eltAB*, which negatively regulate *eltAB* expression. However, the authors did not link occupation of these sites to P*eltAB* repression by CRP. Moreover, it is unlikely that i) CRP centred at position -261 interacts with transcriptional machinery, and ii) that such poor CRP sites would be able to prevent RNAP binding to the promoter.

4.6.3 A model for enterotoxin production in ETEC H10407

Figure 4.9 is a simple model for CRP-dependent regulation of ETEC enterotoxins. CRP directly activates *estA2* and directly represses *estA1*. CRP binding alone at the promoters of these genes is sufficient for the observed regulation. Repression of *eltAB* expression must be indirect since CRP does not bind P*eltAB in vitro*, and the CRP sites are not required for repression *in vivo*. From a wider perspective, one might speculate that initial intoxication of epithelial cells by LT early in the infection process would lead to increasing levels of cAMP in the gut lumen. This would stimulate a switch from LT production to ST production mediated by CRP. Co-opting global transcriptional regulators, such as CRP and H-NS, is an effective strategy to control toxin expression, without major re-wiring of regulatory networks. For example, it is already known that CRP upregulates many genes important for mouse intestinal colonisation (Chang *et al.,* 2004). Futher more, use of CRP as a regulator of toxin expression has also been observed before. For example, the *pet* promoter of EAEC 042, is dependent on CRP for activation (Rossiter *et al.,* 2011). Additionally, exotoxin A production in *Pseudomonas aeruginosa* is dependent on Vfr, a CRP homologue (Kanack *et al.,* 2006, West *et al.,* 1994).
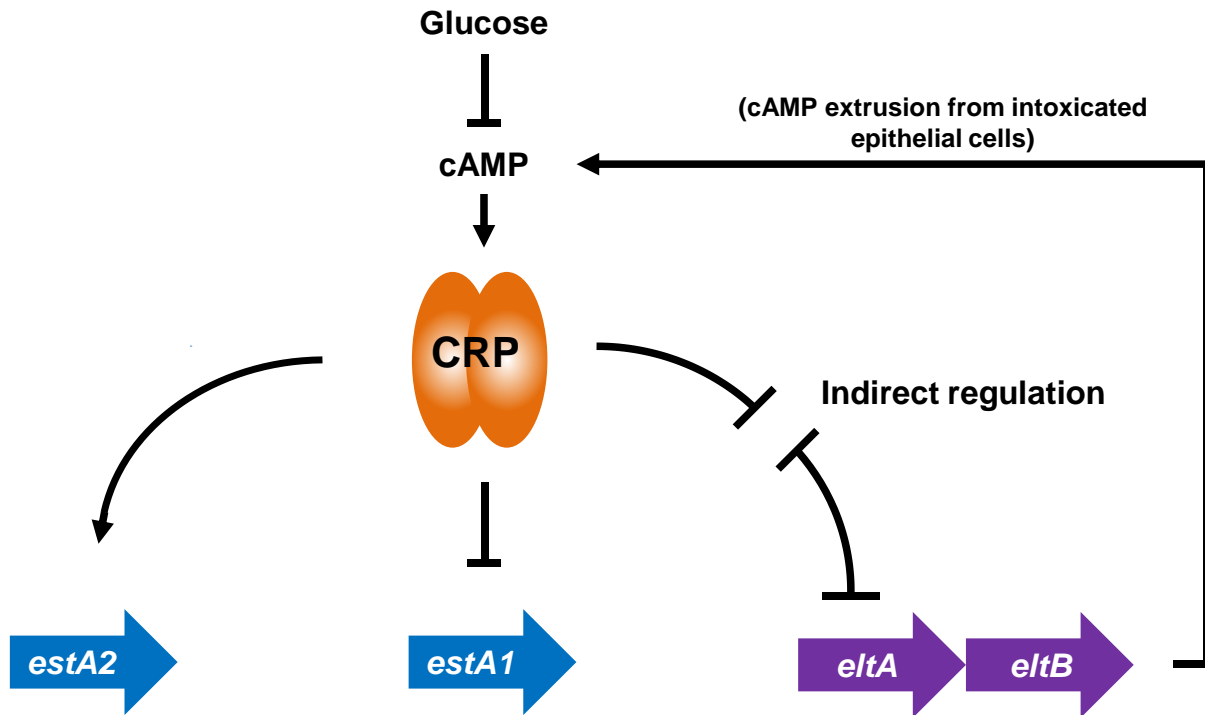
Figure 4.9 Proposed models for CRP-dependent enterotoxin regulation in ETEC

H10407

A model for CRP-dependent regulation of enterotoxin production in ETEC H10407. Arrows indicate positive regulation. Blunt-ended lines indicate repression. Heat-stable toxin genes are shown in blue, the heat-labile toxin gene is shown in purple.

Chapter 5

Characterisation of non-canonical CRP targets in ETEC

H10407

5.1 Introduction

Our bioinformatic screen of plasmids p948 and p666 also a predicted CRP binding site at the *aatPABC* locus (Figure 5.1 and Figure 5.2). The *aatPABCD* operon has been previously characterised in enteroaggregative *E. coli* (EAEC) and encodes the Aat type I secretion system (Nishi *et al.,* 2003). The Aat system is responsible for the secretion of dispersin, a small 10.2 kDa hydrophilic protein that associates with AAF/I and AAF/II fimbriae to maintain their correct positioning (Figure 5.2). In ETEC, the 12.4 kDa CexE protein is a secreted dispersin-protein encoded by the *cexE* gene which is genomically associated with the *aatPABC* locus (Crossman *et al.,* 2010, Pilonieta *et al.,* 2007). Unusually, the CRP site (target '2', from our bioinformatics screen) was located at the 3' end of the *aat* operon, centred 19 bp downstream of *aatC*, and bound CRP with high-affinity *in vitro*. Given the central role of CRP in controlling the expression of ST and LT we were intrigued by the possibility that CRP may also influence expression of the Aat system. The unusual position of the CRP at this locus suggests that i) CRP acts via an unusual mechanism or ii) this CRP site does not have a regulatory function. In this chapter, we have investigated these possibilities.

Our starting point was to search the 3' end of the *aatPABC* operon for an unidentified promoter that may exist in the vicinity of the CRP site. Hence, we searched 100 bp either side of the CRP site for sequences that resembled -10 or -35 hexamers. A potential -10 element with the sequence 5'-TAACCT-3' was identified downstream of the CRP site (highlighted red in Figure 5.1).

ATAA**AGTGA**TAAAAA**TCACA**TAAAATTTTTATTAAAAGGATA**TAACCT**TCATATCACTTGTAATTAAAT

TTGTGTCATGGGTTACCATAACAACTGTTCCACCTTGTTTATAATACTTAAGAAGTATATCCATG
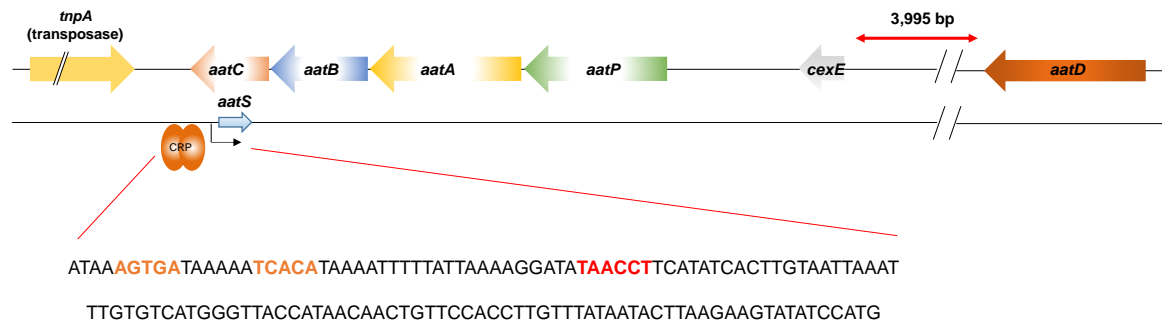
Figure 5.1 Genetic organisation of the *aat* locus, and the position and sequence of target

'2' on plasmid p948 of ETEC H10407

Black lines represent DNA, coloured arrows show genes, the CRP site is shown in orange. The sequence of the CRP site is shown in orange below. The potential -10 element is shown in red. The predicted open reading frame is shown in blue.
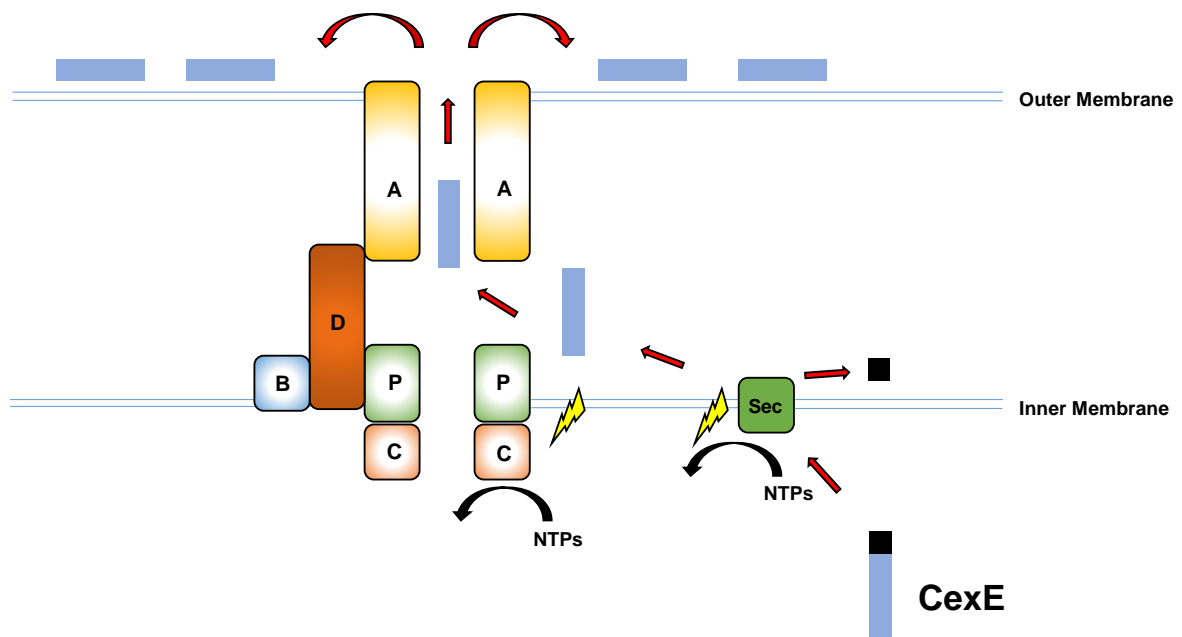
Figure 5.2 Predicted function of the ETEC Aat system based on Aat system of EAEC

Subunits are colour coded with genes in Figure 5.1., CexE is shown as a blue rectangle. Red arrows show the pathway for CexE secretion. The black box on CexE represents a signal peptide. Yellow bolts and black arrows highlight NTP hydrolysis. Adapted from Nishi *et al.,* 2003.

5.2 CRP target '2' lies upstream of a promoter driving transcription anti-sense to the

***aat* operon**

Our next aim was to characterise the promoter lying adjacent to the CRP site. Thus, the 134 bp

fragment shown in Figure 5.1 (that includes the predicted promoter and CRP site) was cloned

into pRW50. M182 cells and the Δ*crp* derivative were transformed with the resulting plasmid

construct. RNA was extracted from log-phase cultures and used as a template for primer

extension assays to i) determine the promoter transcription start site and ii) to determine if the

promoter was CRP dependent. A 166 nucleotide primer extension product was observed (lane

5 in Figure 5.3A) which could be mapped to a thymine residue 39.5 bases downstream of the

CRP site (marked as "+1" in Figure 5.3B). However, note that this transcription start site is 5

bp, rather than 7 bp, downstream of the -10 element. Hence, the CRP site is in a perfect class

II position with respect to the proposed -10 and -35 elements. Consistent with this, almost no

primer extension product was observed in experiments using RNA from Δ*crp* cells, (lane 6 in

Figure 5.3A), suggesting that CRP activates transcription. Intriguingly, we were also able to

identify an open reading frame downstream of this possible promoter using NCBI ORF finder,

which was located within the coding sequence of *aatC* (Figure 5.1). For convenience, we

named this potential open reading frame *aatS* and the promoter P*aatS*.

A
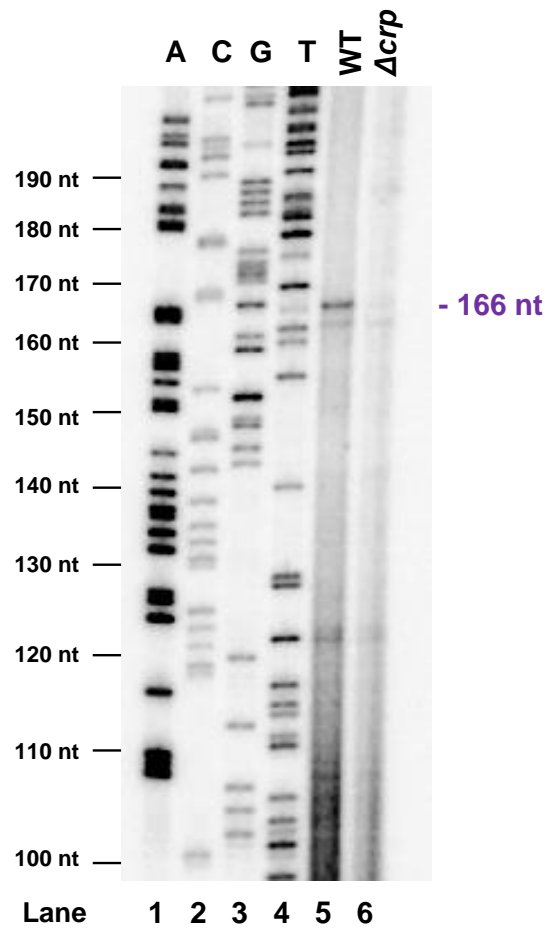
B
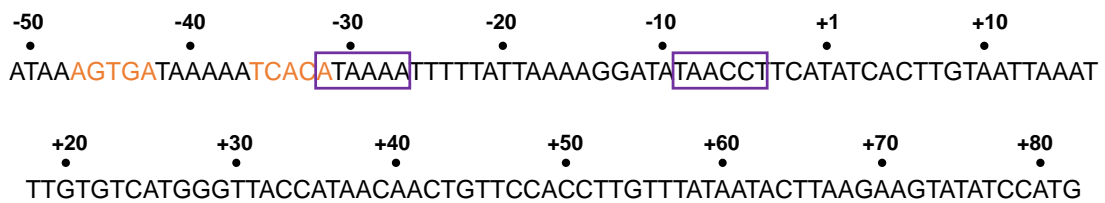
Figure 5.3 Characterisation of the promoter adjacent to target '2' on p948

A) Primer extension analysis of promoter adjacent to target '2'. Lane 5 shows the primer extension product generated from RNA isolated from wildtype cells, lane 6 shows the primer extension product generated from RNA isolated from Δ*crp* cells which labelled in purple. Lanes 1-4 are M13 T7 sequencing reactions ('A', 'C', 'G', and 'T') used to calibrate the gel.

B) Mapping of the transcription start site of the promoter adjacent to target '2'. The transcription start site is labelled as '+1', the proposed -35 and -10 elements are highlighted in purple boxes. CRP half sites are shown in orange.

5.2.1 CRP binds to a site centred at position -39.5 relative to the P*aatS* transcription start site to activate transcription

Next, we used DNAse I footprinting to confirm CRP bound to the predicted site upstream of P*aatS*. The 134 bp P*aatS* DNA fragment was cloned into pSR and a 212 bp *Aat*II-*Hin*dIII fragment was excised from the plasmid for use in DNAse I footprinting experiments. The footprint is shown in Figure 5.4. Lane 1 shows DNAse I cleavage patterns produced at P*aatS* in the absence of CRP. Alterations to this pattern are evident in lanes 2-6 as increasing CRP concentrations are added. CRP induced the appearance of three hypersensitive bands (starred) and protection of two regions of DNA (underlined) (Figure 5.4). We tested for direct CRP-dependent activation using the construct carrying P*aatS* as a template for *in vitro* transcription. We expected the transcript generated from P*aatS* to be 169 nucleotides long. A band of this size was observed in the products of the *in vitro* transcription experiment. Production of this transcript was dependent on CRP (Figure 5.5A). Similar results were obtained in β-galactosidase assays using M182, and the Δ*crp* derivative, transformed with pRW50 containing the P*aatS*::*lac*Z fusion (Figure 5.5B).
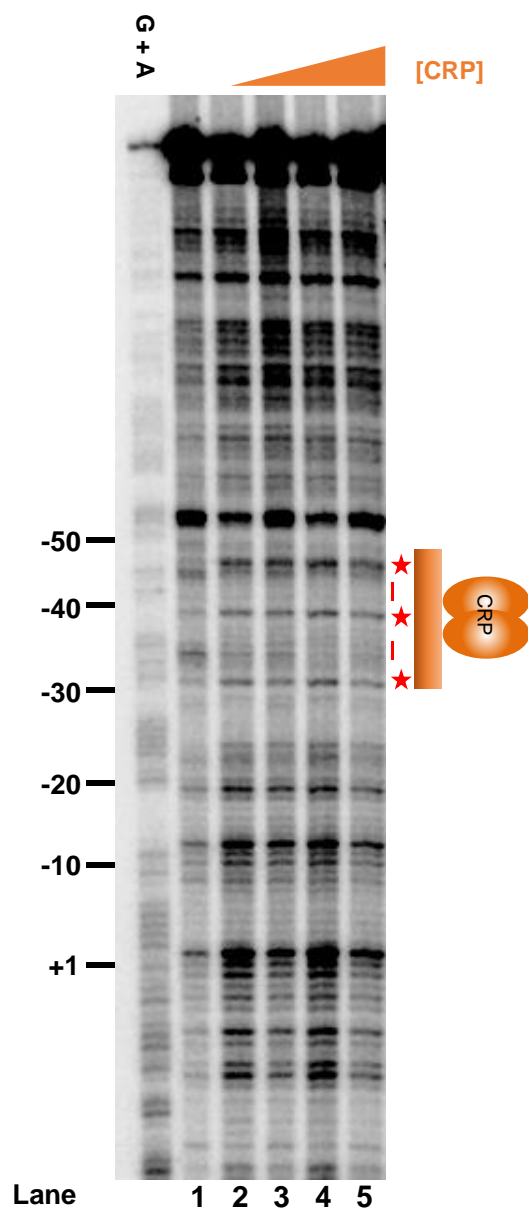
Figure 5.4 DNAse I footprinting analysis of P*aatS*

G + A is a Maxim/Gilbert G + A ladder, which shows the position of pyrimidine bases in the sequence. Numbering refers to distance from the transcription start site, as determined by primer extension. Each band represents a site at which DNA has been cleaved. Lane 1 shows the cleavage pattern of P*aatS* when exposed to DNAse I, in the absence of CRP. Lanes 2 to 5 show cleavage patterns of P*aatS* when exposed to increasing concentrations of CRP. Final CRP concentrations used in lanes 2-5 respectively are; 0.35 µM, 0.7 µM, 1.05 µM and 1.4 µM. All reactions contain 0.2 mM cAMP. Additionally, all reactions contained 12.5 µg/ml Herring Sperm DNA as a non-specific competitor. Zero point two units of DNAse I were used per reaction. Reactions were run on a 6 % (*w/v*) denaturing gel. Red stars highlight hypersensitive bands which form in the presence of CRP, red dashes highlight protection of the DNA from DNAse I cleavage.

150

A

169 nt *aatS* transcript →

108 nt *RNA I* →

[CRP]

(aatS/ RNAI)   0.3   0.6   0.8

B

LacZ activity activity (Miller units)

Wildtype   **Δ*crp***

P*aatS*

SD   lacZ

91   **50**
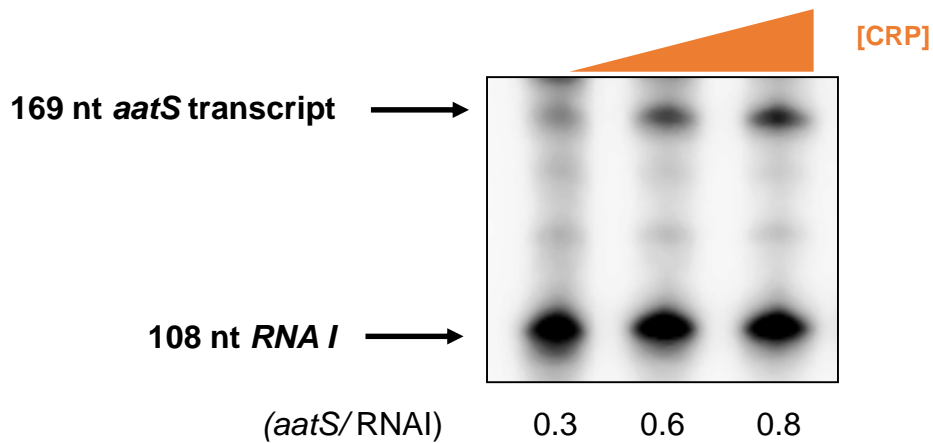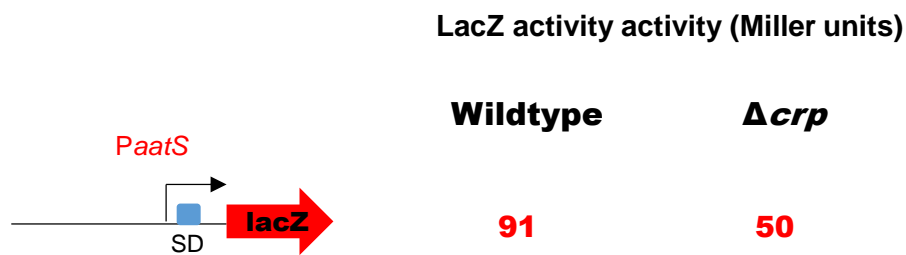
Figure 5.5 CRP activates transcription from P*aatS in vivo* and *in vitro*

A) The panel shows *in vitro* transcription products of P*aatS* cloned upstream of a *λoop* terminator in pSR. RNA polymerase was used at a concentration of 400 nM, CRP was used at a concentration of 219 nM if 438 nM. Reactions were run for 10 minutes at 37 °C before being stopped. Transcription products were run on a 6 % (*w/v*) denaturing gel. The 169 nt P*aatS* transcript and 108 nt RNAI control transcript are labelled. Corresponding relative intensities (P*aatS*/RNAI) are displayed below the panel.

B) β-galactosidase activity of P*aatS* cloned into pRW50, in M182 wildtype and M182 Δ*crp* background. Cells were grown to mid-log phase (OD$_{650}$=0.3-0.6). Standard deviation is less than 15 % for three biological replicates. Activity shown is in Miller units.

5.3 P*aatS* lies upstream of a predicted 62 amino acid open reading frame

As previously mentioned, we identified a small open reading frame downstream of P*aatS*, within the coding region of *aatC*. This small ORF, potentially encoding a 62 amino acid protein, is located 82 bp downstream of P*aatS* transcription start site, within the coding region of *aatC*. Interestingly, this open reading frame encodes a conserved domain of unknown function (DUF1602) found in 37 hypothetical genes in the NCBI database. Strikingly, of these 37 genes, many were genetically associated with operons implicated in transport functions. Furthermore, the 37 genes frequently overlap adjacent genes. Remarkably, possible AatS homologues are found in multiple phylogenetic groups. In addition to finding homologues in Gram positive and Gram negative bacteria (including α, β, δ and γ proteobacteria), we also identified homologs in eukaryotes such as the fungal pathogens *Sclerotinia sclerotiorum* and *Trichophyton rubrum*. Figure 5.6A is an alignment of the predicted AatS sequences. Figure 5.6B shows the genomic position of the hypothetical genes. The predicted *aatS* homologue frequently overlaps a gene encoding the ATP-subunit of a transporter. Note that *aatS* was not found in all sequenced ETEC isolates. For example, in strains E42377A and 1392/75, the start codon of *aatS* is absent (Figure 5.6C). In strain B7A, we found an open reading frame encoding a 90 amino acid protein 85.5 % similar to AatS from H10407 (Figure 5.6D). Note that, like *aatS* from H10407, *aatS* from B7A also lies completely within the *aatC* coding region (Figure 5.6E). The high-affinity CRP site upstream of the *aatS* open reading frame (shown in orange in Figure 5.6C) is not well conserved amongst ETEC isolates.

A

```
AatS       (ETEC H10407)                          ------------------------MIFFLLLLSKLAVGSSASSILGFFIKALASATRCFSP 37
Adeh_1080 [Anaeromyxobacter dehalogenans 2CP-C]   -----MCVTTTAVIPIRTFRSASKAWIPSAVCGSRLPVGSSASSSFGWRHSARARATRCCSP 57
BURPS1710b_2717 [Burkholderia pseudomallei 1710b] --------------------------------MPVGSSASTQAGLVTSARAIATRWRSP 27
RD1_1447  [Roseobacter denitrificans OCh 114]     -----------MVMPSSASSIMVSSTSLIISGSSAEVGSSNNMIRGFMQSERAMATRCCCP 50
Spaf_0664 [Streptococcus parasanguinis FW213]     -------MIPNSSLILFN-----NSIIFCSLSASKLPVGSSAINISGLLIRALAKATLCFSP 50
XNC1_2169 [Xenorhabdus nematophila ATCC 19061]    --------MTIVVPLLFNSLS--DDKILSPFTLSKLPVGSSANKIAGLDIIALAIATRYC-- 52
SYNPCC7002_A1750 [Synechococcus sp. PCC 7002]     MASSTSWVIKRILISVCSQISRRNCCIFSLVRTSKAPKGSSINKIFGWEARARAIATRCFIP 62
```

B

Streptococcus parasanguinis FW21 (NC_017905.1)   **67% identity over a 27 AA region**
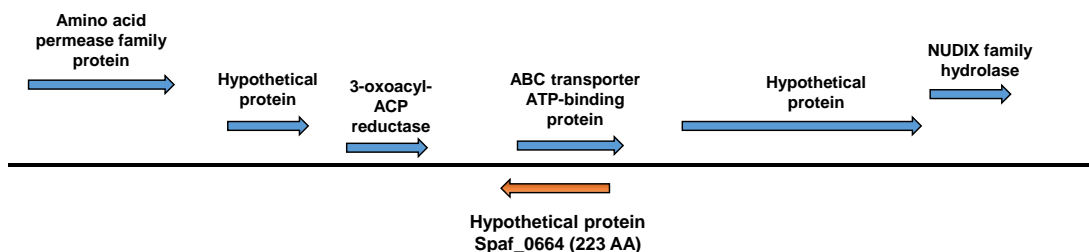
Synechococcus sp. PCC 7002 (NC_010475.1)   **47% identity over a 49 AA region**
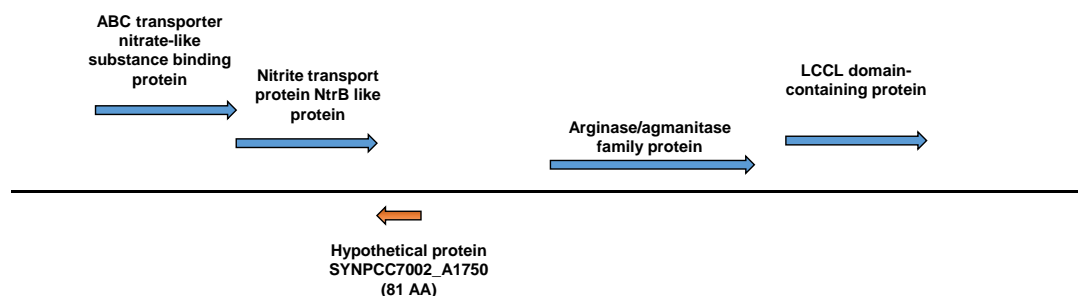
Burkholderia pseudomallei 1710b (NC_007434.1)   **57% identity over a 30 AA region**
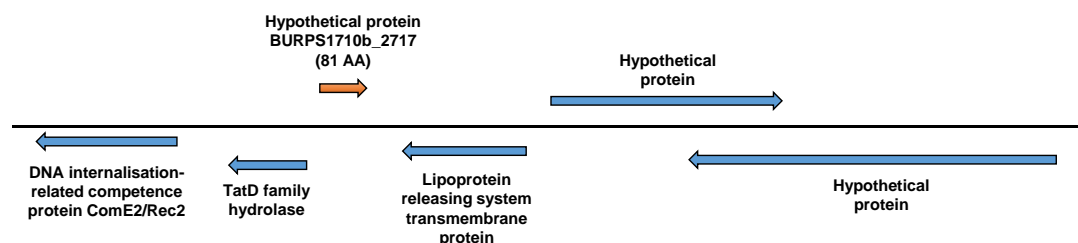
Anaeromyxobacter dehalogenans 2CP-C NC_007760.1   66% identity over a 32 AA region

Lysyl-tRNA synthetase
Hypothetical protein
Surface antigen (D15)
Outer membrane chaperone Skp
Hypothetical protein Adeh_1080 (190 AA)
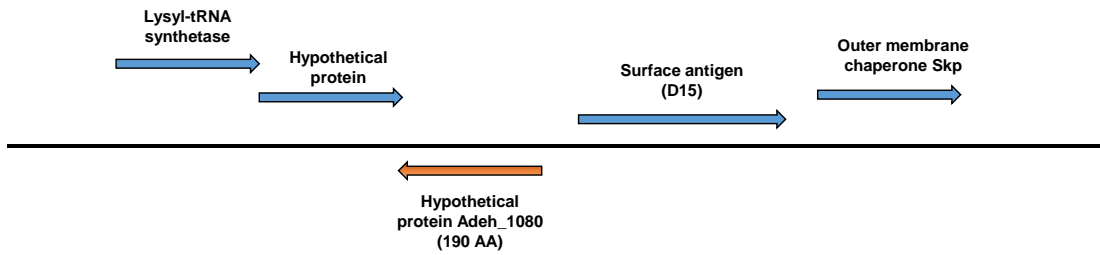
C

```
E24377A    AAAGACCTGAACTTTCAGAACATCATTTTATTGT-TCGCATTAGTA--ATATATAGATAT
1392/75    AAAGACCTGAACTTTCAGAACATCATTTTATTGT-TCGCATTAGTA--ATATATAGATAT
H10407     AGTTGCCGTAGTTATTC----GTGAGGGAAAAATCCAATCATACTTTTTGTATGGTCGCT
B7A        GATCACCCAATCACTACCACCCTAGAAGGTGATTTAGGAATAACAATGTTATATTGTTCT
                 **  *     *         *           *         *                *

E24377A    CTAACCACTGAGAATTACCCATATTAAAACGATAAATTTAAAGCGTTATTACTTTCATAT
1392/75    CTAACCACTGAGAATTACCCATATTAAAACGATAAATTTAAAGCGTTATTACTTTCATAT
H10407     GGCGAAATAAAGTGATAAAAATCACATAAAATTTTTATTAAAAGGATATAACCTTCATAT
B7A        GGCGGAATAAAGTGATAAATAATCTATAGAGAAATTATCAAAAGGAAATGACTCTCATCT
                 *    **   **     *     * *       * ***   *   ** **   **** *

E24377A    CTGTTGTGATTATATTTGTATCATGCGTTATCATAACAACTGTTCCACCGCGCTTATAAT
1392/75    CTGTTGTGATTATATTTGTATCATGCGTTATCATAACAACTGTTCCACCGCGCTTATAAT
H10407     CACTTGTAATTAAATTTGTGTCATGGGTTACCATAACAACTGTTCCACCTTGTTTATAAT
B7A        CATTAGTAATTAAATTTGTGTCATGAGTAACCATAACAACAGTTCCGCCTTGTTTATAAT
                *   *  ** ****  ****** ***** **  *  ********* ***** **   * *******

E24377A    AATTAAGTATTACATTCATAACTTTCTCTTTATTATCATTATCCAAACTCGCTGTTGGTT
1392/75    AATTAAGTATTACATTCATAACTTTCTCTTTATTATCATTATCCAAACTCGCTGTTGGTT
H10407     ACTTAAGAAGTATATCCATGATTTTTTTTTTTGTTATTATTATCTAAGCTTGCCGTTGGTT
B7A        ACTCAAGAAGTATATCCATGATTTTTTTCTTTGTTATTTTTATCTAAACTTGCTGTTGGTT
                *  *  *** * *   ** ** ** ***  * ***  * *** ****  ***** **  ** **  *******
```

D

```
AatS (H10407)   MIFFLLLLSKLAVGSSASSILGFFIKALASATRCFSPPDSKAADASGDMPCNPTFSSSFKTT--------------------------
AatS (B7A)      MIFSLLFLSKLAVGSSASNILGFFIKALASATRCFSPPDNRAADTSGDMPCNPTFSSAFKIILLHSFCFFNSIIDSKFLRIVCFSFISES
                *** **:*********** .*****************  .:***:*************:**
```

154

E

Escherichia coli B7A plasmid pEB4 (CP006002.1)   **85.5% identity over a 62 AA region**
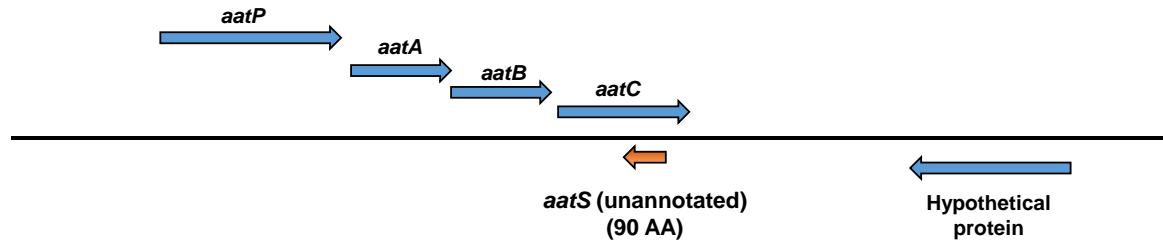
Figure 5.6 Conservation (A) and genomic location (B) of AatS homologues found in other bacteria, and occurrence of *aatS* and P*aatS*-like sequences amongst ETEC strains (C-E)

A. Alignment of predicted protein sequence of AatS with homologues proteins containing DUF1602 determined by NCBI BLAST. Only regions of homology are shown. Alignments were done using Clustal Omega.
B. Genomic context of some *aatS* homologues. In each case, the species, NCBI genome accession number, and the degree of amino acid identity between the protein sequence and AatS is shown. Coloured arrows depict genes, arrows shaded orange represent *aatS* homologues. Protein functions are given above each arrow if known.
C. Alignment of P*aatS*-like sequences in different ETEC strains. The figure shows an alignment of the 3' ends of *aatC* (in the reverse, complemented orientation) from different ETEC isolates. The stop codon of *aatC* is highlighted blue, the predicted start codon of *aatS* is shown in a red box. CRP half-sites are shown in orange, mismatches from the consensus are underlined within these sites. The transcription start site of *PaatS* in H10407 is shown in red and is underlined.
D. Alignment of predicted protein sequence of AatS from strain H10407 with that from ETEC strain B7A. Alignments were done using Clustal Omega.
E. Genomic context of *aatS* from strain B7A. The degree of amino acid identity between the B7A protein sequence and AatS is shown. Coloured arrows depict genes, arrows shaded orange represent *aatS* homologues.

5.3.1 The *aatS* leader contains translation initiation signals

The 5' end of the putative *aatS* mRNA contains the sequence 5'-UAAGAAGU-3', (Figure 5.7A) which bears similarity to the consensus Shine-Dalgarno sequence (consensus UAAGGAGGU) (Shine & Dalgarno, 1974), and is concordant with the sequence logo compiled by Shultzaberger *et al.* (2001) from 4122 translation initiation sites. In addition, the predicted *aatS* ribosome binding site is located 5 bp upstream of the *aatS* start codon, a position optimal for translation initiation (Chen *et al.,* 1994).

To investigate whether the Shine-Dalgarno sequence was functional, we created translational *aatS::lacZ* fusions and explored the effects of mutating the ribosome site on *aatS* expression. Fortuitously, the 134 bp P*aatS* DNA fragment used for characterisation of P*aatS* included the Shine-Dalgarno sequence, and start codon of *aatS* (Figure 5.7A). Additionally, we generated a derivative of this fragment where the entire Shine-Dalgarno sequence was inverted (Figure 5.7B). We cloned both fragments into pRW225 (Islam *et al*., 2012), to translationally fuse *aatS* to *lacZ*. This vector contains a functional *lac* operon downstream of the multiple cloning site, but unlike pRW50, does not contain a Shine-Dalgarno sequence upstream of the *lacZ* gene. Hence, supplied Shine-Dalgarno sequences included on cloned fragments can be tested for functionality. Plasmid constructs were transformed into M182 and the Δ*crp* derivative, and LacZ activities were determined in the transformants (Figure 5.7C). We found LacZ activity was significantly reduced in lysates of cells transformed with pRW225 derivatives where the *aatS* ribosome binding site was inverted.

A

```
    -50         -40         -30         -20         -10         +1          +10
     •           •           •           •           •           •           •
ATAAAGTGATAAAAATCACATAAAATTTTTATTAAAAGGATATAACCTTCATATCACTTGTAATTAAAT

    +20         +30         +40         +50         +60         +70         +80
     •           •           •           •           •           •           •
TTGTGTCATGGGTTACCATAACAACTGTTCCACCTTGTTTATAATACTTAAGAAGTATATCCATG
```

B

```
    -50         -40         -30         -20         -10         +1          +10
     •           •           •           •           •           •           •
ATAAAGTGATAAAAATCACATAAAATTTTTATTAAAAGGATATAACCTTCATATCACTTGTAATTAAAT

    +20         +30         +40         +50         +60         +70         +80
     •           •           •           •           •           •           •
TTGTGTCATGGGTTACCATAACAACTGTTCCACCTTGTTTATAATACTTTTCTTCAATATCCATG
```

C

Figure 5.7 Predicting expression of *aatS* using translational fusions

A. Sequence of P*aatS* cloned into pRW225 containing the native Shine-Dalgarno sequence, shown in blue. CRP half-sites are shown in orange, the transcription start site.

B. Sequence of P*aatS*ΔSD cloned into pRW225, containing an inverted Shine-Dalgarno sequence, shown in blue. CRP half-sites are shown in orange, the transcription start site.

C. β-galactosidase activity of P*aatS* cloned into pRW225, in M182 wildtype and M182 Δ*crp* background. Cells were grown to mid-log phase in M9 minimal media supplemented with 1 % (*w/v*) fructose (OD$_{650}$= 0.3-0.6). Standard deviation is shown for three biological replicates. Activity shown is in Miller units. Note that in M182 wildtype cells, CRP was supplied *in trans*, by transforming cells with plasmid pDCRP, which contains the *crp* gene under the control of its own promoter.

5.3.2 Attempts to identify interactions of AatS with subunits of the Aat system

Since *aatS* is preceded by a functional Shine-Dalgarno sequence, our attention turned to possible roles for AatS. First, we sought to determine if AatS might interact with subunits of the Aat system. To explore this possibility, we used the bacterial two-hybrid system (BACTH). We cloned each Aat subunit separately, in-frame, into pKT25, pUT18 and pUT18C. Figure 5.8 is a schematic showing the different fusion proteins produced from these plasmid contructs. Furthermore, we used Spoctopus software (Viklund *et al.,* 2008) to look for any possible signal peptides. The N-terminal region of AatS was predicted to contain a signal peptide. Thus, we generated a shortened version of *aatS*, named '*aatS* short', which was also used in the BACTH assays. We first looked for interactions between AatC and AatS, due to their genomic association with each other. However, we were unable to detect LacZ activity significantly above background, and thus no measureable interaction (Figure 5.9A). Consequently, we expanded our study to explore interactions between AatC and other subunits in the Aat system. Again, no interaction signals were detected (Figure 5.9B). Furthermore, we could not detect interactions between any of the well-characterised Aat subunits (Figure 5.10A-D). We conclude that the Aat system is not tractable for BACTH analysis.
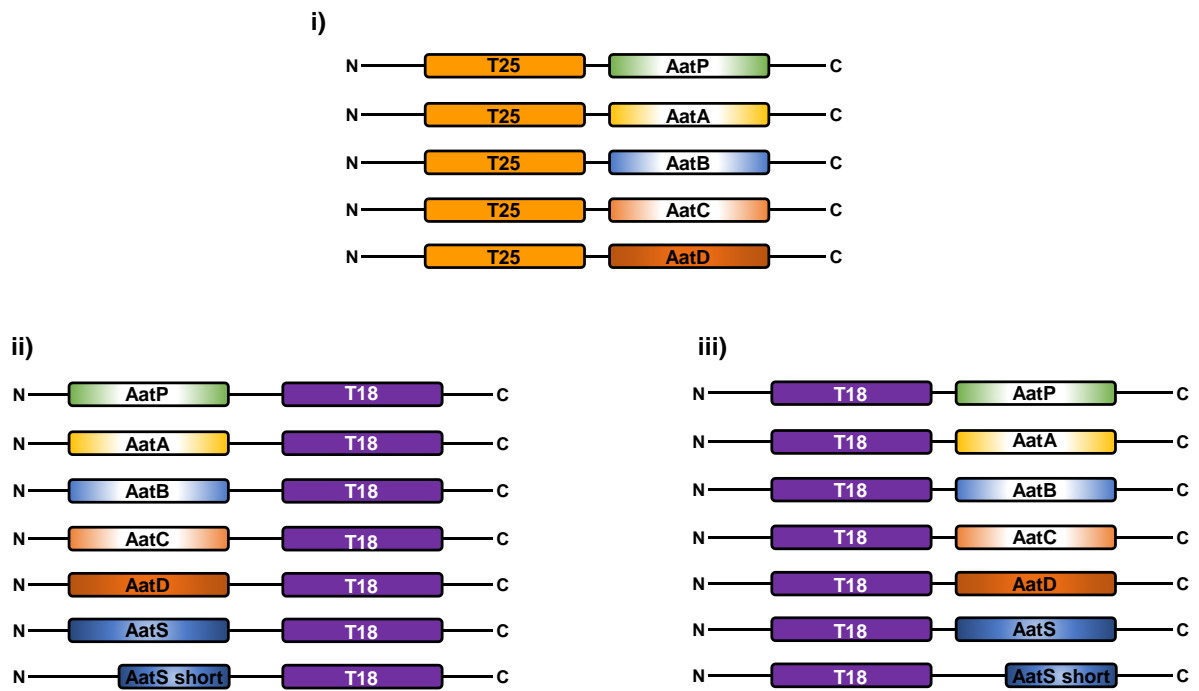
Figure 5.8 BACTH fusions constructed and used for investigating inter-subunit

interactions of the Aat system

Gene fusions were constructed in pKT25 (i), pUT18 (ii), or pUT18C (iii). In experiments, a pKT25 construct was always used with a pUT18 or pUT18C construct containing a different protein.

160
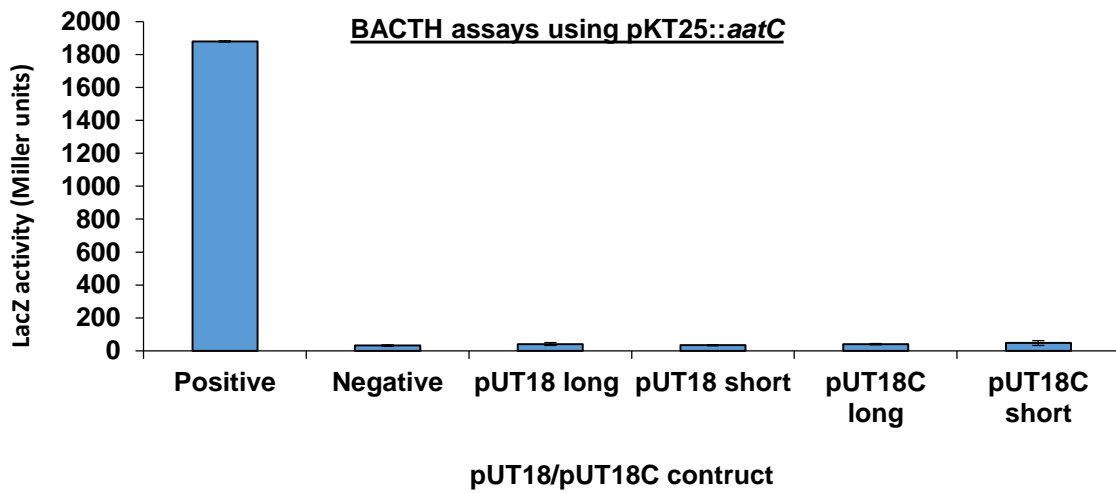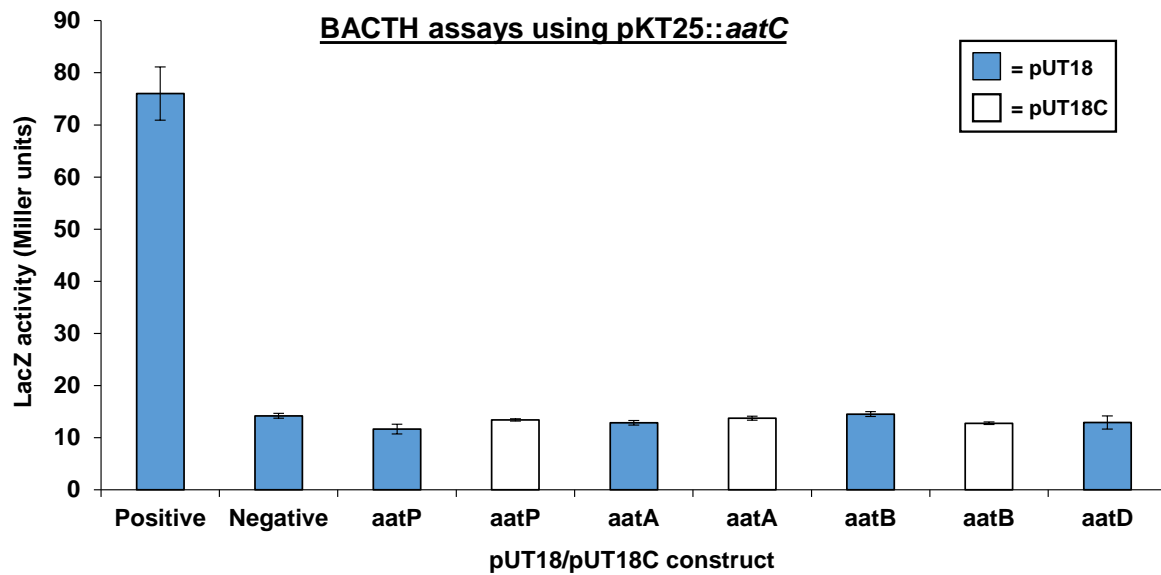
Figure 5.9 BACTH assays measuring interactions between AatC and AatS (A), and AatC and other Aat system components (B)

β-galactosidase activity measured in BACTH assays in BTH101 cells. Cells were grown to mid-log phase (OD_{650}= 0.3-0.6). Standard deviation is shown for three biological replicates. Activity shown is in Miller units. Note that *aatC* was cloned into pKT25, and *aatS*, or other Aat subunits were cloned into pUT18 or pUT18C.
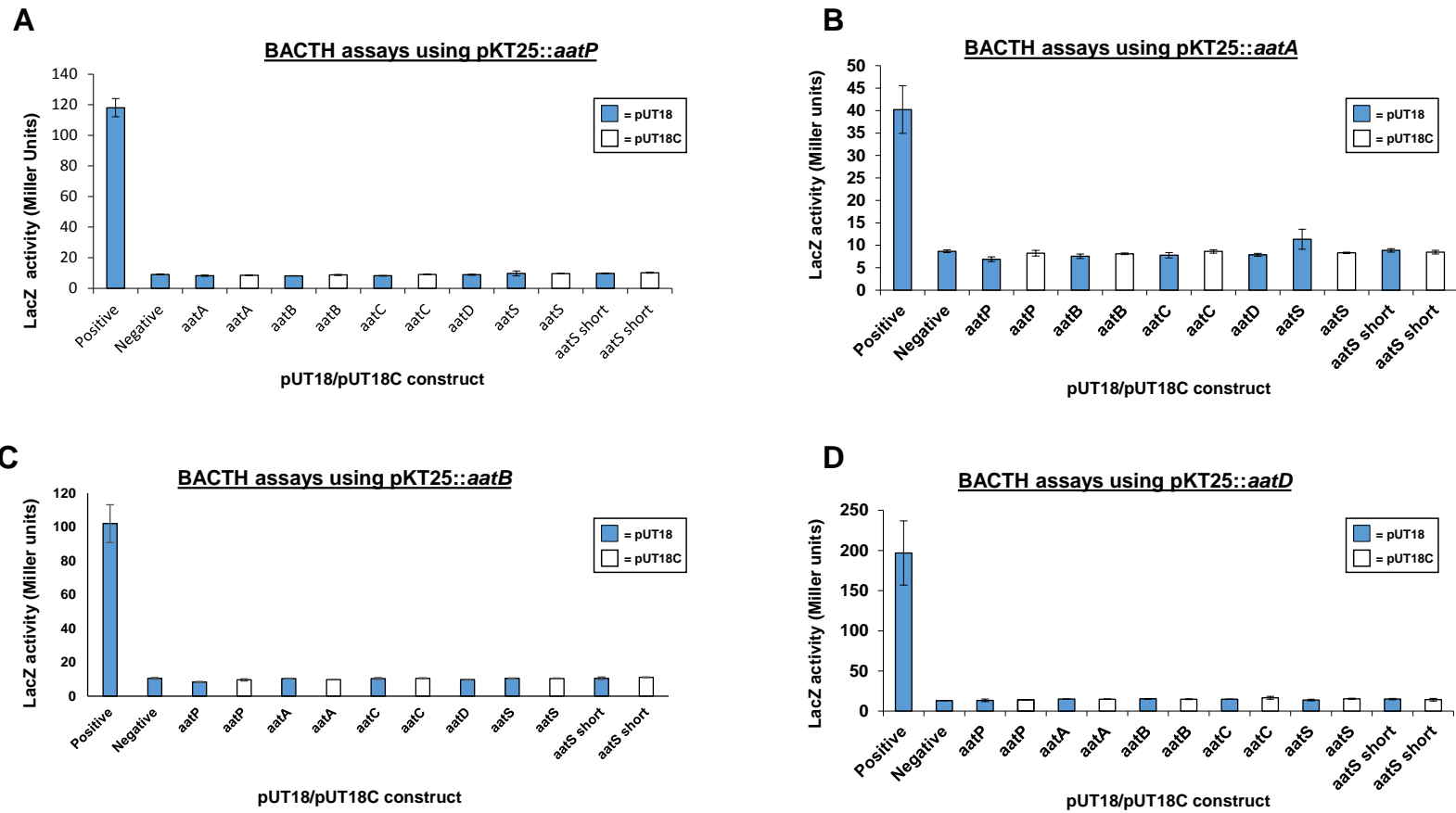
Figure 5.10 BACTH assays measuring interactions between all Aat subunits

β-galactosidase activity measured in BACTH assays in BTH101 cells. Cells were grown to mid-log phase (OD$_{650}$= 0.3-0.6). Standard deviation is shown for three biological replicates. Activity shown is in Miller units. In each panel, a positive and negative control is included. (A-D) are assays were *aatP*, *aatA*, *aatB*, *aatB* or *aatC* were cloned into pKT25, respectively. Other subunits were cloned into pUT18 and pUT18C as indicated.

5.3.3 Attempts to determine the effect of AatS expression on CexE secretion

Since BACTH assays were unsuccessful in yielding clues as to the function of AatS, we decided to investigate the effect of AatS expression on CexE secretion into culture supernatants. We reasoned that this approach would be valid since CexE is presumed to be the substrate of the Aat system in ETEC H10407 (Crossman *et al.,* 2010). In our system plasmid pJ204::*aatS* (Amp^R) encodes *aatS* under the control of the constitutive *lacUV5* promoter. Plasmids pJRH1-4 contain the full *aatPABCD* operon and Shine-Dalgarno sequence. In plasmids pJRH3 and pJRH4, the start codon of the predicted *aatS* open reading frame has been removed. Plasmids pJRH2 & 4 also encodes CexE (the putative substrate for the Aat system). The composition and structure of each plasmid is shown in Figure 5.11A and 5.11B respectively. Note that, in all four constructs, expression of genes in the *aatPABCD* operon is driven by P*lacUV5*.

Constructs were used to transform *E. coli* K-12 JCB387 cells and transformants were grown to stationary phase in M9 minimal media supplemented with 1 % (*w/v*) fructose. Cells were then harvested, and proteins in the culture supernatant precipitated using 10 % (*w/v*) trichloroacetic acid (TCA). Precipitated proteins were run on polyacrylamide gels. Additionally, to detect any CexE protein which had not been secreted, cell lysates were also run. Note that, although the *lacUV5* promoter is subject to LacI binding and repression, the *lacI* locus of JCB387 has been replaced with an X174 phage, and therefore LacI is not expressed. The results of the experiment are shown in Figure 5.12A. We were unable to discern any difference between samples from different constructs. Our ChIP-seq analysis showed that the *aat* operon is bound by H-NS *in vivo*, and hence is likely to be repressed. Hence, we performed a final set of experiments using M182Δ*hns* cells. For these experiments we were able to run purified CexE protein (generously donated by Dr. Tim Wells) alongside our samples. Unfortunately, we were unable to detect CexE in any of the samples (Figure 5.12B).

A

Plasmid	pJRH1	pJRH2	pJRH3	pJRH4
lacUV5	+	+	+	+
cexE	-	+	-	+
aat operon	+	+	+	+
aatS	+	+	-	-

B
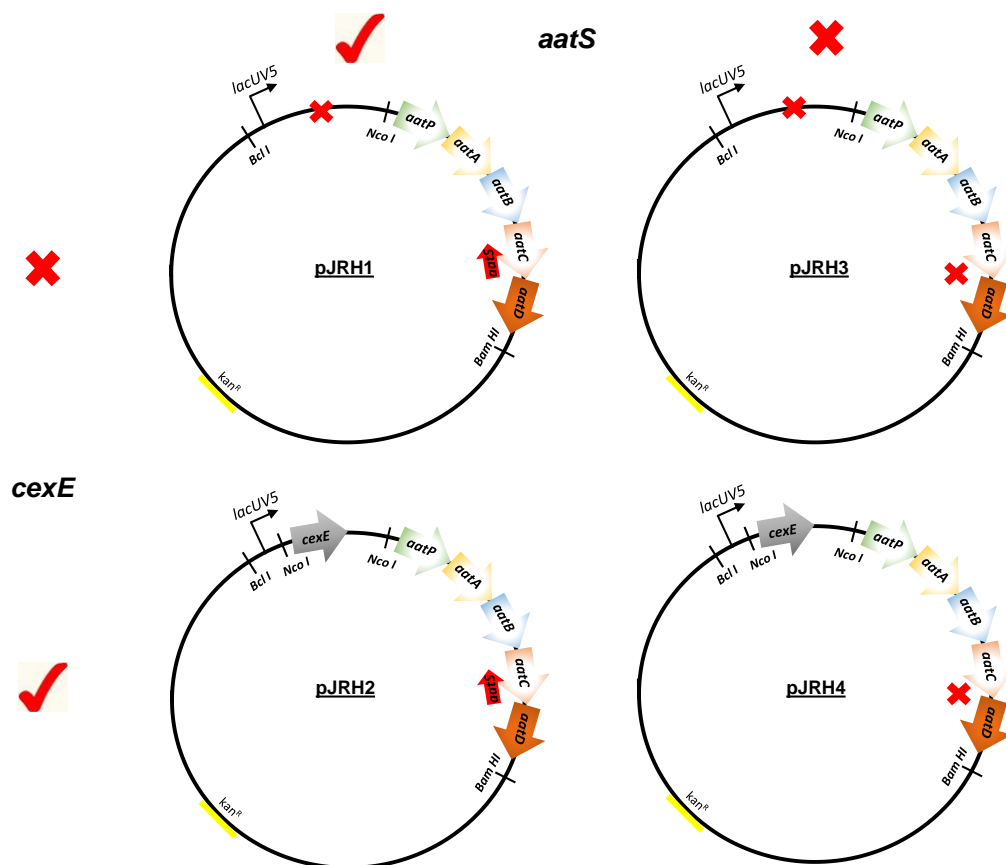
Figure 5.11 Composition (A) and structure (B) of plasmids pJRH1-4

The schematic in (A) shows the composition of each of pJRH derivative. Note that all plasmids contain a *lacUV5* promoter, and the entire *aat* operon. In (B), coloured arrows depict genes, solid lines depict restriction sites.

A
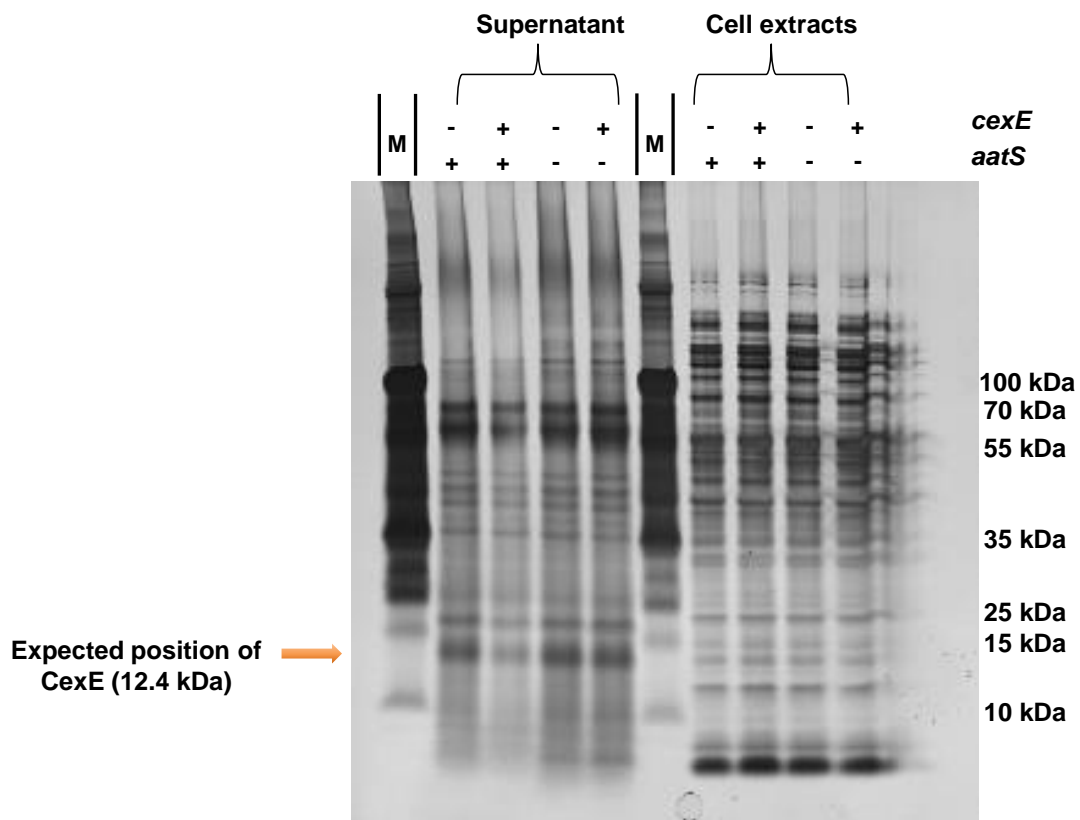
B
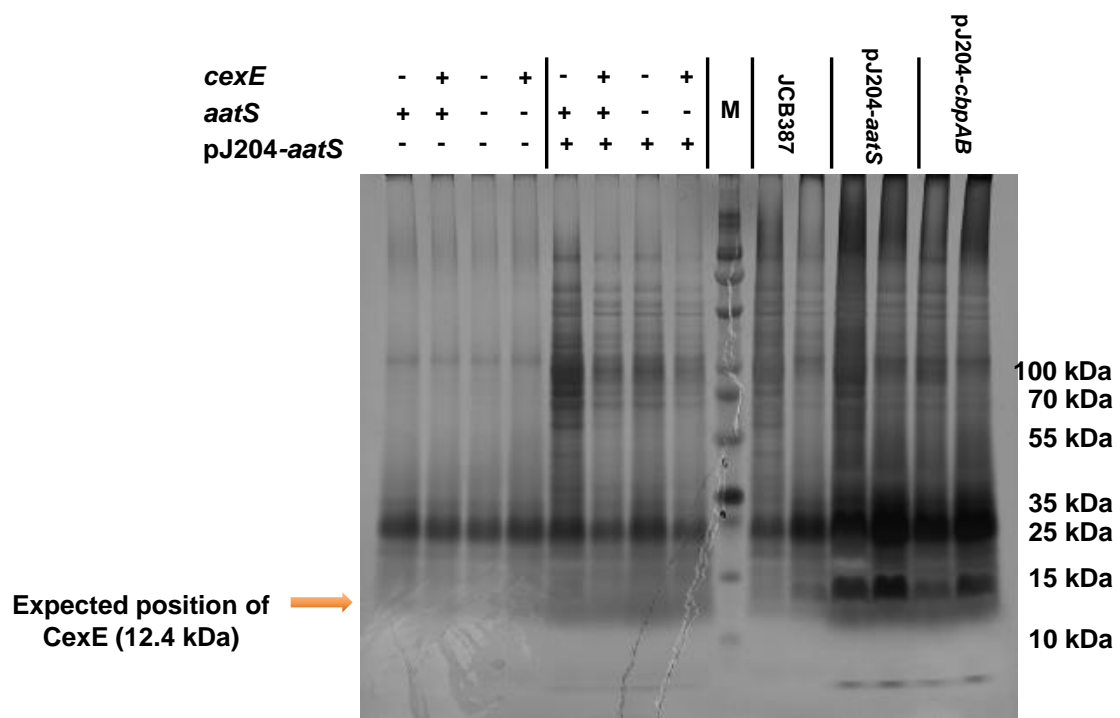
C
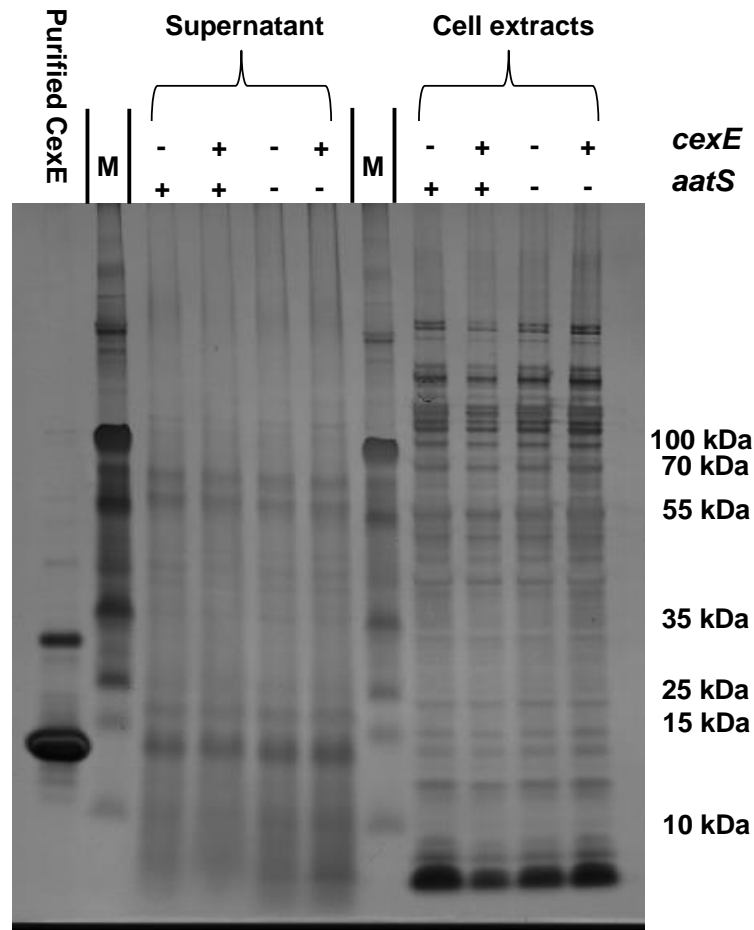
Figure 5.12 TCA precipitation of CexE from culture supernatants and cell extracts

The gels show TCA precipitated protein and cell extract samples extracted from JCB387 cells grown in M9 minimal media supplemented with 1 % (*w/v*) fructose to stationary phase. All gels are calibrated with PageRuler Plus prestained protein ladder (Thermo Scientific) ("M").

A. TCA precipitations and cell extracts from JCB387 cells transformed with a pJRH derivative, genes present in the construct are indicated above in each case.

B. TCA precipitations from JCB387 cells transformed with a pJRH1-4, or transformed with a pJRH derivative and pJ204*::aatS* as indicated above the gel. Additionally, control TCA precipitations are shown to the right of the marker. These include control precipitations from supernatants of JCB387 cells not containing any plasmids, cells containing pJ204*::cbpAB*, or pJ204*::aatS*, without any pJRH derivative.

C. TCA precipitations and cell extracts from M182Δ*hns* cells transformed with pJRH1-4. Ten µl of purified CexE (12.4 kDa) has been run alongside these samples in the well indicated.

5.4 Discussion

We have identified a gene (*aatS*) within the coding sequence of *aatC*. Expression of *aatS* is activated by CRP via a class II mechanism. Our efforts to elucidate a role for AatS have, as yet, not provided insight into AatS function. In any case, the CRP-dependent regulation of P*aatS/aatS* appears to be specific to H10407. Thus, although strain B7A also contains an AatS homologue, there is no CRP site upstream of *aatS*. We speculate that ETEC H10407 represents a snap-shot of evolution in the *aat* gene cluster. Thus different combinations of P*aatS*, the associated CRP site, and *aatS* are found in different ETEC isolates.

Regardless of AatS functionality, the transcription from the P*aatS* promoter, which drives RNAP into the 3' end of the *aatC* gene, is likely to interfere with transcription of the *aatPABC* operon. Elongating RNAP transcribing the *aatPABC* operon could collide with RNAP from P*aatS*, with transcriptional termination driven by the build up of supercoiling between the two RNAP enzymes. This transcriptional interference phenomenon is well documented in the *ubiGmccBA* operon of *Clostridium acetobutylicum* (André *et al.,* 2008). The products *ubiGmccBA* operon convert methionine to cysteine. In the presence of cysteine, an S-box-dependent promoter (which senses S-adenosyl-methionine, a co-enzyme in sulphur metabolism) is activated, and drives transcription into the 3' end of the *ubiGmccBA* operon, reducing the abundance of full length *ubiGmccBA* mRNA trancripts.

Currently, it is difficult to speculate on a role for AatS, since the domain of unknown function (DUF1602) has not been studied in detail. Small proteins can be encoded on anti-sense RNAs. An example of this is SgrT, which is encoded by the small 227 nucleotide non-coding RNA SgrS, from *Salmonella typhimurium* (Wadler and Vanderpool, 2007). The 3' region of SgrS binds to the translation initiation region of the *ptsG* mRNA and silences translation. The 5' region of SgrS encodes a 43 amino acid protein called SgrT, which inhibits existing EIIBC^{Glc}, through an unknown mechanism. SgrT expression does not affect *ptsG* mRNA expression or

EIIBC^{Glc} protein abundance, so it has been suggested that SgrT acts as a 'plug' to block glucose import. Hence, the combined action of SgrS and SgrT is thought to reduce the toxic effects of glucose-6-phosphate accumulation in the cell when glycolysis is blocked.

In the case of the *aatPABC* operon, it is possible that the anti-sense P*aatS* promoter and the AatS protein have redundant effects on the expression of the Aat system in ETEC H10407. The fact that the P*aatS* CRP site is not present in all strains might suggest that control of the P*aatS* promoter is in the process of being recruited to (or lost from) the CRP regulon in ETEC. The substrate of the Aat system, CexE, is thought to promote dispersal of ETEC cells through the mucus layer surrounding the gut lumen. Given that *E. coli* strains utilise mucus, and not luminal contents, as a nutrient source, it is possible that penetration of ETEC cells into the mucus lining will modulate P*aatS* activity through metabolic cues sensed by CRP (Le Bouguénec and Schouler, 2011). Downregulation of the Aat system (through P*aatS* transcriptional interference or AatS expression) may be beneficial to ETEC pathogenesis. Perhaps preventing CexE secretion promotes more intimate interactions with intestinal epithelial cells.

Chapter 6

Final conclusions

The aim of this study was to define the regulons of CRP, H-NS and σ^{70} in Enterotoxigenic *Escherichia coli* strain H10407. Here, we show that all three factors are central to virulence regulation. Thus both *estA2* and *estA1* are directly regulated by CRP. We show that repression of *eltAB* by CRP occurs through an indirect mechanism. Intriguingly, all three toxin genes are occupied by H-NS not CRP *in vivo*, which suggests that H-NS may occlude CRP from its binding sites. This was also the case for other characterised binding sites that bound CRP with high-affinity *in vitro*.

Our observations have implications for the treatment of ETEC-mediated disease. Briefly, the treatment for ETEC-mediated diarrhoea is oral rehydration with a 1 % (*w/v*) solution of glucose and 60 mM NaCl (Nalin *et al.,* 1968). Ingestion of these solutions results in rehydration of the patient (Guerrant *et al.,* 2003). Importantly, CRP is a sensor of glucose concentration and H-NS is a sensor of osmotic flux. Hence, it is likely that toxin production can be influenced by sugar and salt concentrations in solutions used to rehydrate patients. This is an area which will require further direct study.

We speculate that cAMP extrusion from epithelial cells mediated by LT is likely to activate CRP, and stimulate ST production. Future experiments might involve exchanging the regulatory regions of the *estA2* gene and *eltAB* operon, and studying an effects on pathogenicity. ETEC strain H10407 is unusual in expressing LT, STp, STh and many of the additional minor virulence factors; including EtpA, CexE, Tia, TibA, LeoA and EatA. The presence of these additional virulence factors is thought to contribute to the unique severity of the strain (Crossman *et al.,* 2010, Porter *et al.,* 2011, Turner *et al.,* 2006). It is interesting to speculate that the differential regulation of the expression of the two *estA* genes might enable ETEC H10407 to produce ST irregardless of intracellular cAMP levels. Our study highlights the fact that TFs that apparently bind at activatory positions at promoters (for example, P*estA1*) may not behave in the expected way, and are sensitive to subtle changes in promoter

architecture. Hence, TF binding sites identified in genome-wide screens must be directly studied in detail, before inferences can be drawn about their possible function.

Finally, a central paradigm of gene regulation is that TF binding sites, and promoters, are located at the 5' end of genes. The use of genome-wide approaches often leads to identification of anomalies that do not fit these expectations. It is often unclear what function these anomalies might have. For example, our bioinformatic screen identified a high-affinity CRP-binding site, located at the 3' end of the *aatPABC* operon, which controlled the activity of a nearby promoter. However, the CRP site, promoter, and downstream gene were only partially conserved in other ETEC genomes. Given that any genome can only be considered as being at a "mid-point" in evolution it is perhaps not surprising that such anomalies are common place. ETEC H10407 is the only strain in which we were able to identify both an *aatS* open reading frame, in combination with a CRP binding site which resembled the consensus. It is possible that the presence of these elements may contribute to strain H10407 disease severity in some way.

Future work in this area will focus on the interplay between CRP and H-NS in the regulation of ETEC toxin genes, and the response of the promoter regions controlling them to different sugar and salt concentrations. Since CRP represses *eltAB* expression indirectly, further studies should be directed at finding the mechanism of regulation. For example, in *Vibrio cholerae*, CRP negatively regulates the *ctxAB* operon, which encodes Cholera toxin. This regulation occurs through three sets of regulators; TcpP and TcpH (membrane-bound), ToxR and ToxS (membrane-bound), and ToxT, which directly activates P*ctxAB* (Kovacikova and Skorupski, 2001). Beyond these experiments, changes in the ETEC transcriptome will be need to be measured *in vivo* using a suitable model organism.

Finally, the environmental conditions found in the mucus lining of the intestine are not well studied. Hence, it is difficult to predict the transcriptional responses of bacteria to this

environment. Determining these conditions will be important for the study of virulence factor expression.

References

Aiba, H. (1985). Transcription of the *Escherichia coli* adenylate cyclase gene is negatively regulated by cAMP-cAMP receptor protein. *Journal of Biological Chemistry* **260**: 3063-3070.

Alba, B.M., and Gross, C.A. (2004). Regulation of the *Escherichia coli* σ^E-dependent envelope stress response. *Molecular Microbiology* **52**: 613-619.

Alderete, J.R. and Robertson, D.C. (1977). Repression of heat-stable enterotoxin synthesis in enterotoxigenic *Escherichia coli*. *Infection and Immunity* **17**: 629-133.

Ali Azam, T., Iwata, A., Nishimura, A., Ueda, S. and Ishihama, A. (1999). Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *Journal of Bacteriology* **181:** 6361-6370.

Anantha, R.P., McVeigh, A.L, Lee, L.H., Agnew, M.K., Cassels, F.J., Scott, D.A., Whittam, T. and Savarino, S.J. (2004). Evolutionary and functional relationships of colonisation factor antigen I and other class 5 adhesive fimbriae of enterotoxigenic *Escherichia coli*. *Infection and Immunity* **72**: 7190-7201.

Ansari, A.Z., Chael, M.L. and O'Halloran, T.V (1992). Allosteric underwinding of DNA is a critical step in positive control of transcription by Hg-MerR. *Nature* **355**: 87-89.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W. and Noble, W.S. (2009). MEME Suite: tools for motif discovery and searching. *Nucleic Acids Research* 1-7.

Bar-Nahum, G., Epshtein, V., Ruckenstein, A.E., Rafikov, R., Mustaev, A. and Nudler, E. (2005). A ratchet mechanism of transcription elongation and its control. *Cell* **120**: 183-193.

Barker, M.M., Gaal, T., Josaitis, C.A., Gourse, R.L. (2001). Mechanism of regulation of transcription initiation by ppGpp. I. Effects of ppGpp of transcription initiation *in vivo* and *in vitro*. *Journal of Molecular Biology* **305**: 673-688.

Barne, K.A., Bown, J.A., Busby, S.J.W, and Minchin, S.D. (1997). Region 2.5 of the *Escherichia coli* RNA polymerase σ^{70} subunit is responsible for the recognition of the'extended –10' motif at promoters. *The EMBO Journal* **16**: 4034-4040.

Beatty, C.M., Browning, D.F., Busby, S.J.W., and Wolfe, A.J. (2003). Cyclic AMP receptor protein-dependent activation of the *Escherichia coli acs*P2 promoter by a synergistic class III mechanism. *Journal of Bacteriology* **185**: 5148-5157.

Bell, A., Gaston, K., Williams, R., Chapman, K., Kolb, A., Buc, H., Minchin, S., Williams, J. and Busby, S. (1990). Mutations that alter the ability of the *Escherichia coli* cyclic AMP receptor protein to activate transcription. *Nucleic Acids Research* **18**: 7243-7250.

Bell, C.E. and Lewis, M. (2001). The Lac repressor: a second generation of structural and functional studies. *Current Opinion in Structural Biology* **11**: 19-25.

Benoff, B., Yang, H., Lawson, C.L., Parkinson, G., Liu, J., Blatter, E., Ebright, Y.W., Berman, H.M. and Berman, R.H. (2002). Structural basis of transcription activation: the CAP-alpha CTD-DNA complex. *Science* **297**: 1562-1566.

Bertin, P., Benhabiles, N., Krin, E., Laurent-Winter, C., Tendeng, C., Turlin, E., Thomas, A., Danchin, A. and Brasseur, R. (1999). The structural and functional organization of H-NS-like proteins is evolutionarily conserved in Gram-negative bacteria. *Molecular Microbiology* **31**: 319-329.

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

Bodero, M. and Munson, G.P. (2009). Cyclic AMP Receptor Protein-Dependent Repression of Heat-labile Enterotoxin. *Infection and Immunity* **77**: 791-798.

Borukhov, S. and Nudler, E. (2003). RNA polymerase holoenzyme: structure, function and biological implications. *Current Opinion in Microbiology* **6**: 93-100.

Borukhov, S. and Nudler, E. (2007). RNA polymerase: the vehicle of transcription. *Trends in Microbiology* **16**: 126-134.

Botsford, J.L. and Harman, J.G. (1992). Cyclic AMP in prokaryotes. *Microbiological Reviews* **56**: 100-122.

Braun, V., Mahren, S. and Ogierman, M. (2003). Regulation of the FecI-type ECF sigma factor by transmembrane signalling. *Current Opinion in Microbiology* **6**: 173-180.

Brown, E.A. and Hardwidge, P.R. (2007). Biochemical characterization of the enterotoxigenic *Escherichia coli* LeoA protein. *Microbiology* **153**: 3776-3784.

Browning, D.F. and Busby, S.J.W. (2004). The Regulation of Bacterial Transcription Initiation. *Nature Reviews Microbiology* **2**: 57-64**.**

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

Buck, M.J., Lieb, J.D. (2004). ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipiation experiments. *Genomics* **83**: 349-360.

Burr, T., Mitchell, J., Kolb, A., Minchin, S. and Busby, S. (2000). DNA sequence elements located immediately upstream of the -10 hexamer in *Escherichia coli* promoters: a systemic study. *Nucleic Acids Research* **28**: 1864-1870.

Busby, S.J.W. and Ebright, R.H. (1994). Promoter Structure, Promoter Recognition, and Transcription Activation in Prokaryotes. *Cell* **79**: 743-746.

Busby, S.J.W. and Ebright, R.H. (1997). Transcription activation at Class II CAP-dependent promoters. *Molecular Microbiology* **23**: 853-859.

Busby, S.J.W. and Ebright, R.H. (1999). Transcription Activation by Catabolite Activator Protein (CAP). *Journal of Molecular Biology* **293**: 199-213.

Busby, S., Kotlarz, D. and Buc, H. (1983). Deletion mutagenesis of the *Escherichia coli* galactose operon promoter region. *Journal of Molecular Biology* **167**: 259-274.

Busque, P., Letellier, A., Harel, J. and Dubreuil, J. (1995). Production of *Escherichia coli* STb enterotoxin is subject to catabolite repression. *Microbiology* **141**: 1621-1627.

Campbell, E.A., Muzzin, O., Chlenov, M., Sun, J.L, Olson, C.A., Weinman, O., Trester-Zedlitz, M.L. and Darst, S.A. (2002). Structure of the bacterial RNA polymerase promoter specificity σ subunit. *Molecular Cell* **9**: 527-539.

Campbell, E.A., Westblade, L.F., and Darst, S.A. (2008). Regulation of bacterial RNA polymerase σ factor activity: a structural perspective. *Current Opinion in Microbiology* **11**: 121-127.

Caron, J., Coffield, L.M. and Scott, J.R. (1989). A plasmid-encoded regulatory gene, *rns*, required for expression of the CS1 and CS2 adhesins of enterotoxigenic *Escherichia coli*. *Proceedings of the National Academy of Sciences* **86**: 963-967.

Caron, J., Maneval, D.R., Kaper, J.B. and Scott, J.R. (1990). Association of *rns* homologs with colonization factor antigens in clinical *Escherichia coli* isolates. *Infection and Immunity* **58**: 3442-3444.

Carver, T., Thomson, N., Bleasby, A., Berriman, M. and Parkhill, J. (2009). DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* **25**: 119-120.

Casadaban, M.J. and Cohen, S.N. (1980). Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. *Journal of Molecular Biology* **138**: 179-207.

Chang, D., Smalley, D.J., Tucker, D.L., Leatham, M.P., Norris, W.E., Stevenson, S.J., Anderson, A.B., Grissom, J.E., Laux, D.C., Cohen, P.S. and Conway, T. (2004). Carbon nutrition of *Escherichia coli* in the mouse intestine. *Proceedings of the National Academy of Sciences* **101**: 7427-7432.

Chao, A.C., de Sauvage, F.J., Dong, Y.J., Wagner, J.A., Goeddel, D.V. and Gardner, P. (1994). Activation of intestinal CFTR Cl- channel by heat-stable enterotoxin and guanylin via cAMP-dependent protein kinase. *The EMBO Journal* **13**: 1065–1072.

Chaudhuri, R.R. and Henderson, I.R. (2012). The evolution of the *Escherichia coli* phylogeny. *Infection, Genetics and Evolution* **12**: 214-226.

Chen, H., Bjerknes, M., Kumar, R. and Jay, E. (1994). Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of *Escherichia coli* mRNAs. *Nucleic Acids Research* **22**: 4953-4957.

Chen, H., Tang, H. and Ebright, R.H. (2003). Function interaction between RNA polymerase α subunit C-terminal domain and σ^{70} in UP-element- and activator-dependent transcription. *Molecular Cell* **11**: 1621-1633.

Cheng, S.H., Rich, D.P., Marshall, J., Gregory, R.J., Welsh, M.J. and Smith, A.E. (1991). Phosphorylation of the R domain by cAMP-dependent protein kinase regulates the CFTR chloride channel. *Cell* **66**: 1027-1036.

Chilcott, G.S., and Hughes, K.T. (2000). Coupling of flagellar gene expression to flagellar assembly in *Salmonella enterica* serovar Typhimurium and *Escherichia coli*. *Microbiology and Molecular Biology Reviews* **64**: 694-708.

Cho, B.K., Kim, D., Knight, E.M., Zenger, K. and Palsson, B.Ø. (2014). Genome-Scale reconstruction of the sigma factor network in *Escherichia coli*: topology and functional states. *BMC Biology* **12**: 4.

Choy, H.E. and Adhya, S. (1992). Control of *gal* transcription through DNA looping: Inhibition of the initial transcribing complex. *Proceedings of the National Academy of Sciences* **89**: 11264-11268.

Crossman, L.L., Chaudhuri, R.R., Beatson, S.A., Wells, T.J., Desvaux, M., Cunningham, A.F., Petty, N.K., Mahon, V., Brinkley, C., Hobman, J.L., Savarino, S.J., Turner, S.M., Pallen, M.J., Penn, C.W., Parkhill, J., Turner, A.K., Johnson, T.J., Thomson, N.R., Smith, S.G.J. and Henderson, I.R. (2010). A Commensal Gone Bad: Complete Genome Sequence of the Prototypical Enterotoxigenic *Escherichia coli* Strain H10407. *Journal of Bacteriology* **192**: 5822-5831.

Croxen, M.A. and Finlay, B.B. (2010). Molecular mechanisms of *Escherichia coli* pathogenicity. *Nature Reviews Microbiology* **8**: 26-38.

Dame, R.T., Wyman, C., Wurm, R., Wagner, R. and Goosen, N. (2002). Structural basis for H-NS-mediated trapping of RNA polymerase in the open initiation complex at the *rrnB* P1. *The Journal of Biological Chemistry* **277**: 2146-2150.

Darst, S.A. (2001). Bacterial RNA polymerase. *Current Opinion in Structural Biology* **11**: 155-162.

Datsenko, K.A. and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences* **97**: 6640-6645.

Dieci, G., Hermann-Le Denmat, S., Lukhtanov, E., Thuriaux, P., Werner, M. and Sentenac, A. (1995). A universally conserved region of the largest subunit participates in the active site of RNA polymerase III. *The EMBO Journal* **14**: 3766-3776.

Dillon, S.C. and Dorman, C.J. (2010). Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nature Review Microbiology* **8**: 185-195.

Dole, S., Nagarajavel, V. and Schnetz, K. (2004). The histone-like nucleoid structuring protein H-NS represses the *Escherichia coli bgl* operon downstream of the promoter. *Molecular Microbiology* **52**: 589-600.

Dorman, C.J. (2004). H-NS: a universal regulator for a dynamic genome. *Nature Reviews Microbiology* **2**: 391-400.

Ebright, R.H. (1993). Transcriptional activation at Class I CAP-dependent promoters. *Molecular Microbiology* **8**: 797-802.

Ebright, R.H. and Busby, S.J.W. (1995). The *Escherichia coli* RNA polymerase α subunit: structure and function. *Current Opinion in Genetics and Development* **5**: 197-203.

Ebright, R.H., Ebright, Y.W. and Gunasekera, A. (1989). Consensus DNA site for the *Escherichia coli* catabolite gene activator protein (CAP): CAP exhibits a 450-fold higher affinity for the consensus DNA site than for the *E.coli lac* DNA site. *Nucleic Acids Research* **17**: 10295-10305.

Einhauer, A. and Jungbauer, A. (2001). The FLAG[TM] peptide, a versatile fusion tag for the purification of recombinant proteins. *Journal of Biochemical and Biophysics Methods* **49**: 455-465.

Elsinghorst, E.A. and Weitz, J.A. (1994). Epithelial cell invasion and adherence directed by the enterotoxigenic *Escherichia coli tib* locus is associated with a 104-kilodalton outer membrane protein. *Infection and Immunity* **62**: 3463-3471.

Estrem, S.T., Gaal, T., Ross, W., and Gourse, R.L. (1998). Identification of an UP element consensus sequence for bacterial promoters. *Proceedings of the National Academy of Sciences* **95**: 9761-9766.

Evans, D.G., Evans, D.J., Clegg, S. and Pauley, J.A. (1979). Purification and characterization of the CFA/I antigen of enterotoxigenic *Escherichia coli*. *Immunity and Infection* **25**: 738.

Evans, D.G., Evans, D.J. and Pierce, N.F. (1973). Differences in the response of rabbit small intestine to heat-labile and heat-stable enterotoxins of *Escherichia coli*. *Infection and Immunity* **7**: 873-890.

Evans, D.G., Silver, R.P., Evans, D.J., Chase, D.G. and Gorbach, S.L. (1975). Plasmid-controlled colonization factor associated with virulence in *Escherichia coli* enterotoxigenic for humans. *Infection and Immunity* **12**: 656-667.

Fabich, A.J., Jones, S.A., Chowdhury, F.Z., Cernosek, A., Anderson, A., Smalley, D., McHargue, J.W., Hightower, G.A., Smith, J.T., Autieri, S.M., Leatham, M.P., Lins, J.J., Allen, R.L., Laux, D.C., Cohen, P.S. and Conway, T. (2008). Comparison of carbon nutrition for pathogenic and commensal *Escherichia coli* strains in the mouse intestine. *Infection and Immunity* **76**: 1143-1152.

Feklistov, A. and Darst, S.A. (2011). Structural basis for promoter -10 element recognition by the bacterial RNA polymerase σ subunit. *Cell* **147**: 1257-1269.

Feklistov, A., Sharon, B.D., Darst, S.A. and Gross, C.A. (2014). Bacterial sigma factors: a historical, structural, and genomic perspective. *Annual Reviews in Microbiology* **68**: 357-376.

Fenton, M.S. and Gralla, J.D. (2001). Function of the bacterial TATAAT -10 element as a single-stranded DNA during RNA polymerase isomerisation. *Proceedings of the National Academy of Sciences* **98**: 9020-9025.

Fic, E., Bonarek, P., Gorecki, A., Kedracka-Krok, S., Mikolajczak, J., Polit, A., Tworzydlo, M., Dziedzicka-Wasylewska, M. and Wasylewski, Z. (2009). cAMP receptor protein from *Escherichia coli* as a model of signal transduction in proteins- a review. *Journal of Molecular Microbiology and Biotechnology* **17**: 1-11.

Figueroa-Bossi, N., Guérin, M., Rahmouni, R., Leng, M. and Bossi, L. (1998). The supercoiling sensitivity of a bacterial tRNA promoter parrelels its responsiveness to stringent control. *The EMBO Journal* **17**: 2359-2367.

Fitzgerald, D.M., Bonocora, R.P. and Wade, J.T. (2014). Comprehensive mapping of the *Escherichia coli* flagellar regulatory network. *PLoS Genetics* **10**: 10.

Fleckenstein, J.M., Hardwidge, P.R., Munson, G.P., Rasko, D.A., Sommerfelt, H. and Steinsland, H. (2010). Molecular mechanisms of enterotoxigenic *Escherichia coli* infection. *Microbes and Infection* **12**: 89-98.

Fleckenstein, J.M., Kopecko, D.J., Warren, R.L. and Elsinghorst, E.A. (1996). Molecular characterization of the *tia* invasion locus from enterotoxigenic *Escherichia coli*. *Infection and Immunity* **64**: 2256-2265.

Fleckenstein, J.M., Roy, K., Fischer, J.F. and Burkitt, M. (2006). Identification of a two-partner secretion locus of enterotoxigenic *Escherichia coli*. *Infection and Immunity* **74**: 2245-2258.

Gaastra, W. and Svennerholm, A.M. (1996). Colonization factors of human enterotoxigenic *Escherichia coli* (ETEC). *Trends in Microbiology* **4**: 444-452.

Gaal, T., Ross, W., Blatter, E.E., Tang, H., Jia, X., Krishnan, V.V., Assa-Munt, N., Ebright, R.H. and Gourse, R.L. (1996). DNA binding determinants of the alpha subunit of RNA polymerase: novel DNA-binding domain architecture. *Genes and Development* **10**: 16-26.

Gaston, K., Bell, A., Kolb, A., Buc, H. and Busby, S. (1990). Stringent spacing requirements for transcription activation by CRP. *Cell* **62**: 733-743.

Gentry, D.R. and Burgess, R.R. (1989). *rpoZ*, encoding the omega subunit of *Escherichia coli* RNA polymerase, is in the same operon as *spot*. *Journal of Bacteriology* **171**: 1271-1277.

Ghosaini, L.R., Brown, A.M. and Sturtevant, J.M. (1988). Scanning calorimetric study of the thermal unfolding of catabolite activator protein from *Escherichia coli* in the absence and presence of cyclic mononucleotides. *Biochemistry* **27**: 5257-5261.

Gibert, I., Villegas, V. and Barbé, J. (1990). Expression of heat-labile enterotoxin genes is under cyclic AMP control in *Escherichia coli*. *Current Microbiology* **20**: 83-90.

Goldman, S.R., Ebright, R.H. and Nickels, B.E. (2009). Direct detection of abortive RNA transcripts *in vivo*. *Science* **324**: 927-928.

Gordon, B.R.G., Imperial, R., Wang, L., Navarre, W.W. and Liu, J. (2008). Lsr2 of *Mycobacterium* represents a novel class of H-NS-like proteins. *Journal of Bacteriology* **190**: 7052-7059.

Gordon, B.R.G., Li, Y., Wang, L., Sintsova, A., van Bakel, H., Tian, S., Navarre, W.W., Xia, B. and Liu, J. (2010). Lsr2 is a nucleoid-associated protein that targets AT-rich sequences and virulence genes in *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences* **107**: 5154-5159.

Görke, B., Stűlke, J. (2008) Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nature Reviews Microbiology* **6**: 613-624.

Gosset, G., Zhang, Z., Nayyar, S., Cuevas, W.A. and Saier, M.H. (2004). Transcriptome analysis of Crp-dependent catabolite control of gene expression in *Escherichia coli*. *Journal of Bacteriology* **186**: 3516-3524.

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

Grainger, D.C., Hurd, D., Harrison, M., Holdstock, J. and Busby, S.J.W. (2005). Studies of the distribution of *Escherichia coli* cAMP-receptor protein and RNA polymerase along the *E. coli* chromosome. *Proceedings of the National Academy of Sciences* **102**: 17693-17698.

Grainger, D.C., Webster, C.L., Belyaeva, T.A., Hyde, E.I., and Busby, S.J.W. (2004). Transcription activation at the *Escherichia coli melAB* promoter: interactions of MelR with its DNA target site and with domain 4 of RNA polymerase σ subunit. *Molecular Microbiology* **51**: 1297-1309.

Green, J., Stapleton, M.R., Smith, L.J., Artymiuk, P.J., Kahramanoglou, C., Hunt, D.M., Buxton, R.S. (2014). Cyclic-AMP and bacterial cyclic-AMP receptor proteins revisited: adaption for different ecological niches. *Current Opinion in Microbiology* **18**: 1-7.

Gruber, T.M. and Gross, C.A. (2003). Multiple sigma subunits and the partitioning of bacterial transcriptional space. *Annual Reviews in Microbiology* **57**: 441-66.

Guerrant, R.L., Carneiro-Filho, B.A. and Dillingham, R.A. (2003). Cholera, diarrhea, and oral rehydration therapy: triumph and indictment. *Clinical Infectious Diseases* **37**: 398-405.

Hamilton, D., Johnson, M., Forsyth, G., Roe, W. and Nielsen, N. (1978). The effect of cholera toxin and heat labile and heat stable *Escherichia coli* enterotoxin on cyclic AMP concentrations in small intestinal mucosa of pig and rabbit. *Canadian Journal of Comparative Medicine* **42**: 327-331.

Harman, J.G. (2001). Allosteric regulation of the cAMP receptor protein. *Biochimica et Biophysica Acta* **1547**: 1-17.

Helmann, J.D. and Chamberlin, M.J. (1988). Structure and function of bacterial sigma factors. *Annual Reviews in Biochemistry* **57**: 839-872.

Hengge-Aronis, R. (2002). Signal transduction and regulatory mechanisms involved in control of the of σ^S (RpoS) subunit of RNA polymerase. *Microbiology and Molecular Biology Reviews* **66**: 373-395.

Hiard, S., Marée, R., Colson, S., Hoskisson, P.A., Titgemeyer, F., van Wezel, G.P., Joris, B., Wehenkel, L. and Rigali, S. (2007). PREDetector: a new tool to identify regulatory elements in bacterial genomes. *Biochemical and Biophysical Research Communications* **357**: 861-864.

Hobman, J.L. (2007). MerR family transcription activators: similar designs, different specificities. *Molecular Microbiology* **63**:1275-1278.

Hogema, B.M., Arents, J.C., Bader, R., Eikkemans, K., Yoshida, H., Takahashi, H., Aiba, H. and Postma, P.W. (1998). Inducer exclusion in *Escherichia coli* by non-PTS substrates: the role of the PEP to pyruvate ratio in determining the phosphorylation state of enzyme IIA^{Glc}. *Molecular Microbiology* **30**: 487-498.

Hogema, B.M., Arents, J.C., Bader, R. and Postma, P.W. (1999). Autoregulation of lactose uptake through the LacY permease by enzyme IIA^{Glc} of the PTS in *Escherichia coli* K-12. *Molecular Microbiology* **31**: 1825-1833.

Hollands, K., Busby, S.J.W. and Lloyd, G.S. (2007). New targets for the cyclic AMP receptor protein in the *Escherichia coli* K-12 genome. *FEMS Microbiology Letters* **274**: 89-94.

Hook-Barnard, I.G. and Hinton, D.M. (2009). The promoter spacer influences transcription initiation via σ^{70} region 1.1 of *Escherichia coli*. *Proceedings of the National Academy of Sciences* **106**: 737-742.

Horstman, A.L. and Kuehn, M.J. (2002). Bacterial surface association of heat-labile enterotoxin through lipopolysaccharide after secretion via the general secretion pathway. *The Journal of Biological Chemistry* **277**: 32538-32545.

Huerta, A.M., Salgado, H., Thieffry, D. and Collado-Vides J. (1998). RegulonDB: a database on transcriptional regulation in *Escherchia coli*. *Nucleic Acids Research* **26**: 55-59.

Igarashi, K., Fujita, N. and Ishihama, A. (1989). Promoter selectivity of *Escherichia coli* RNA polymerase: omega factor is responsible for ppGpp sensitivity. *Nucleic Acids Research* **17**: 8755-8765.

Inada, T., Kimata, K. and Aiba, H. (1996). Mechanism responsible for glucose-lactose diauxie in *Escherichia coli*: challenge to the cAMP model. *Genes to Cells* **1**: 293-301.

Ishihama, A. (1990). Molecular assembly and functional modulation of *Escherichia coli* RNA polymerase. *Advances in Biophysics* **26**: 19-31.

Ishihama, A. (2010). Prokaryotic genome regulation: multifactor promoters, multitarget regulator and hierarchic networks. *FEMS Microbiology Reviews* **34**: 628-645.

Ishihama, A. (1993). Protein-protein communication within the transcription apparatus. *Journal of Bacteriology* **175**: 2483-2489.

Isidean, S.D., Riddle, M.S., Savarino, S.J. and Porter, C.K. (2011). A systematic review of ETEC epidemiology focusing on colonisation factor and toxin expression. *Vaccine* **29**: 6167-6178.

Islam, M.D., Shaw, R.K., Frankel, G., Pallen, M.J. and Busby, S.J.W. (2012). Translation of a minigene in the 5' leader sequence of the enterohaemorrhagic *Escherichia coli LEE1* transcription unit affects expression of the neighbouring downstream gene. *Biochemical Journal* **441**: 247-253.

Jishage, M., Gasgupta, D. and Ishihama, A. (2001). Mapping of the Rsd contact site on the sigma 70 subunit of *Escherichia coli* RNA polymerase. *Journal of Bacteriology* **183**: 2952-56.

Johnson, A.M., Kaushik, R.S., Francis, D.H., Fleckenstein, J.M. and Hardwidge, P.R. (2009). Heat-Labile Enterotoxin Promotes *Escherichia coli* Adherence to Intestinal Epithelial Cells. *Journal of Bacteriology* **191**: 178-186.

Jordi, B.J.A.M, van der Zeijst, B.A.M. and Gaastra, W. (1994). Regions of the CFA/I promoter involved in the activation by the transcriptional activator CfaD and repression by the histone-like protein H-NS. *Biochimie* **76**: 1052-1054.

Jordi, B.J., Willshaw, G.A., van der Zeijst, B.A. and Gaastra, W. (1992). The complete nucleotide sequence of region 1 of the CFA/I fimbrial operon of human enterotoxigenic *Escherichia coli*. *DNA sequence* **2**: 257-263.

Jordi, B.J.A.M, Dagberg, B., de Haan, L.A.M., Hamers, A.M., van der Zeijst, B.A.M., Gaastra, W. and Erin Uhlin, B. (1992). The positive regulator CfaD overcomes the repression mediated by histone-like protein H-NS (H1) in the CFA/I fimbrial operon of *Escherichia coli*. *The EMBO Journal* **11**: 2627-2632.

Kahramanoglou, C., Seshasayee, A.S.N., Prieto, A.I., Ibberson, D., Sachmidt, S., Zimmermann, J., Benes, V., Fraser, G.M., and Luscombe, N.M. (2011). Direct and indirect effects of H-NS and Fis on global gene expression control in *Escherichia coli*. *Nucleic Acids Research* **39**: 2073-2091.

Kanack, K.J., Runyen-Janecky, L.J., Ferrell, E.P., Suhll, S. and West, S.E.H. (2006). Characterization of DNA-binding specificity and analysis of binding sites of the *Pseudomonas aeruginosa* global regulator, Vfr, a homologues of the *Escherichia coli* cAMP receptor protein. *Microbiology* **152**: 3485-3496.

Kapanidis, A.N., Margeat, E., On Ho, S., Kortkhonjia, E., Weiss, S. and Ebright, E.H. (2006). Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* **314**: 1144-1147.

Karimova, G., Pidoux, J., Ullmann, A. and Ladant, D. (1998). A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proceedings of the National Academy of Sciences* **95**: 5752-5756.

Karimova, G., Ullmann, A. and Ladant, D. (2001). Protein-protein interaction between *Bacillus stearothermophilus* tyrosyl-tRNA synthetase subdomains revealed by a bacterial two-hybrid system. *Journal of Molecular Microbiology and Biotechnology* **3**: 73-82.

Karjalainen, R.K., Evans, D.G., Evans D.J., Graham, D.Y. and Lee, C.H. (1991). Catabolite repression of the colonization factor antigen I (CFA/I) operon of *Escherichia coli*. *Current Microbiology* **23**: 307-313.

Keseler, I.M., Collado-Vides, J., Gama-Castro, S., Ingraham, J., Paley, S., Paulsen, IT., Peralta-Gil, M. and Karp, P.D. (2005). EcoCyc: a comprehensive databse resource for *Escherchia coli*. *Nucleic Acids Research* **33**: 334-337.

Kim, D., Hong, J.S.J., Qiu, Y., Nagarajan, H., Seo, J.H., Cho, B.K., Tsai, S.F. and Palsson, B.Ø. (2012). Comparative analysis of regulatory elements between *Escherichia coli* and *Klebsiella pneumonia* by genome-wide transcription start site profiling. *PLOS Genetics* **8:**(8).

Kimata, K., Takahashi, H., Inada, T., Postma, P. and Aiba, H. (1997). cAMP receptor protein-cAMP plays a crucial role in glucose-lactose diauxie by activating the major glucose tranporter gene in *Escherichia coli*. *Proceedings of the National Academy of Sciences* **94**: 12914-12919.

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

Kovacikova, G. and Skorupski, K. (2001). Overlapping binding sites for the virulence gene regulators AphA, AphB, and cAMP-CRP at at the *Vibrio cholerae tcpPH* promoter. *Molecular Microbiology* **41**: 393-407.

Kustu, S., Santero, E., Keener, J., Popham, D. and Weiss, D. (1989). Expression of σ^{54} (*ntrA*)-dependent genes is probably united by a common mechanism. *Microbiological Reviews* **53**: 367-376.

Lang, B., Blot, N., Bouffartigues, E., Buckle, M., Geertz, M., Gualerzi, C.O., Mavathur, R., Muskhelishvilli, G., Pon, C.L., Rimsky, S., Stella, S., Babu, M.M. and Travers, A. (2007). High-affinity DNA binding sites for H-NS provide a molecular bases for selective silencing within proteobacterial genomes. *Nucleic Acids Research* **35**: 6330-6337.

Lawson, C.L., Swigon, D., Murakami, K.S., Darst, S.A., Berman, H.M. and Ebright, R.H. (2004). Catabolite activator protein: DNA binding and transcription activation. *Current Opinion in Structural Biology* **14**: 10-14.

Le Bouguénec, C. and Schouler, C. (2011). Sugar metabolism, an additional virulence factor in enterobacteria. *International Journal of Medical Microbiology* **301**:1-6.

Lee, D.J., Busby, S.J.W. and Lloyd, G.S. (2003). Exploitation of a chemical nuclease to investigate the location and orientation of the *Escherichia coli* RNA polymerase α subunit C-terminal domains at simple promoters that are activated by cyclic AMP receptor protein. *The Journal of Biological Chemistry* **278**: 52944-52952.

Lee, D.J., Minchin, S.D. and Busby, S.J.W. (2012). Activating Transcription in Bacteria. *Annual Reviews in Microbiology* **66**: 125-152.

Levine, M.M., Caplan, E.S., Waterman, D., Cash, R.A., Hornick, R.B. and Snyder, M.J. (1977). Diarrhoea caused by *Escherichia coli* that produce only heat-stable enterotoxin. *Infection and Immunity* **17**: 8-82.

Lewis, D.E.A. and Adhya, S. (2002). *In vitro* repression of the *gal* promoters by GalR and HU depends of the proper helical phasing of the two operators. *Journal of Biological Chemistry* **277**: 2498-2504.

Lim, C.J., Lee, S.Y., Kenney, L.J. and Yan, J. (2012). Nucleoprotein filament formation is the structural basis for bacterial protein H-NS gene silencing. *Scientific Reports* **2**: (509).

Lindenthal, C. and Elsinghorst, E.A. (1999). Identification of a glycoprotein produced by enterotoxigenic *Escherichia coli*. *Infection and Immunity* **67**: 4084-4091.

Lindenthal, C. and Elsinghorst, E.A. (2001). Enterotoxigenic *Escherichia coli* TibA glycoprotein adheres to human intestinal epithelial cells. *Infection and Immunity* **69**: 5257.

Liu, Y., Chen, H., Kenny, L.J. and Yan, J. (2010). A divalent switch drives H-NS/DNA-binding conformations between stiffening and bridging modes. *Genes and Development* **24**: 339-344.

Lloyd, G.S., Niu, W., Tebbutt, J., Ebright, R.H. and Busby, S.J.W. (2002). Requirement for two copies of RNA polymerase α subunit C-terminal domain for synergistic transcription activation at complex bacterial promoters. *Genes and Development* **16**: 2557-2565.

Lodge, J., Fear, J., Busby, S., Gunasekaran, P. and Kamini, N.R. (1992). Broad host range plasmids carrying the *Escherichia coli* lactose and galactose operons. *FEMS Microbiology Letters* **95**: 271-276.

Lucchini, S., Rowley, G., Goldberg, M.D., Hurd, D., Harrison, M. and Hinton, J.C.D. (2006). H-NS mediates the silencing of laterally acquired genes in bacteria. *PLoS Pathogens* **2**:(8).

Lucht, J.M., Dersch, P., Kempf, B. and Bremer, E. (1994). Interactions of the nucleoid-associated DNA-binding protein H-NS with the regulatory region of the osmotically controlled *proU* operon of *Escherichia coli*. *The Journal of Biological Chemistry* **269**: 6578-6596.

MacLellan, S.R., Eiamphungporn, W. and Helmann, J.D. (2009). ROMA: an *in vitro* approach to defining target genes for transcription regulators. *Methods* **47**: 73-77.

Maeda, H., Fujita, N. and Ishihama, A. (2000). Competition among seven *Escherichia coli* σ subunits: relative binding affinities to the core RNA polymerase. *Nucleic Acids Research* **28**: 3497-3503.

Malan, T.P., Kolb, A., Buc, H. and McClure, W.R. (1984). Mechanism of CRP-cAMP activation of *lac* operon transcription initiation activation of the P1 promoter. *Journal of Molecular Biology* **180**: 881-909.

Mammarappallil, J.G. and Elsinghorst, E.A. (2000). Epithelial cell adherence mediated by the enterotoxigenic *Escherichia coli* Tia protein. *Infection and Immunity* **68**: 6595-6601.

Mardis, E.R. (2007). The impact of next-generation sequencing technology on genetics. *Trends in Genetics* **24**: 133-141.

Mathew, R. and Chatterji, D. (2006). Evolving story of the RNAP ω subunit. *TRENDS in Microbiology*, **14**: 450-455.

McKay, D.B., Weber, I.T. and Steitz, T.A. (1982). Structure of catabolite gene activator protein at 2.9Å resolution. *The Journal of Biological Chemistry* **257**: 9518-9524.

Mekler, V., Kortkhonjia, E., Mukhopadhyay, J., Knight, J., Revyakin, A., Kapanidis, A.N., Niu, W., Ebright, Y.W., Levy, R. and Ebright, R.H. (2002). Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell* **108**: 599-614.

Miller, J.H. (1972). Experiements in molecular genetics. *Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press*.

Minakhin, L., Bhagat, S., Brunning, A., Campbell, E.A., Darst, S.A., Ebright, R.H. and Severinov, K. (2001). Bacterial RNA polymerase subunit ω and eukaryotic RNA polymerase subunit RPB6 are sequence, structural and functional homologs and promoter RNA polymerase assembly. *Proceedings of the National Academy of Sciences* **98**: 892-897.

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

Mockler, T.C. and Ecker, J.R. (2005). Applications of DNA tiling arrays for whole-genome analysis. *Genomics* **85**: 1-15.

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

Monsalve, M., Calles, B., Mencía, M., Rojo, F. and Salas, M. (1998). Binding of phage φ29 protein p4 to the early A2c promoter: recruitment of a repressor by the RNA polymerase. *The Journal of Molecular Biology* **283**: 559-569.

Morita, M.T., Tanaka, Y., Kodama, T.S., Kyogoku, Y., Yanagi, H., Yura, T. (1999). Translational induction for heat shock transcription factor σ^{32}: evidence for a built-in RNA thermosensor. *Genes and Development* **13**: 655-665.

Moseley, S.L., Samadpour-Matolebi, M. and Falkow, S. (1983). Plasmid associated and nucleotide sequence relationships of two genes encoding heat-stable enterotoxin production in *Escherichia coli* H10407. *Journal of Bacteriology* **156**: 441-443.

Moss, J. and Richardson, S.H. (1978). Activation of adenylate cyclase by heat-labile toxin. Evidence for ADP-ribosyltransferase activity similar to that of choleragen. *Journal of Clinical Investigation* **62**: 281-285.

Munson, G.P. and Scott, J.R. (1999). Binding site recognition by Rns, a virulence regulator in the AraC family. *Journal of Bacteriology* **181**: 2110-2117.

Munson, G.P., Holcomb, L.G., Alexander, H.L. and Scott, J.R. (2002). *In vitro* identification of Rns- regulated genes. *Journal of Bacteriology* **184:** 1196-1199.

Murakami, K.S. (2013). X-ray crystal structure of *Escherichia coli* RNA polymerase σ^{70} holoenzyme. *The Journal of Biological Chemistry* **288**: 9126-9134.

Murakami, K.S. and Darst, S.A. (2003). Bacterial RNA polymerases: the wholo story. *Current Opinion in Structural Biology* **13**: 31-39.

Murakami, K.S., Masuda, S. and Darst, S.A. (2002). Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science* **296**: 1280-1284.

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

Myers, K.S., Yan, H., Ong, I.M., Chung, D., Liang, K., Tran, F., Keleş, S., Landick, R. and Kiley, P.J. (2013). Genome-scale analysis of *Escherichia coli* FNR reveals complex features of transcription factor binding. *PLOS Genetics* **9**: (6).

Nalin, D.R., Cash, R.A., Islam, R., Molla, M. and Phillips, R.A. (1968). Oral maintenance therapy for Cholera in adults. *Lancet* **2**: 370-373.

Narang, A. (2009). Quantitative effect and regulatory function of cyclic adenosine 5'-phosphate in *Escherichia coli*. *Journal of Biosciences* **34**: 445-463.

Nataro, J. and Kaper, J. (1998). Diarrhoeagenic *Escherichia coli*. *Clinical Microbiology Review* **11**: 142-201.

Navarre, W.W., McClelland, M., Libby, S.J. and Fang, F.C. (2007). Silencing of xenogenic DNA by H-NS- facilitation of lateral gene transfer in bacteria by a defense system that recognises foreign DNA. *Genes and Development* **21**: 1456-1471.

Navarre, W.W., Porwollik, S., Wang, Y., McClelland, M., Rosen, H., Libby, S.J. and Fang, F.C. (2006). Selective silencing of foregin DNA with low GC content by the H-NS protein in *Salmonella*. *Science* **313**: 236-238.

Nelson, S.O., Wright, J.K. and Postma, P.W. (1983). The mechanism of inducer exclusion. Direct interaction between purified III^{Glc} of the phosphoenolpyruvate: sugar phophotransferase system and the lactose carrier of *Escherichia coli*. *The EMBO Journal* **2**:715-720.

Newsome, P.M., Burgess, M.N. and Mullan, N.A. (1978). Effect of *Escherichia coli* heat-stable enterotoxin on cyclic GMP levels in mouse intestine. *Infection and Immunity* **22**: 290-291.

Nishi, J., Sheikh, J., Mizuguchi, K., Luisi, B., Burland, V., Boutin, A., Rose, D.J., Blattner, F.R. and Nataro, J.P. (2003). The export of coat protein from enteroaggregative *Escherichia coli* by a specific ATP-binding cassette transporter system. *The Journal of Biological Chemistry* **278**: 45680-45689.

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

Niu, W., Zhou, Y., Dong, Q., Ebright, Y.W. and Ebright, R.H. (1994). Characterisation of the activating region of *Escherichia coli* catabolite gene activator protein (CAP) I. Saturation and Alanine-scanning mutagenesis. *Journal of Molecular Biology* **243**: 595-602.

Notley-McRobb, L., Death, A. and Ferenci, T. (1997). The relationship between external glucose and concentration of cAMP levels in *Escherichia coli*: implications for models of phosphotransferase-mediated regulation of adenylate cyclase. *Microbiology* **143**: 1909-1938.

Oehler, S., Eismann, E.R., Kramer, H. and Müller-Hill, B. (1990). The three operators of the *lac* operon cooperate in repression. *The EMBO Journal* **9**: 973-979.

Okamoto, K., Baba, T., Yamanaka, H., Akashi, N. and Fujii, Y. (1995). Dusulphide bond formation and secretion of *Escherichia coli* heat-stable enterotoxin II. *Journal of Bacteriology* **177**: 4579-4586.

Olsén, A., Arnqvist, A., Hammar, M., Sukupolvi, S. and Normark, S. (1993). The RpoS sigma factor relieves H-NS-mediated transcriptional repression of *csgA*, the subunit gene of fibronectin-binding curli in *Escherichia coli*. *Molecular Microbiology* **7**: 523-536.

Park PJ (2009). ChIP-seq: advantages and challenges of a maturing technology. *Nature Reviews Genetics* **10**: 669-680.

Parkinson, G., Wilson, C., Gunasekera, A., Ebright, Y.W., Ebright, R.E. and Berman, H.M. (1996). Structure of the CAP-DNA complex at 2.5Å resolution: a complete picture of the protein-DNA interface. *The Journal of Molecular Biology* **260**: 395-408.

Passner, J.M., Schultz, S.C. and Steitz, T.A. (2000). Modeling the cAMP-induced allosteric transition using the crystal structure of CAP-cAMP at 2.1Å resolution. *Journal of Molecular Biology* **304**: 847-859.

Passner, J.M. and Steitz, T.A. (1997). The structure of a CAP-DNA complex having two cAMP molecules bound to each monomer. *Proceedings of the National Academy of Sciences* **94**: 2843-2847.

Paul, B.J., Barker, M.M., Ross, W., Schneider, D.A., Webb, C., Foster, J.W. and Gourse, R.L. (2004). DksA: A critical component of the transcription initiation machinery that potentiates the regulation of rRNA promoters by ppGpp and the initiating NTP. *Cell* **118**: 311-322.

Paul, B.J., Berkman, M.B., Gourse, R.L. (2005). DksA potentiates direct activation of amino acid promoters by ppGpp. *Proceedings of the National Academy of Sciences* **102**: 7823-7828.

Peters, J.M., Mooney, R.A., Grass, J.A., Jessen, E.D., Tran, F. and Landick, R. (2012). Rho and NusG supress pervasive antisense transcription in *Escherichia coli*. *Genes and Development* **26**: 2621-2633.

Pilonieta, M.C., Bodero, M.D. and Munson, G.P. (2007). CfaD-Dependent Expression of a Novel Extracytoplasmic Protein from Enterotoxigenic *Escherichia coli*. *Journal of Bacteriology* **189**: 5060-5067.

Popovych, N., Tzeng, S.R., Tonelli, M., Ebright, R.H. and Kalodimos, C.G. (2009). Structural basis for cAMP-mediated allosteric control of the catabolite activator protein. *Proceedings of the National Academy of Sciences* **106**: 6927-6932.

Porter, C.K., Riddle, M.S., Tribble, D.R., Louis Bougeois, A., McKenzie, R., Isidean, S.D., Sebeny, P. and Savarino, S.J. (2011). A systematic review of experimental infections with enterotoxigenic *Escherichia coli* (ETEC). *Vaccine* **29**: 5869-5885.

Potrykus, K. and Cashel, M. (2008). ppGpp: still magical?. *Annual Reviews in Microbiology* **62**: 35-51.

Pul, U., Wurm, R., Arslan, Z., Geiβen, R., Hofmann, N. and Wagner, R. (2010). Identification and characterization of *E. coli* CRISPR-*cas* promoters and their silencing by H-NS. *Molecular Microbiology* **75**: 1495-1512.

Qadri, F., Svennerholm, A.M., Faruque, A.S.G. and Sack, R.B. (2005). Enterotoxigenic *Escherichia coli* in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention. *Clinical Microbiology Reviews* **18:** 465-483.

Ramseier, T.M. and Saier, M.G. (1995). cAMP-CRP receptor protein complex: five binding sites in the control region of the *Escherichia coli* mannitol operon. *Microbiology* **141**: 1901-1907.

Reppas, N.B., Wade, J.T., Church, G.M.. and Stuhl, K. (2006). The transition between transcriptional initiation and elongation in *E. coli* is highly variable and often rate limiting. *Molecular Cell* **24**: 747-757.

Revyakin, A., Liu, C., Ebright, R.H. and Strick, T.R. (2006). Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* **314**: 1139-1143.

Rhodius, V.A. and Busby, S.J.W. (2000). Interactions between activating region 3 of the *Escherichia coli* cyclic AMP receptor protein and region 4 of the RNA polymerase sigma (70) subunit: application of suppression genetics. *Journal of Molecular Biology* **299**: 311-324.

Rhodius, V.A. and Busby, S.J.W. (2000). Transcription activation by the *Escherichia coli* cyclic AMP receptor protein: determinants within activating region 3. *Journal of Molecular Biology* **299**: 295-310.

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

Richet, E., Vidal-Ingigliardi, D. and Raibaud, O. (1991). A new mechanism for coactivation of transcription initiation: repositioning of an activator triggered by the binding of a second activator. *Cell* **66**:1185-1195.

Rimsky, S., Florent, Z., Buckle, M. and Henri, B. (2001). A molecular mechanism for the repression of transcription by the H-NS protein. *Molecular Microbiology* **42**: 1311-1323.

Rivera, F.P., Medina, A.M., Aldasoro, E., Sangil, A., Gascon, J., Ochoa, T.J., Vila, J. and Ruiz, J. (2013). Genotypic characterization of enterotoxigenic *Escherichia coli* strains causing traveller's diarrhoea. *Journal of Clinical Microbiology* **51**: 633-635.

Robison, K., McGuire, A. and Church, G.M. (1998). A comprehensive library of DNA-binding site matrices for 55 proteins applied to the complete *Escherichia coli* K-12 genome. *The Journal of Molecular Biology* **284**: 241-254.

Ross, W., Gosink, K.K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K. and Gourse, R.L. (1993). A Third Recognition Element in Bacterial Promoters: DNA Binding by the a Subunit of RNA Polymerase. *Science* **262**: 1407-1413.

Ross, W. and Gourse, R.L. (2005). Sequence-independent upstream DNA- αCTD interactions strongly stimulate *Escherichia coli* RNA polymerase-*lacUV5* promoter association. *Proceedings of the National Academy of Sciences* **102**: 291-296.

Ross, W., Schneider, D.A., Paul, B.J., Mertens, A. and Gourse, R.L. (2003). An intersubunit contact stimulating transcription initiation by *E.coli* RNA polymerase: interaction of the α C-terminal domain and σ region 4. *Genes and Development* **17**: 1293-1307.

Ross, W., Vrentas, C.E, Sanchez-Vasquez, P., Gaal, T. and Gourse, R.L. (2013). The Magic Spot: A ppGpp binding site on *E.coli* RNA polymerase responsible for regulation of transcription initiation. *Molecular Cell* **50**: 420-429.

Rossiter, A.E., Browning, D.F., Leyton, D.L., Johnson, M.D., Godfrey, R.E., Wardius, C.A., Desvaux, M., Cunningham, A.F., Ruiz-Perez, F., Nataro, J.P., Busby, S.J.W. and Henderson, I.R. (2011). Transcription of the plasmid-encoded toxin gene from enteroaggregative *Escherichia coli* is regulated by a novel co-activation mechanism involving CRP and Fis. *Molecular Microbiology* **81**: 179-191.

Roy, K., Hamilton, D.J. and Fleckenstein, J.M. (2012). Co-operative role of antibodies against heat-labile toxin and the EtpA adhesin in preventing toxin delivery and intestinal colonization by enterotoxigenic *Escherichia coli*. *Clinical and Vaccine Immunology* **19**: 1603-1608.

Roy, K., Hilliard, G.M., Hamilton, D.J., Luo, J., Ostmann, M.M. and Fleckenstein, J.M. (2009). Enterotoxigenic *Escherichia coli* EtpA mediates adhesion between flagella and host cells. *Nature* **457**: 594-598.

Sack, R.B., Gorbach, S.L., Banwell, J.G., Jacobs, B., Chatterjee, B.D. and Mitra, R.C. (1971). Enterotoxigenic *Escherichia coli* Isolated from Patients with Severe Cholera-like Disease. *The Journal of Infectious Diseases* **123**: 378-385.

Saecker, R.M., Record, M.T. and deHaseth, P.L. (2011). Mechanism of bacterial transcription initiation: RNA polymerase-promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *Journal of Molecular Biology* **412**: 754-771.

Saier Jr. MH (1998). Multiple mechanisms controlling carbon metabolism in bacteria. *Biotechnology and Bioengineering* **58**: 170-174.

Sato, T. and Shimonishi, Y. (2003). Structural features of *Escherichia coli* heat-stable enterotoxin that activates membrane-associated guanylyl cyclase. *Journal of Peptide Research* **63**: 200-206.

Savelkoul, P.H.M., Willshaw, G.A., McConnell, M.M., Smith, H.R., Hamers, A.M., van der Zeijst, B.A.M. and Gaastra, W. (1990). Expression of CFA/I fimbriae is positively regulated. *Microbial Pathogenesis* **8**: 91-99.

Savery, N.J., Lloyd, G.S., Busby, S.J.W., Thomas, M.S., Ebright, R.H. and Gourse, R.L. (2002). Determinants of the C-terminal domain of the *Escherichia coli* RNA polymerase α subunit important for transcription at Class I cyclic AMP receptor protein-dependent interactions. *Journal of Bacteriology* **184**: 2273-2280.

Savery, N.J., Lloyd, G.S., Kainz, M., Gaal, T., Ross, W., Ebright, R.H., Gourse, R.L. and Busby, S.J.W. (1998). Transcription activation at class II CRP-dependent promoters: identification of determinants in the C-terminal domain of the RNA polymerase α subunit. *The EMBO Journal* **17**: 3439-3447.

Schneider, R., Lurz, R., Lüder, G., Tolksdorf, G., Travers, A. and Muskhelishvili, G. (2001). An architectural role of the *Escherichia coli* chromatin protein Fis in organising DNA. *Nucleic Acids Research* **29**: 5107-5114.

Schultz, S.C., Shields, G.C. and Steitz, T.A. (1991). Crystal structure of CAP-DNA complex: The DNA is bent by 90º. *Science* **253**: 1001-1007.

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

Sellitti, M.A., Pavco, P.A. and Steege, D.A. (1987). *lac* repressor blocks *in vivo* transcription of *lac* control region DNA. *Proceedings of the National Academy of Sciences* **84**: 3199-3203.

Semsey, S., Geanacopoulos, M., Lewis, D.E.A. and Adhya, S. (2002). Operator bound GalR dimmers close DNA loops by direct interaction: tetramerization and inducer binding. *The EMBO Journal* **21**: 4349-4356.

Sette, M., Spurio, R., Trotta, E., Brandizi, C., Brandi, A., Pon, C.L., Barbato, G., Boelens, R. and Gualerzi, C.O. (2009). Sequence-specific recognition of DNA by the C-terminal domain of nucleoid-associated protein H-NS. *The Journal of Biological Chemistry* **284**: 30453-30462.

Sharp, M.M., Chan, C.L., Lu, C.Z., Marr, M.T., Nechaev, S., Merritt, E.W., Severinov, K., Roberts, J.W. and Gross, C.A. (1999). The interface of σ with core RNA polymerase is extensive, conserved, and functionally specialised. *Genes and Development* **13**: 3015-3026.

Sheikh, J., Czeczulin, J.R., Harrington. S., Hicks. S., Henderson, I.R., Le Bouguénec, C., Gounon, P., Phillips, A. and Nataro, J.P. (2002). A novel dispersin protein in enteroaggregative *Escherichia coli*. *Journal of Clinical Investigation* **110**: 1329-1337.

Shimada, T., Fujita, N., Maeda, M. and Ishihama, A. (2005). Systemic search for the Cra-binding promoters using genomic SELEX. *Genes to Cells* **10**: 907-918.

Shimada, T., Fujita, N., Yamamoto, K. and Ishihama, A. (2011). Novel Roles of cAMP Receptor Protein (CRP) in Regulation of Transport and Metabolism of Carbon Sources. *PLOS ONE* **6**: (6).

Shin, M., Kang, S., Hyun, S.J., Fujita, N., Ishihama, A., Valentin-Hansen, P. and Choy, H.E. (2001). Repression of *deoP2* in *Escherichia coli* by CytR: conversion of a transcription activator into a repressor. *The EMBO Journal* **20**: 5392-5399.

Shin, M., Song, M., Rice, J.H., Hong, Y., Kim, Y., Seok, Y., Ha, K., Jung, S. and Choy, H.E. (2005). DNA looping-mediated repression by histone-like protein H-NS: specific requirement of Eσ^{70} as a cofactor for looping. *Genes and Development* **19**: 2388-2398.

Shine, J. and Dalgarno, L. (1974). The 3'-Terminal sequence of *Escherichia coli* 16S ribosomal RNA: Complementarity to nonsense triplets and ribosome binding sites. *Proceedings of the National Academy of Sciences* **71**: 1342-1346.

Shultzaberger, R.K., Bucheimer, R.E., Rudd, K.E. and Schneider, T.D. (2001). Anatomy of *Escherichia coli* ribosome binding sites. *Journal of Molecular Biology* **313**: 215-228.

Shultzaberger, R.K., Chen, Z., Lewis, K.A. and Schneider, T.D. (2006). Anatomy of *Escherichia coli* σ^{70} promoters. *Nucleic Acids Research* **35**:771-788.

Singh, S.S., Singh, N., Bonocora, R.P., Fitzgerald, D.M., Wade, J.T. and Grainger, D.C. (2014). Widespread suppression of intragenic transcription initiation by H-NS. *Genes and Development* **28**: 214-219.

Sixma, T.K., Pronk, S.E., Kalk, K.H., Wartna, E.S., van Zanten, B.A.M., Witholt, B. and Hol, W.G.J. (1991). Crystal structure of a cholera toxin-related heat-labile enterotoxin from *E. coli*. *Nature* **351**: 371-377.

Soto, G.E. and Hultgren, S.J. (1999). Bacterial adhesins: common themes and variations in architecture and assembly. *Journal of Bacteriology* **181**: 1059-1071.

Soutourina, O., Kolb, A., Krin, E., Laurent-Winter, C., Rimsky, S., Danchin, A. and Bertin, P. (1999). Multiple control of flagellum biosynthesis in *Escherichia coli*: Role of H-NS protein and the cyclic AMP-catabolite activator protein complex in transcription of the *flhDC* master operon. *Journal of Bacteriology* **181**: 7500-7508.

Spangler, B.D. (1992). Structure and function of cholera toxin and related *Escherichia coli* heat-labile toxin. *Microbiology and Molecular Biology Reviews* **56**: 622-647.

Steinsland, H., Valentiner-Branth, P., Perch, M., Dias, F., Fischer, T.K., Aaby, P., Mølbak, K. and Sommerfelt, H. (2002). Enterotoxigenic *Escherichia coli* infections and diarrhoea in a cohort of young children in Guinea-Bissua. *The Journal of Infectious Diseases* **186**: 1740-1747.

Stringer, A.M., Singh, N., Yermakova, A., Petrone, B.L., Amarasinghe, J.J., Reyes-Diaz, L., Mantis, N.J. and Wade, J.T. (2012). FRUIT, a scar-free system for targeted chromosomal mutagenesis, epitope tagging, and promoter replacement in *Escherichia coli* and *Salmonella enterica*. *PLOS ONE* **7**: (9).

Tan, K., Moreno-Hagelsieb, G., Collado-Vides, J. and Stormo, G.D. (2001). A comparative genomics approach to prediction of new members of regulons. *Genome Research* **11**: 566-584.

Tauschek, M., Gorrell, R.J., Strugnell, R.A. and Robins-Browne, R.M. (2002). Identification of a protein secretory pathway for the secretion of heat-labile enterotoxin by an enterotoxigenic strain of *Escherichia coli*. *Proceedings of the National Academy of Sciences* **99**: 7066-7071.

Taxt, A., Aasland, R., Sommerfelt, H., Nataro, J. and Puntervoll, P.Â. (2010). Heat-Stable Enterotoxin of Enterotoxigenic *Escherichia coli* as a Vaccine Target. *Infection and Immunity* **5**: 1824-1831.

Tebbutt, J., Rhodius, V.A., Webber, C.L. and Busby, S.J.W. (2002). Architectural requirements for optimal activation by tandem CRP molecules at a class I CRP-dependent promoter. *FEMS Microbiology Letters* **210**: 55-60.

Traxler, M.F., Summers, S.M., Nguyen, H.T., Zacharia, V.M., Hightower, G.A., Smith, J.T. and Conway, T. (2008). The global, ppGpp-mediated stringent response to amino acid starvation in *Escherichia coli*. *Molecular Microbiology* **68**: 1128-1148.

Turner, S.M., Chaudhuri, R.R., Jiang, Z.D., DuPont, H., Gyles, C., Penn, C.W., Pallen, M.J. and Henderson, I.R. (2006). Phylogenetic comparisons reveal multiple acquisitions of the toxin genes by enterotoxigenic *Escherichia coli* strains of different evolutionary lineages. *Journal of Clinical Microbiology* **44**: 4528-4536.

Turner, S.M., Scott-Tucker, A., Cooper, L.M. and Henderson, I.R. (2006). Weapons of mass destruction: virulence factors of the global killer enterotoxigenic *Escherichia coli*. *FEMS Microbiology Letters* **263**: 10-20.

Tutar, Y. (2008). *Syn*, *anti*, and finally both conformations of cyclic AMP are involved in the CRP-dependent transcription initiation mechanism in *E. coli lac* operon. *Cell Biochemistry Function* **26**: 399-405.

Typas, A. and Hengge, R. (2006). Role of the spacer between the -35 and -10 regions in σ^s promoter selectivity in *Escherichia coli*. *Molecular Microbiology* **59**: 1037-1051.

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

Vassylyev, D.G., Vassylyeva, M.N., Zhang, J., Palangat, M., Artsimovitch, I., Landick, R. (2007). Structural basis for substrate loading in bacterial RNA polymerase. *Nature* **448**: 163-169.

Viklund, H., Bernsel, A., Skwark, M. and Elofsson, A. (2008). SPOCTOPUS: a combined predictor of signal peptides and membrane protein topology. *Bioinformatics* **24**: 2928-2929.

Vrentas, C.E., Gaal, T., Ross, W., Ebright, R.H., Gourse, R.L. (2005). Response of RNA polymerase to ppGpp: requirement for the ω subunit and relief of this requirement by DksA. *Genes and Development* **19**: 2378-2387.

Wade JT, Belyaeva TA, Hyde EL and Busby SJW (2001). "A simple mechanism for co-dependence on two activators at an *Escherichia coli* promoter". *The EMBO Journal* **20**: 7160-7167.

Wade, J.T., Castro Roa, D., Grainger, D.C., Hurd, D., Busby, S.J.W., Struhl, K. and Nudler, E. (2006). Extensive functional overlap between σ factors in *Escherichia coli*. *Nature Structural and Molecular Biology* **13**: 806-814.

Wade, J.T. and Struhl, K. (2004). Association of RNA polymerase with transcribed regions in *Escherichia coli*. *Proceedings of the National Academy of Sciences* **101**: 17777-17782.

Wadler, C.S. and Vanderpool, C.K. (2007). A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide. *Proceedings of the National Academy of Sciences* **104**: 20454-20459.

Walker, R.L., Steele, D., Aguade, T., Ad Hoc ETEC Technical expert committee (2007). Analysis of strategies to successfully vaccinate infants in developing countries against enterotoxigenic *E.coli* (ETEC) disease. *Vaccine* **25**: 2545-2566.

Weber, I.T. and Steitz, T.A. (1987). Structure of a complex of catabolite gene activator protein and cyclic AMP refined at 2.5 Å resolution. *Journal of Molecular Biology* **198**: 311-326.

West, S.E., Sample, A.K., and Runyen-Janecky, L.J. (1994). The *vfr* gene product, required for *Pseudomonas aeruginosa* exotoxin A and protease production, belongs to the cyclic AMP receptor protein family. *Journal of Bacteriology* **176**: 7532-7542.

West, D., Williams, R., Rhodius, V., Bell, A., Sharma, N., Zou, C., Fujita, N., Ishihama, A. and Busby, S.J.W. (1993). Interactions between the *Escherichia coli* cyclic AMP receptor protein and RNA polymerase at class II promoters. *Molecular Microbiology* **10**: 789-797.

Williams, R., Bell, A., Sims, G. and Busby, S.J.W. (1991). The role of two surface exposed loops in transcription activation by the Escherichia coli CRP and FNR proteins. *Nucleic Acids Research* **19**: 6705-6712.

Wilson, C.J., Zhan, H., Swint-Kruse, L., and Matthews, K.S. (2007). The lactose repressor system: paradigms for regulation, allosteric behaviour and protein folding. *Cellular and Molecular Life Sciences* **64**: 3-16.

Yamanaka, H., Kameyama, M., Baba, T., Fujii, Y. and Okamoto, K. (1994). Maturation pathway of *Escherichia coli* heats-stable enterotoxin I: requirement of DsbA for disulphide bond formation. *Journal of Bacteriology* **176**: 2906-2913.

Yamanaka, H., Nomura, T., Fujii, Y. and Okamoto, K. (1998). Need for TolC, and *Escherichia coli* outer membrane protein, in the secretion of heat-stable enterotoxin I across the outer membrane. *Microbial Pathogenesis* **25**: 111-120.

Yang, J., Tauschek, M., Strugnell, R. and Robins-Browne, R.M. (2005). The H-NS protein represses transcription of the *eltAB* operon, which encodes heat-labile enterotoxin in enterotoxigenic *Escherichia coli*, by binding to regions downstream of the promoter. *Microbiology* **151**: 1199-1208.

Yu, J., Webb, H. and Hirst, T.R. (1992). A homologue of the *Escherichia coli* DsbA protein involved in disulphide bond formation is required for enterotoxin biogenesis in *Vibrio cholerae*. *Molecular Microbiology* **6**: 1949-1958.

Zhang, X. and Bremer, H. (1996). Effects of Fis on ribosome synthesis and activity and on rRNA promoter activities in *Escherichia coli*. *Journal of Molecular Biology* **259**: 27-40.

Zhou, Y., Kolb, A., Busby, S.J.W. and Wang, Y.P. (2014). Spacing requirements for class I transcription activation in bacteria are set by promoter elements. *Nucleic Acids Research* **42**: 9209-9216.

Zhou, Y., Zhang, X. and Ebright, R.H. (1993). Identification of the activating region of catabolite gene activator protein (CAP): isolation and characterization of mutants of CAP specifically

defective in transcription activation. *Proceedings of the National Academy of Sciences* **90**: 6081-6085.

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

Zhang, Y., Feng, Y., Chatterjee, S., Tuske, S., Ho, M.X., Arnold, E. and Ebright, R.H. (2012). Structural basis of transcription initiation. *Science* **338**: 1076-1080.

Zheng, D., Constantinidou, C., Hobman, J.L., Minchin, S.D. (2004). Identification of the CRP regulon using in vitro and in vivo transcriptional profiling. Nucleic Acids Research **32**: 5874-5893.

Zuo, Y., Wang, Y. and Steitz, T.A. (2013). The mechanism of *E.coli* RNA polymerase regulation by ppGpp is suggested by the structure of their complex. *Cell* **50**: 430-436.

Published papers

The Molecular Basis for Control of ETEC Enterotoxin Expression in Response to Environment and Host

James R. J. Haycocks[1], Prateek Sharma[1], Anne M. Stringer[2], Joseph T. Wade[2,3], David C. Grainger[1]*

1 Institute of Microbiology and Infection, School of Biosciences, University of Birmingham, Edgbaston, Birmingham, United Kingdom, 2 Wadsworth Center, New York State Department of Health, Albany, New York, United States of America, 3 Department of Biomedical Sciences, School of Public Health, University at Albany, SUNY, Albany, New York, United States of America

Abstract

Enterotoxigenic *Escherichia coli* (ETEC) cause severe diarrhoea in humans and neonatal farm animals. Annually, 380,000 human deaths, and multi-million dollar losses in the farming industry, can be attributed to ETEC infections. Illness results from the action of enterotoxins, which disrupt signalling pathways that manage water and electrolyte homeostasis in the mammalian gut. The resulting fluid loss is treated by oral rehydration. Hence, aqueous solutions of glucose and salt are ingested by the patient. Given the central role of enterotoxins in disease, we have characterised the regulatory trigger that controls toxin production. We show that, at the molecular level, the trigger is comprised of two gene regulatory proteins, CRP and H-NS. Strikingly, this renders toxin expression sensitive to both conditions encountered on host cell attachment and the components of oral rehydration therapy. For example, enterotoxin expression is induced by salt in an H-NS dependent manner. Furthermore, depending on the toxin gene, expression is activated or repressed by glucose. The precise sensitivity of the regulatory trigger to glucose differs because of variations in the regulatory setup for each toxin encoding gene.

Introduction

ETEC are Gram negative bacteria that cause severe diarrhoea, known as non-*vibrio* cholera, in humans [1,2]. First isolated in 1971, ETEC are responsible for 210 million infections annually, mostly in developing countries, leading to 380,000 deaths [3]. Disease results primarily from the action of two enterotoxins. The heat-labile toxin (LT) is similar in structure and function to cholera toxin [4,5]. The heat-stable toxin (ST) mimics the human hormone guanylin [6]. Both toxins are secreted by ETEC during infection. Made up of two subunits, encoded by the *eltAB* operon, LT has the configuration AB_5 [5,7]. In the gut, LT binds to host cell GM1 gangliosides and is endocytosed [8,9]. This triggers constitutive cAMP production in the affected cell [8]. The ST toxin, encoded by the *estA* gene, also interferes with cell signalling [6]. Hence, ST binds to the guanylate cyclase C receptor and stimulates overproduction of cGMP. The combined actions of LT and ST cause loss of H_2O, and electrolytes, from epithelial cells into the gut lumen [4]. Oral Rehydration Therapy (ORT) is used to redress the resulting electrolyte imbalance and rehydrate the patient [10]. In its most simple form, ORT requires only an aqueous solution of glucose and salt. Hence, the availability of metabolites and cations are a central theme of ETEC mediated disease. The effect of ORT on human physiology is well understood: glucose and Na_2^+ are transported across the epithelial membrane, along with water, to promote rehydration [11]. Surprisingly, despite the existence of molecular mechanisms that allow bacteria to respond to these signals, the consequences for ETEC are unknown.

In *E. coli*, the transcriptional response to glucose is controlled by cAMP receptor protein (CRP) [12]. In the absence of glucose, intracellular cAMP levels increase and CRP binds DNA targets with the consensus sequence 5′-TGTGA-n_6-TCACA-3′ [13]. Subsequently, gene expression is reprogrammed to make use of alternative carbon sources [14]. Note that the gene regulatory network managed by CRP includes many indirect pathways [14,15]. Hence, CRP is also a pleiotropic regulator of transcription. Whilst indirect regulatory effects are difficult to characterise, genes that are directly controlled by CRP can be divided into distinct classes [12]. At Class II targets, CRP binds to a site overlapping the promoter -35 element and interacts directly with both the N-terminal and C-terminal domains of the RNA polymerase α subunit (αNTD and αCTD). At Class I targets, CRP binds further upstream and interacts only with αCTD. This interaction can be further stabilised by UP-elements, AT-rich DNA sequences, adjacent to the CRP site, that facilitate αCTD-DNA interactions [12]. At both classes of promoter, the various contacts enhance gene expression by stabilising the transcription

Author Summary

Diarrheagenic illness remains a major disease burden in the developing world. Enterotoxigenic *Escherichia coli* (ETEC) are the leading bacterial cause of such disease; hundreds of millions of cases occur every year. The severe watery diarrhoea associated with ETEC infections results from the action of enterotoxins. The toxins target human gut epithelial cells and trigger the loss of water and electrolytes into the gut lumen. Oral rehydration therapy can counteract this process. Hence, glucose and salt solutions promote rehydration of the patient. In this work we show that the gene regulatory mechanisms controlling toxin expression respond directly to sugar and salt. Furthermore, we describe a molecular mechanism to explain these effects. Hence, we provide a starting point for the optimisation of oral rehydration solutions to reduce toxin expression over the course of an ETEC infection.

initiation complex. Unsurprisingly, most genes regulated by CRP encode proteins involved in metabolism. However, in some bacteria, CRP has been co-opted as a virulence regulator [16].

The Histone-like Nucleoid Structuring (H-NS) factor is a component of bacterial nucleoprotein. Consequently, H-NS also influences gene expression on a global scale [17]. Briefly, H-NS targets sections of the genome with a low GC content [17]. Depending on H-NS conformation, the resulting nucleoprotein complexes can be filamentous or bridged in organisation [18]. Filamentous complexes favour gene regulation by excluding RNA polymerase, and transcriptional regulators, from their targets [19,20]. Bridged complexes favour RNA polymerase trapping [21]. In all scenarios, it is thought that H-NS acts primarily to silence transcription [22]. The conformation of H-NS, and hence the way in which it modulates DNA topology, can be controlled by divalent cations. Consequently, H-NS mediated repression can be relieved by increased osmolarity [23]. Like CRP, H-NS has been incorporated into the virulence gene regulatory networks of many bacteria [17].

In this work we define the molecular trigger that controls toxin expression in ETEC. We show that CRP and H-NS are key regulatory factors. Strikingly, this allows ETEC to integrate extracellular signals of osmolarity and metabolism to control toxin production. Hence, we propose that ETEC toxicity responds directly to osmo-metabolic flux. Interestingly, the precise regulatory settings are different for each toxin encoding gene. The differences result from i) varying promoter configurations and ii) competition between CRP and H-NS for overlapping DNA targets. This is significant since fluctuations in osmolarity, and changes in the availability of metabolites, are central to ETEC infection and its treatment.

Results

Binding of CRP and H-NS across the ETEC H10407 genome

The prototypical ETEC strain H10407 reproducibly elicits diarrhoea in human volunteers and has a well-defined genome that shares 3,766 genes with *E. coli* K-12 [1]. Pathogenicity arises from 599 ancillary genes encoded by 25 discrete chromosomal loci and 4 plasmids. The plasmids, named p948, p666, p58 and p52, encode the enterotoxins. Derivatives of the *estA* gene are found on plasmids p666 (*estA1*) and p948 (*estA2*). A single copy of the *eltAB* operon is encoded by plasmid p666. We used Chromatin Immunoprecipitation (ChIP) coupled with next-generation DNA

sequencing (ChIP-seq) to map CRP and H-NS targets across the ETEC H10407 genome. The binding profiles are shown in Fig. 1A. In each plot genes are illustrated by blue lines (tracks 1 and 2), DNA G/C content by a cyan and pink graph (track 3), H-NS binding is in green (track 4) and CRP binding is shown in orange (track 5). As expected, H-NS binding is inversely correlated with DNA G/C content (compare tracks 3 and 4). Similarly, CRP binding occurs in expected locations; 96% of the CRP binding sites are associated with the DNA logo shown in Fig. 1B (i.e. the known CRP consensus sequence (13–15)). We identified a total of 111 high-confidence CRP targets (Table 1). Of these targets 93% were present in the genome sequences of both ETEC H10407 and *E. coli* K-12. The most common location for CRP sites was in intergenic regions (66% of targets) whilst a smaller number of targets were found within genes (34%). Consistent with expectations, CRP sites were most frequently located ~40.5 bp, or ~92.5 bp, upstream of experimentally determined transcription start sites (TSSs). Surprisingly, CRP binding was restricted to the ETEC chromosome (Fig. 1Ai). Conversely, H-NS bound to chromosomal and plasmid loci (Fig. 1Ai), including all toxin encoding genes (Fig. 1Aii).

Unoccupied high-affinity CRP binding targets on p948 and p666 are bound by H-NS

To better understand the lack of CRP binding to p948 and p666 we took a bioinformatic approach. CRP targets were aligned to generate a position weight matrix (PWM). The PWM was then used to search p948 and p666 for CRP sites. A continuum of over 100 potential CRP targets was identified. However, we recognise that the vast majority of these are likely to be false positives. Hence, we next sought to differentiate between genuine CRP sites and spurious predictions. To do this, predicted sites were scored, grouped, and ranked on the basis of their match to the PWM (Fig. 2A, S1 Table). Electrophoretic mobility shift assays (EMSA) were then used to measure binding of CRP to a target from each group so that a meaningful cut-off could be established. The result is illustrated graphically in Fig. 2B. The raw data are shown in S1A Fig. We found that predicted sites with a score<10 did not bind CRP. To assess the affinity of CRP for all predicted targets scoring >10 a second set of EMSA experiments was done (S1B Fig.). Hence, we identified a total of 5 potential CRP targets on p666 and p948. Interestingly, the *estA1* and *estA2* genes, which both encode ST, were amongst the 5 targets (Fig. 2C). Remarkably, all 5 of the plasmid borne CRP targets identified *in silico*, and bound tightly by CRP *in vitro*, were occupied by H-NS *in vivo* (Fig. 2C).

The *estA2* gene is transcribed from a Class I CRP dependent promoter

To understand if CRP could regulate ST production we focused first on *estA2*. This derivative of the toxin is more commonly associated with human disease and ETEC H10407 is somewhat unusual in also encoding *estA1* [24]. The sequence of the *estA2* regulatory region is shown in Fig. 3A. A 93 bp DNA fragment, containing the regulatory region, was cloned into the *lacZ* reporter plasmid pRW50 to generate a *lacZ* fusion (S2A Fig.). The *estA2* TSS was then determined using mRNA primer extension analysis. We detected a single extension product, of 109 nucleotides (nt) in length (Fig. 3B). The position of the TSS is labelled "+1" in Fig. 3A. Promoter -10 (5′-TTAAAT-3′) and -35 (5′-TTGCGC-3′) elements were observed at the expected positions upstream of the TSS. Throughout this work we refer to this promoter, highlighted purple in Fig. 3A, as P*estA2*. To confirm CRP binding
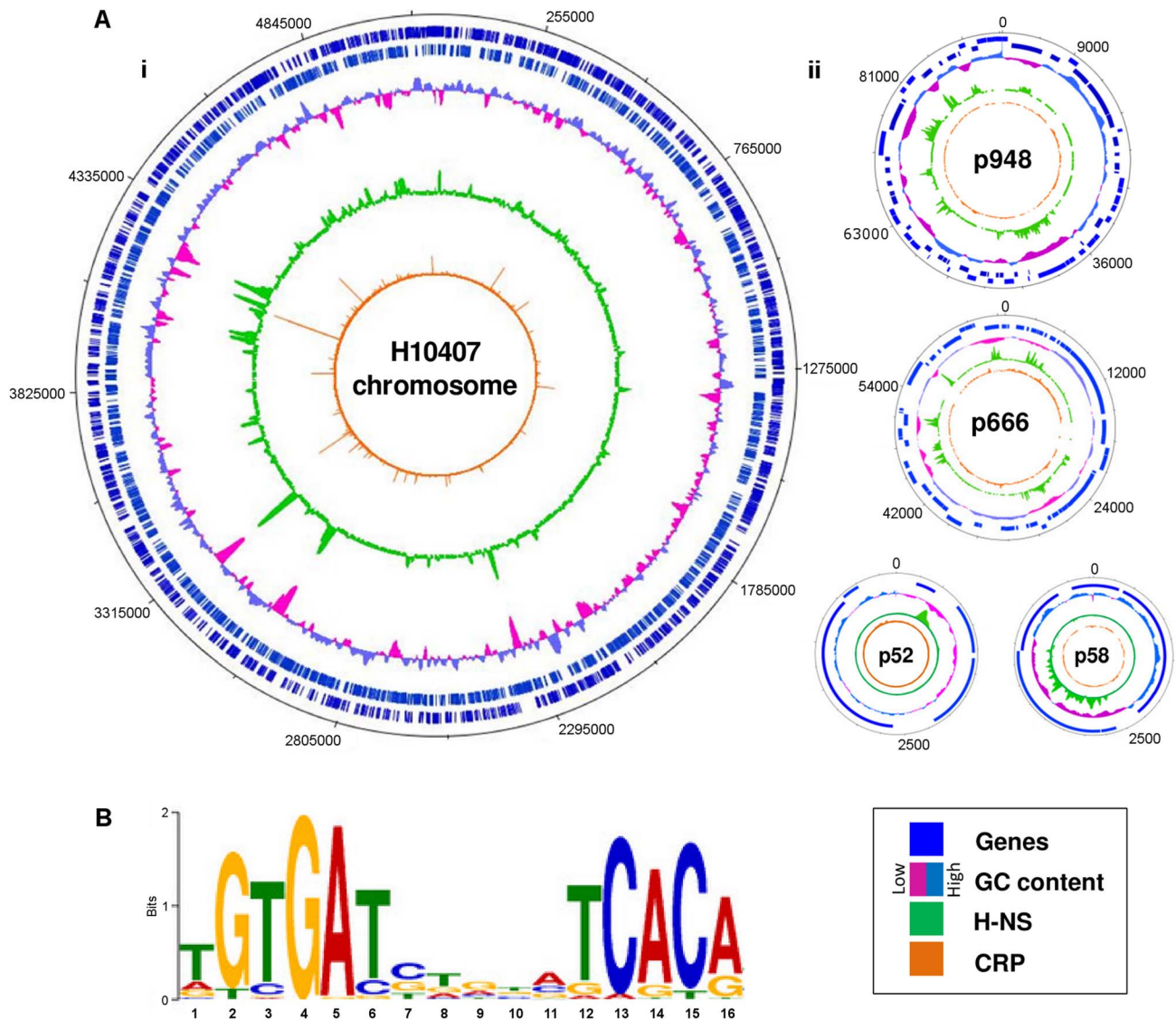
Fig. 1. Distribution of CRP and H-NS across the ETEC H10407 genome. A) The panel shows maps of the ETEC H10407 chromosome (i) and associated plasmids (ii). In each plot, tracks 1 and 2 (blue lines) show the position of genes, track 3 (purple and cyan graph) is a plot of DNA GC content, track 4 (green) is the H-NS binding profile and track 5 (orange) is the CRP binding profile. B) A DNA sequence motif generated by aligning regions of the ETEC H10407 chromosome bound by CRP.
doi:10.1371/journal.ppat.1004605.g001

at the predicted site we used DNase I footprinting (Fig. 3C). As expected, CRP protected the predicted target from digestion. Additionally, CRP induced DNase I hypersensitivity in the centre of the site. Note that the CRP site is centred 59.5 bp upstream of the TSS and adjacent to an AT-rich sequence that may be an UP element (Fig. 3A). Thus, we hypothesised that P*estA2* is a class I CRP activated promoter. To test our hypothesis we first determined whether CRP could indeed activate P*estA2*. To do this, we compared LacZ expression in M182Δ*lac* and M182Δ*lac*Δ*crp* cells carrying the P*estA2::lacZ* fusion. The data show that loss of CRP results in a 3-fold decrease in LacZ expression from P*estA2* (Fig. 3D). We next tested the ability of CRP to activate P*estA2 in vitro*. The 93 bp DNA fragment was cloned upstream of the λ*oop* terminator in plasmid pSR. In the context of this construct a 112 nt transcript is generated by RNA polymerase from P*estA2 in vitro*. The amount of transcript can then be quantified by electrophoresis. The result of the analysis,

with and without CRP, is shown in Fig. 3E. As expected, an intense band corresponding to the 112 nt transcript was observed. Production of the transcript was stimulated by CRP. Note that CRP had no effect on production of the 108 nt control RNAI transcript from the plasmid replication origin. Finally, we examined the AT-rich DNA sequence (highlighted blue in Fig. 3A) located between the CRP site and the promoter -35 element. We found that increasing the GC content of the putative UP-element altered migration of the 93 bp DNA fragment on an agarose gel, consistent with a change in DNA topology (S3A Fig.). Moreover, these changes to the UP-element rendered P*estA2* insensitive to CRP *in vivo* and *in vitro* (S3B Fig.).

H-NS excludes CRP from the *estA2* promoter and represses *estA2* transcription

Promoters can be liberated from H-NS repression if separated from flanking, H-NS bound, DNA [25]. We reasoned that this

Table 1. High-confidence CRP binding sites on the ETEC H10407 chromosome identified by ChIP-seq.

Peak Centre[a]	Binding Site(s)[b]	Gene(s)[c]	K-12 Homologues[d]
45284	**TGTGA**TTGGTAT**CACA**	ETEC_0040	caiT
92014	A**GTGA**TGGATG**TCAC**G	(ETEC_0078)	(cra)
176905	A**GCG**TTCCACG**TCACA**	(ETEC_0150)	(hemL)
408885	**TGTGA**TCTCTC**TCG**CA	ETEC_0385/ETEC_0386	yahN/yahO
461874	**TGTG**CGCAAGA**TCACA**	ETEC_0434	ddlA
463095	TT**TG**CGCGAGG**TCACA**	(ETEC_0436)	(phoA)
468009	A**GG**GATCTGCG**TCACA**	ETEC_0443	aroM
492973	ATC**GA**TTGCGT**TCAC**G	ETEC_0464	**tsx**
540805	**TGTGA**TCTTTAT**CACA**	ETEC_0511	maa
574230	GA**TGA**CGACGAT**CACA**	(ETEC_0538)	(ybaT)
683187	A**GTGA**TCGAGTTA**ACA**	ETEC_0628	**cstA**
697540	A**GTGA**TTTGCG**TCACA**	ETEC_0639	rnk
739223	C**GTT**ACCCTTG**TCG**CA	ETEC_0680	rihA
941002	**TGTGA**TGAGTAT**CAC**G	ETEC_0869	ybiJ
958866	**TGTGT**ACGAAAT**CACA**	ETEC_0886/ETEC_0887	**ybiS/ybiT**
1128472	n.d.	(ETEC_1030)	(yccS)
1205350	A**GTGA**TGTAGAT**CACA**	ETEC_1101	ycgZ
	TGAGATCGAGCA**CACA**		
1263558	T**TTGA**CGGCTAT**CAC**G	ETEC_1166	**ptsG**
1274886	**TGTGA**TCTGGAT**CACA**	ETEC_1176/ETEC_1177	**ycfQ/bhsA**
1301786	GA**TGA**TCCGCAT**CACA**	(ETEC_1206)/ETEC_1207	ETEC-specific/ETEC-specific
1348166	AT**TGA**ACAGGAT**CACA**	(ETEC_1259)/ETEC_1260	(rluE)/icd
1376374	G**GTGA**GCTGGC**TCACA**	ETEC_1292/ETEC_1293	**ycgB/dadA**
1388620	A**GTGA**GCCAGTTA**ACA**	(ETEC_1303)	(dhaI)
1541732	CG**TGA**ACCGGG**TCACA**	ETEC_1443/ETEC_1444	ycjZ/mppA
1567885	GTTA**AGT**AAAAT**CACA**	ETEC_1462/ETEC_1463	**paaZ/paaA**
1701402	**TGTGA**TGGATG**TCACT**	ETEC_1568	ydeN
1767726	**TGTGA**TTAACAG**CACA**	ETEC_1628	**mlc**
1777143	**TGTGA**TCTAGCG**CCAA**	ETEC_1637	pntA
1811426	CG**TGA**TCAAGAT**CAC**G	(ETEC_1668A)	(ETEC specific)
1859265	AT**TGA**GCGGGAT**CACA**	(ETEC_1713)	(sufS)
1887513	A**GTGA**TGCGCAT**CAC**G	ETEC_1737	aroH
	TGCGAGGTGTG**TCACA**		
2126754	**TGTG**GCGTGCAT**CACA**	n.a.	n.a.
2201816	G**GTGA**CGCGCG**TCACA**	ETEC_2057	yedP
2210222	C**GTGA**TCTCGCG**CACA**	ETEC_2065/ETEC_2066	yedR/ETEC-specific
2458348	**TGTGA**TCTGAAT**CTCA**	ETEC_2278	**cdd**
	TGCGATGCGTCG**CGCA**		
2492757	AT**TGA**TCGCCC**TCACA**	ETEC_2309	yeiQ
2555083	CG**TGA**CCAAAG**TCTCA**	(ETEC_2360)	(yfaQ)
2729713	TT**TGA**AGCTTG**TCACA**	ETEC_2510/ETEC_2511	mntH/nupC
2735124	A**GTT**ATTCATG**TCAC**G	ETEC_2514	yfeC
2795423	**TGTGA**GCCATGA**CACA**	(ETEC_2572)/ETEC_2573	**(aegA)/narQ**
2810983	C**GTGA**TCAAGA**TCACA**	ETEC_2586	**hyfA**
2887131	T**TTGA**TCTCGC**TCACA**	(ETEC_2666)/ETEC_2665	(xseA)/**guaB**
3012645	**TGTGA**TCCCCACA**ACA**	(ETEC_2793)	(**ung**)

Table 1. Cont.

Peak Centre[a]	Binding Site(s)[b]	Gene(s)[c]	K-12 Homologues[d]
3048307	TT**TGA**CGAGCAT**CACC**	(ETEC_2822)	(emrB)
3132920	G**GTGA**CCGGTTT**CACA**	ETEC_2905/ETEC_2906	ascG/ascF
3161660	**TGTGA**CCGTGG**TCGCA**	(ETEC_2933)	(nlpD)
3184337	CG**TGA**TGCGTGTA**ACA**	(ETEC_2956)/ETEC_2955	(cysI)/cysH
3196088	**TGTGA**TTACGAT**CACA**	ETEC_2966/ETEC_2967	ygcW/yqcE
3223792	A**GTGA**TCTTGAT**CTCA**	ETEC_2986	sdaC
	A**GT**T**A**TGTATC**T**AT**CA**		
3234980	**TG**C**GA**TCGTTA**TCACA**	(ETEC_2994)/ETEC_2995	(fucU)/fucR
3265047	**TGTGA**CCTGGG**TCAC**G	ETEC_3017	rppH
3324543	**TGTG**GGCTACGTA**ACA**	(ETEC_3075)	(ydhD)
3361162	n.d.	ETEC_3105	**serA**
3368992	**TTTGA**TGCACCG**CACA**	(ETEC_3113)	(ygfI)
3382158	**TGTGA**TCTACAA**CAC**G	ETEC_3126	cmtB
3390811	**TGTGA**TTTGCTT**CACA**	ETEC_3133	**galP**
3408173	**TGTGA**TGTGGATA**ACA**	ETEC_3154	**nupG**
3442697	**TGTGA**TGATTGT**CGCA**	ETEC_3186	ETEC-specific
3558573	A**GTGA**TTTGGC**TCACA**	ETEC_3291	ygiS
3580767	A**GTGA**CTTGCAT**CACA**	(ETEC_3318)	(yqiH)
3635301	AT**TGA**TCTAAC**TCAC**G	ETEC_3362	**uxaC**
3642302	CT**TGA**AGTGGG**TCACA**	(ETEC_3372)	(yqjG)
3665634	**TGTGA**TCAATGT**CA**AT	ETEC_3393/ETEC_3394	garP/garD
	TGTGCTTTAGCG**CGCA**		
3721308	G**GTGA**TTGATGT**CAC**C	(ETEC_3446)	(greA)
3785700	CG**TG**GGTCGCAT**CACA**	(ETEC_3510)	(mreC)
3878729	G**GTGA**TTTTGAT**CAC**G	ETEC_3614/ETEC_3615	ppiA/tsgA
3908574	G**GTGA**TCGCGC**TCACA**	(ETEC_3645)	(hofM)
3918861	**TGTGA**GTGGAA**TCGCA**	ETEC_3652/ETEC_3653	yhgE/pck
3986400	C**GTGA**TTTTATC**CACA**	ETEC_3707	**rpoH**
4105040	AG**TA**AGGCAAG**TCCC**T	n.a.	n.a.
4111116	**TGTGA**CGGGGCTA**ACA**	(ETEC_3806)	(wecH)
4153055	**TGTGA**TCTGAAT**CACA**	ETEC_3840	**yibI**
	TGTGATCTACAG**CA**TG		
4153191	**TGTGA**TTGATAT**CACA**	ETEC_3841	mtlA
	TGTGATGAACG**TCAC**G		
4158433	n.d.	ETEC_3846	lldP
4196869	**TG**CAATCGATA**TCACA**	ETEC_3886	dinD
4251326	CT**TA**CTCCTGC**TCACA**	ETEC_3938	ETEC specific
4266125	G**GTGA**TGGCATC**CGC**G	(ETEC_3956)	(nepI)
4290730	GG**TGA**GCAAAAC**CAC**G	(ETEC_3979)	(yidR)
4322430	AT**TGA**CCTGAGT**CACA**	(ETEC_4010)	(yieL)
4340544	CT**TGA**CCACGG**TCAG**A	(ETEC_4025)/ETEC_4024	(atpA)/atpG
4344649	**TGTGA**TCTGAAG**CAC**G	ETEC_4030	atpI
4373517	**TGT**AATGCTGGTA**ACA**	(ETEC_4051)	(ilvG)
4402013	C**GTG**CTGCATA**TCAC**G	(ETEC_4077)	(rffM)
4412999	C**GTGA**TCAATTTA**ACA**	ETEC_4085/ETEC_4085	hemC/cyaA
4438352	G**GTGA**TGAGTA**TCAC**G	ETEC_4107/ETEC_4108	ysgA/udp

Table 1. Cont.

Peak Centre[a]	Binding Site(s)[b]	Gene(s)[c]	K-12 Homologues[d]
	TGTGATTTGAA**TCAC**T		
4508745	**TGTGA**TATTTG**TCACA**	(ETEC_4165)/ETEC_4164	(fdhD)/fdoG
4517442	C**GTGA**TCGCTG**TCCCA**	(ETEC_4173)	(rhaA)
4564670	**TG**CGATCCGCC**TCAT**A	ETEC_4216/ETEC_4217	ptsA/frwC
4668870	**TGT**AACAGAGA**TCACA**	ETEC_4289/ETEC_4290	**malE/malK**
4725047	**TGTG**CGGATGA**TCACA**	n.a.	n.a.
4731402	**TGTGA**TCTTGCG**CAT**A	(ETEC_4365)	(aphA)
4761367	C**GTGA**TGGCTG**TCAC**G	ETEC_4389	fdhF
4846352	n.d.	ETEC_4464	ETEC-specific
4848117	C**GTGA**GTTCTG**TCACA**	n.a.	n.a.
4863253	T**TTGA**TCAACA**TCGCA**	(ETEC_4478)	(ETEC-specific)
4873926	G**GTGA**TCTATT**TCACA**	ETEC_4486/ETEC_4487	aspA/fxsA
4930149	**TGTGA**TGAACT**TCAA**A	ETEC_4545/ETEC_4546	yjfY/rpsF
4940903	**TGTGA**TCACTA**TCGCA**	ETEC_4557/ETEC_4558	ETEC-specific/ytfA
4993073	**TGTGA**CTGGTA**TCTC**G	(ETEC_4604)	(valS)
5002854	**TGT**AACCTTTG**TCACA**	ETEC_4610/tRNA-Leu	yjgB/tRNA-Leu
5030724	**TG**CGATGAATG**TCACA**	ETEC_4633/ETEC_4634	gntP/uxuA
5129400	CG**TA**CCGTCGG**TCACA**	(ETEC_4736)	(yjjI)
5129944	**TGTGA**TGTATA**TCGA**A	ETEC_4736/ETEC_4737	yjjI/deoC

[a]Chomosome coordinate of the ChIP-seq peak in H10407. Underlined text indicates that the ChIP-seq peak maps to sequence that is not conserved in *E. coli* K-12.
[b]CRP binding site sequence predicted by MEME. "n.d." indicates that MEME did not detect a putative binding site.
[c]Genes in parentheses indicate that the ChIP-seq peak is located within that gene. Downstream genes are only listed if the annotated gene start is ≤300 bp downstream of the CRP ChIP-seq peak. "n.a." indicates that no genes starts are ≤300 bp from the CRP ChIP-seq peak.
[d]*E. coli* K-12 homologues are listed for the ETEC genes in the previous column. Genes in parentheses indicate that the ChIP-seq peak is located within that gene. "n.a." indicates that no genes starts are ≤300 bp from the CRP ChIP-seq peak. "ETEC-specific" indicates that there is no K-12 homologue. Underlined genes have been identified as CRP targets in a previous ChIP-chip study [15]. Bold genes are listed as CRP targets in the Ecocyc database.
doi:10.1371/journal.ppat.1004605.t001

might be why, when isolated on the 93 bp fragment, P*estA2* was active and dependent on CRP. To test this logic we generated a further two P*estA2::lacZ* fusions using the pRW50 plasmid system. The additional P*estA2* DNA fragments were both 460 bp in length and include the full *estA2* gene that was entirely bound by H-NS in our ChIP-seq assay (Fig. 2C). The CRP site was ablated in one of the additional fragments by introducing point mutations that are predicted to disrupt CRP binding. The sequence of the DNA fragments is shown in S2A Fig. The *lacZ* fusions are illustrated graphically in Fig. 4A. Our expectation was that the longer 460 bp fragment would bind H-NS whilst the starting 93 bp fragment would not. To test this prediction we used ChIP. Thus, we compared H-NS binding to the different P*estA2* containing fragments *in vivo*. Fig. 4B shows results of a PCR analysis to measure enrichment of the P*estA2* locus. As expected, P*estA2* was only enriched in anti-H-NS immunoprecipitates when in the context of the 460 bp fragment. Crucially, enrichment is specific because, in a set of control PCR reactions, there was no enrichment of the *yabN* locus in any immunoprecipitate.

Our ChIP analysis suggests that the 460 bp fragment containing P*estA2* is subject to regulation by H-NS. To confirm that this was the case, the various pRW50 derivatives were used to transform M182Δ*lac* and M182Δ*lac*Δ*hns* cells. We then measured LacZ activity, driven by P*estA2*, in the transformants. Consistent with our expectations the data show that P*estA2* is repressed 5-fold by H-NS only in the context of the 460 bp DNA fragment (Fig. 4C).

Importantly, mutations in the CRP binding site abolish P*estA2* activity in the absence of H-NS. Hence, the measured LacZ expression must be driven by P*estA2* rather than any spurious promoters located within the *estA2* gene. Taken together our ChIP-seq and LacZ activity data show that H-NS prevents CRP from activating P*estA2*.

The *estA2* and *estA1* promoters are differently regulated by CRP but similarly regulated by H-NS

The *estA1* regulatory region, located on plasmid p666, contains a sequence similar to P*estA2* (Fig. 5A). We expected that this sequence would be the *estA1* promoter (P*estA1*). To test this expectation we created a 92 bp P*estA1::lacZ* fusion, equivalent to the 93 bp P*estA2::lacZ* fusion described above, and mapped the 5′ end of the resulting mRNA. As expected, the primer extension product was 109 nt in length (Fig. 5B). Hence, P*estA1* and P*estA2* use equivalent TSSs. However, we were surprised that the intensity of the P*estA1* primer extension product *increased* in cells lacking CRP (Fig. 5B). Closer examination of the alignment in Fig. 5A shows that, whilst P*estA1* and P*estA2* are similar, there are differences in the sequence and position of key promoter elements. To try and understand which changes result in the aberrant behaviour of P*estA1* we made a set of hybrid promoters. The hybrid constructs are derived from the CRP-activated *estA2* promoter. In each hybrid, named P*estA2.1* through P*estA2.7*, a region of P*estA2* was replaced with the equivalent region from
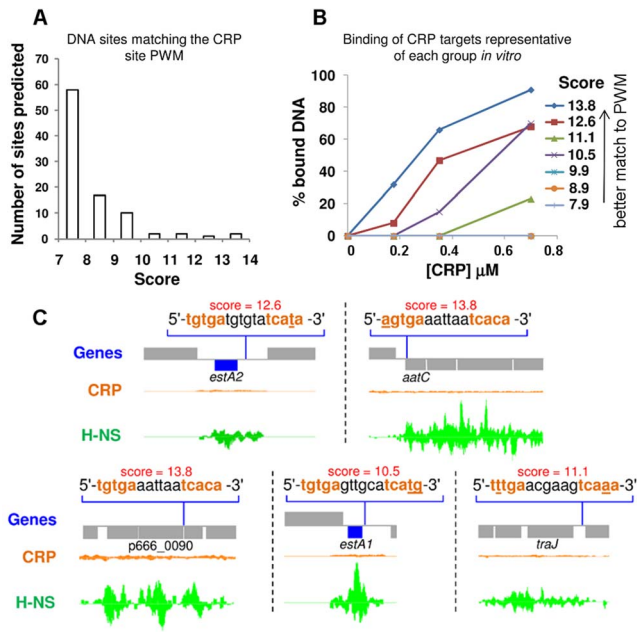
Fig. 2. Unoccupied CRP sites on p666 and p948 align with H-NS bound regions. A) A histogram showing the number of putative CRP binding sites in each of 7 discrete bins. Each bin is delineated by the "score" of the putative CRP site. A high score indicates a better match to the Position Weight Matrix that represents the consensus for CRP binding. B) The graph illustrates binding of CRP to a target from each of the bins shown in Panel A. CRP was used at concentrations of 0, 175, 350 or 700 nM. C) ChIP-seq data for CRP and H-NS binding at five regions of plasmids p666 and p948 that contain unoccupied CRP targets bound by CRP *in vitro*. The CRP and H-NS binding profiles are plots of sequence read counts at each position of the genome on both the top (above the central line) and bottom (below the central line) strand of the DNA. The y-axis scale is the same in each panel. The scale for H-NS binding is 1,785 reads on each strand and for CRP binding is 14,000 reads on each strand.

doi:10.1371/journal.ppat.1004605.g002

P*estA1* (see underlined sequences in Fig. 5C). The ability of the different hybrid promoters to drive *lacZ* expression, with and without CRP, was then tested. The results are shown in Fig. 5D. Note that, in Fig. 5D, the composition of each hybrid promoter is indicated in the grid below the graph. For example, P*estA2.1* is derived from P*estA2* but contains the P*estA1* CRP site. As expected, both P*estA1* and P*estA2* were able to drive *lacZ* expression but CRP had opposite effects. Moreover, maximal expression from P*estA1* was 3-fold lower than from P*estA2*. Only P*estA2.3* and P*estA2.5*, which both carried the same changes in the promoter -35 element, exhibited a reversed dependence on CRP. Hence, the P*estA1* -35 element must be responsible for the altered CRP dependence. All other hybrid promoters exhibited an overall reduction in activity compared to the parent P*estA2* construct. We conclude that this combination of changes results in the lower activity of P*estA1*. Note that both P*estA1* and P*estA2* were bound by H-NS in our ChIP-seq analysis (Fig. 2C). We reasoned that cloning P*estA1*, with flanking DNA, would reveal H-NS mediated repression. We generated a derivative of the P*estA1::lacZ* fusion where the downstream boundary was extended to include the entire *estA1* gene (S2B Fig., Fig. 5Ei). As expected, transcription from P*estA1* was repressed by H-NS in the presence of downstream DNA (Fig. 5Eii).

The *eltAB* operon is indirectly repressed by CRP and directly repressed by H-NS

We next turned our attention to the LT toxin promoter (P*eltAB*) [26,27]. Previously, Bodero and Munson [27] showed that transcription from this promoter was repressed by CRP. A mechanism for repression was proposed whereby CRP acted directly by binding three DNA targets overlapping P*eltAB* [27]. Even so, no CRP binding at P*eltAB* was identified by our ChIP-seq analysis (Fig. 6A). It is possible that this is because H-NS also excludes CRP from this locus (Fig. 6A). However, we also failed to identify CRP targets at P*eltAB* in our bioinformatic screen, even below the stringent cut-off (Fig. 2, S1 Table). In retrospect, this appears to be because all of three P*eltAB* CRP binding sites contain at least 4 mismatches to the consensus for CRP binding (Fig. 6A). Hence, we measured the affinity of CRP for P*eltAB* using EMSA assays. In parallel, we tested CRP binding to P*estA2* as a control. As expected, CRP bound tightly to P*estA2* at low concentrations (Fig. 6B, lanes 1–6). At high CRP concentrations further non-specific binding was observed (evidenced by a conspicuous "smear" in DNA migration in lane 7). In the equivalent experiment, with P*eltAB*, no specific binding of CRP was observed (lanes 8–13). However, non-specific CRP binding was again detectable at high protein concentrations (lane 14). Hence, CRP does not bind specifically to P*eltAB*. We hypothesised that previously observed changes in P*eltAB* activity, in cells lacking CRP, may occur indirectly. To test this, we cloned a 359 bp DNA fragment, containing P*eltAB*, into our pRW50 *lacZ* expression system. We also made a truncated 118 bp derivative of this construct where two of the three putative CRP targets were removed. A derivative of the truncated 118 bp construct, where the remaining CRP site was completely ablated by point mutations, was also made. The DNA sequences of the different constructs are shown in S2C Fig. They are illustrated graphically in Fig. 6Ci. Consistent with previous measurements, we found that transcription from P*eltAB* increased 2.5 fold in the absence of CRP. However, the response of P*eltAB* was identical when the CRP binding sites were removed (Fig. 6Cii). Hence, although CRP represses transcription from P*eltAB*, this must occur indirectly.

Given the configuration of H-NS binding at the *eltAB* locus (Fig. 6A) we reasoned that P*eltAB* would be repressed by H-NS in the presence of sufficient flanking DNA. As we had done previously for P*estA1* and P*estA2*, we compared the binding of H-NS to P*eltAB* in the presence and absence of the downstream flanking sequence. The different DNA constructs are illustrated in Fig. 7A and results of ChIP experiments to measure H-NS binding are shown in Fig. 7B. As predicted, enrichment of P*eltAB*, in immunoprecipitations with anti-H-NS, was only observed in the presence of downstream DNA. Importantly, this enrichment was specific to P*eltAB* and not observed for the control locus *yabN*. Corresponding LacZ activities, for the different DNA constructs, measured in M182 or the Δ*hns* derivative, are shown in Fig. 7C. Incorporation of flanking DNA downstream of P*eltAB* resulted in a 15-fold reduction in LacZ activity that was largely relieved in the absence of H-NS.

CRP and H-NS allow the *estA1*, *estA2* and *eltAB* promoters to respond to glucose and salt

Given the established regulatory connections between CRP and glucose, and between H-NS and salt, we next measured changes in the activity of P*estA1*, P*estA2* and P*eltAB* in response to glucose and salt. A complete description of assay conditions is provided in the Materials and Methods section. Briefly, to establish the range
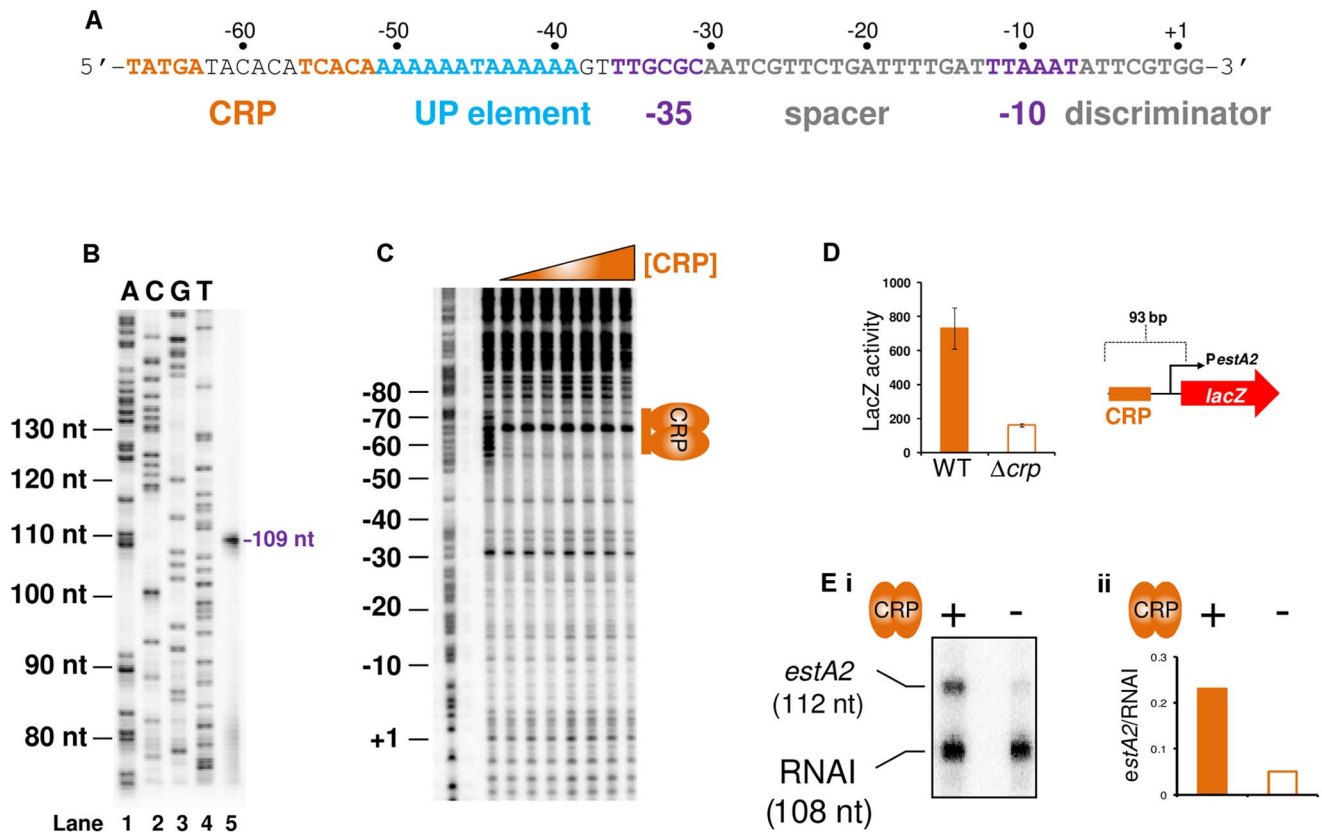
Fig. 3. The *estA2* promoter is activated by a Class I CRP dependent mechanism. A) Sequence of the *estA2* gene regulatory region. The CRP binding site is shown in orange, the UP element is blue and the promoter -10 and -35 elements are shown in purple. The different promoter positions are numbered relative to the transcription start site (+1). B) Location of the P*estA2* transcription start site. The gel shows the product of an mRNA primer extension analysis to determine the *estA2* transcription start site (Lane 5). The gel was calibrated using arbitrary size standards (A, C, G and T in Lanes 1–4). C) Binding of CRP to P*estA2*. The panel shows the result of a DNAse I footprint to monitor binding of CRP to the 93 bp P*estA2* DNA fragment. The gel is calibrated with a Maxim-Gilbert DNA sequencing reaction. CRP was added at concentrations of 0.35–2.1 µM. D) CRP is required for transcription from P*estA2 in vivo*. The panel shows a cartoon representation of the 93 bp P*estA2::lacZ* fusion and a bar chart illustrates LacZ activity in lysates of cells carrying this fusion. Assays were done in LB medium. E) i) Stimulation of P*estA2* by CRP *in vitro*. The figure shows the results of an *in vitro* transcription reaction. The 112 nt transcript initiates from P*estA2* and the 108 nt RNAI transcript is an internal control. CRP was added at a concentration of 350 nM and RNA polymerase was added at a concentration of 400 nM. ii) quantification of band intensities from the *in vitro* transcription analysis.
doi:10.1371/journal.ppat.1004605.g003

of conditions across which the promoters were able to respond, we examined the effect of titrating glucose or salt into the growth medium individually. In all experiments, we used the promoter::*lacZ* fusions that included downstream flanking DNA. This was to ensure that signals sensed by both CRP and H-NS could be integrated. As expected, the activity of P*estA1* was low. Consequently, the effects of glucose and salt were negligible (S4A Fig.). Conversely, the activity of P*estA2* was sensitive to both glucose and salt (S4B Fig.). Thus, *lacZ* expression driven by P*estA2* was repressed by glucose (orange line) and enhanced by salt (green line). As expected, P*eltAB* activity increased in the presence of both salt and glucose, but induction by salt was more prominent (S4C Fig.). We hypothesised that, for P*estA2*, the inhibitory effect of glucose should override the stimulatory effect of salt. Our reasoning was that, although H-NS can repress P*estA2*, the promoter is ultimately dependent on CRP for activity. Hence, we examined the effect of adding salt and glucose, to cells carrying the P*estA2::lacZ* fusion, separately and in combination (Fig. 8A). As predicted, the inhibitory effect of glucose was dominant (Fig. 8Ai) and was still observed in the absence of H-NS (Fig. 8Aii). Conversely, the stimulatory effect of salt required H-NS (compare

green bars in Fig. 8). Importantly, in a separate experiment, we also showed that the effect of glucose on P*estA2* activity requires that the CRP site is intact (S4D Fig.). The combined effect of salt and glucose on P*eltAB* was more difficult to predict because CRP acts via an undefined, and indirect, mechanism. The result of the analysis (Fig. 8B) shows that the stimulatory effects of salt and glucose on transcription from P*eltAB* are not additive. Moreover, the stimulatory effect of glucose requires H-NS.

The response of P*eltAB* and P*estA2* to CRP and H-NS is conserved in other ETEC isolates and during host cell attachment

Examination of all sequenced ETEC genomes reveals slight variations in the sequence of the *eltAB* and *estA2* promoter sequences (recall that ETEC H10407 is somewhat anomalous in also encoding *estA1*). Thus, we next sought to understand if our model for regulation of LT and ST expression was broadly applicable. We focused our efforts on ETEC E24377A since i) the genome has been sequenced and ii) a vast array of independently generated transcriptomic data are available for this organism
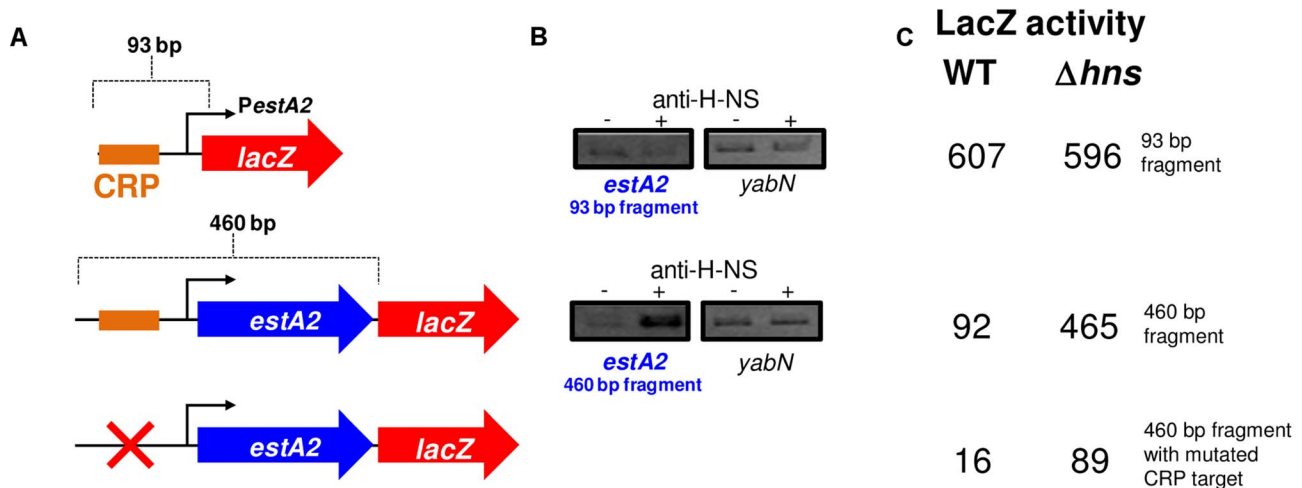
Fig. 4. The *estA2* promoter is repressed by H-NS. A) The panel shows different P*estA2::lacZ* fusions. The *lacZ* gene is shown as a red arrow and the *estA2* gene is shown as a blue arrow. P*estA2* is illustrated using a bent arrow and the CRP binding site is shown as an orange box. B) H-NS binds to P*estA2* only in the presence of flanking DNA. ChIP-PCR was used to measure binding of H-NS to the different P*estA2* derivatives cloned in pRW50. PCR products were generated using primers that could detect P*estA2* in the context of both the 93 bp fragment and the longer 460 bp fragment. C) The values are β-galactsidase activity values for lysates of M182, or M182Δ*hns*, carrying the different P*estA2* derivatives. Assays were done in LB medium.
doi:10.1371/journal.ppat.1004605.g004

[28,29]. Using ETEC E24377A DNA as a template, we generated a 460 bp P*estA2*, and 1126 bp P*eltAB* DNA fragment. The sequences are shown in S2D Fig. The DNA fragments were cloned into pRW50 and the ability of the promoters to drive *lacZ* expression in response to CRP and H-NS was measured. As expected, transcription from P*estA2* was repressed by H-NS and activated by CRP whilst P*eltAB* was repressed by H-NS (Fig. 9A). We observed no effect of CRP on transcription from *PeltAB* in the context of the 1126 bp ETEC E24377A fragment. This is not unexpected because CRP acts indirectly and these indirect CRP effects have only previously been observed in the context of short DNA fragments containing P*eltAB* that are not subject to direct repression by H-NS. We note that Sahl and Rasko previously examined the global transcriptome response of E24377A to glucose levels and bile salts [28]. In exact agreement with our model for toxin regulation, and the data in Fig. 9A, this study confirmed that i) salt induced expression of both toxins and ii) glucose inhibited expression of *estA2* [28]. Fortuitously, changes in the ETEC E24377A transcriptome, prompted by ETEC attachment to human gut epithelial cells, have also been quantified comprehensively [29]. Briefly, in these experiments, ETEC were added to sets of Caco-2 intestinal epithelial cell tissue cultures. Over a time course, ETEC that had adhered to host cells were separated from non-adhered ETEC. The transcriptomes of adhered and non-adhered ETEC were then compared. By mining these data, we next sought to determine if our model was consistent with observed changes in the transcription of *crp*, *hns*, *eltA* and *estA* during host cell attachment. Briefly, our data predict that changes in *estA* expression should be directly correlated to changes in the level of CRP and inversely correlated with changes in levels of H-NS. Conversely, levels of *eltA* expression should be inversely correlated with levels of H-NS. The result of the analysis is illustrated in Fig. 9B. The data show that the relative levels of *crp* transcription in attached and unattached cells are similar (orange line). However, levels of *hns* transcription change dramatically (green line) 60 minutes after host cell attachment. As predicted by our model, levels of *estA2* and *eltA* transcription (dashed lines) inversely track changes *hns* transcript levels. When undertaking this analysis we noticed that, although there was little

change in the relative level of *crp* mRNA between attached and unattached ETEC cells, the absolute level of *crp* mRNA did fluctuate across the time course of the experiment and between biological replicates. Strikingly, when these absolute mRNA levels are compared there is a clear linear relationship between *crp* and *estA2* expression (Fig. 9C). Note that in Fig. 9C the absolute level of *hns* mRNA has been added in parenthesis for each data point. Remarkably, the only two outlying data points in this plot correspond to the two samples with increased *hns* expression. We conclude that regulation of *estA2* and *eltA* by CRP and H-NS is important during the attachment of ETEC to human intestinal epithelial cells, and that the regulatory control of ETEC toxins is conserved across different strains.

Disrupting the regulatory switch attenuates ETEC virulence

Taken together, our data suggest that CRP and H-NS form a regulatory switch that controls ETEC toxicity. We next sought to examine the effect of disabling the switch on virulence. This is not straightforward because no animal model faithfully mimics the disease caused by ETEC in humans. However, intranasal mouse models have been used as a proxy for measuring *E. coli* pathogenicity [30]. Importantly, pathogenic *E. coli* cause more severe disease in this model than non-pathogenic strains [30]. Furthermore, ETEC strains lacking genes encoding toxins and known colonisation factors are less virulent in this model [31]. We opted to disrupt the regulatory switch by removing the *crp* rather than the *hns* gene. This was a deliberate decision since *E. coli* strains lacking *hns* are severely attenuated for growth in laboratory conditions. Conversely, the *crp* null derivative of ETEC H10407 was only mildly compromised for growth in liquid culture. Hence, we compared pathogenicity of ETEC H10407, and the *crp* derivative, using the intranasal mouse model [30]. Note that the outcome of this experiment is difficult to predict since the effects of CRP on pathogenicity likely go far beyond the control of toxin expression. However, it is reasonable to assume that ETEC virulence should differ in cells lacking *crp*. The median survival of mice challenged with wild type ETEC was 53 hours and the
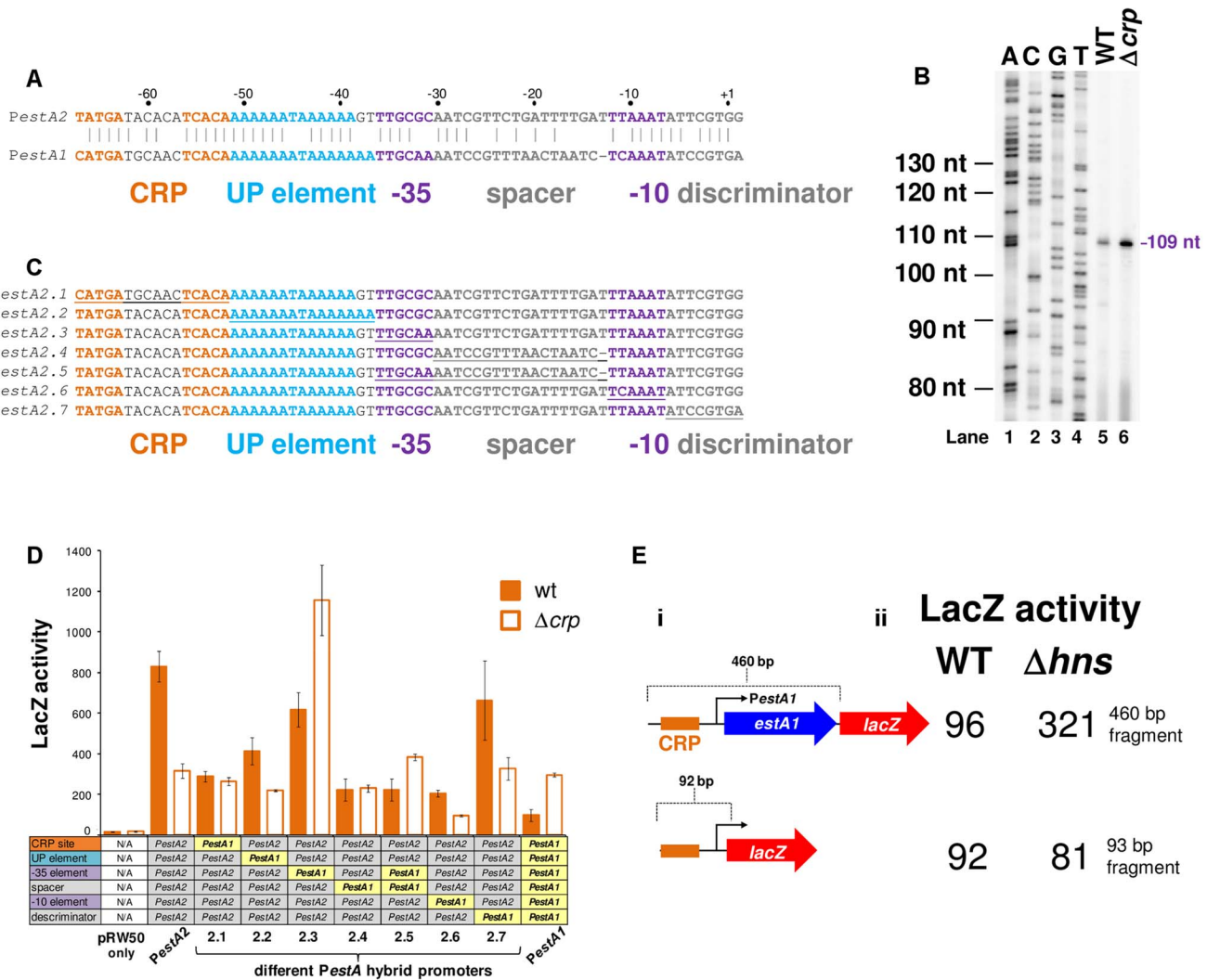
Fig. 5. Comparison of P*estA1* and P*estA2* reveals differential activity and regulation by CRP. A) Comparison of P*estA1* and P*estA2*. The panel shows the DNA sequences of P*estA1* and P*estA2*. Bases that are identical are highlighted by a solid vertical line. The CRP sites are shown in orange, the UP element in blue and the core promoter elements in purple. The sequences are numbered with respect to the transcription start site (+ 1). B) Location of the P*estA1* transcription start site. The gel shows products from an mRNA primer extension analysis (Lanes 5 and 6). The gel was calibrated using arbitrary size standards (A, C, G and T in Lanes 1–4). C) Sequences of hybrid *estA* promoters. The sequences labelled *estA2.1* through *estA2.7* are derivatives of the 93 bp P*estA2* DNA fragment where different sequence elements have been replaced with the equivalent sequence from P*estA1*. D) The bar chart shows β-galactosidase activity measurements for lysates obtained from cultures of M182, or the Δ*crp* derivative, containing the indicated hybrid promoter fragment was fused to *lacZ*. E) The panel shows different P*estA1::lacZ* fusions. The *lacZ* gene is shown as a red arrow and the *estA1* gene is shown as a blue arrow. P*estA1* is illustrated using a bent arrow and the CRP binding site is shown as an orange box. Assays were done in LB medium.
doi:10.1371/journal.ppat.1004605.g005

mortality rate was 100%. Conversely, the median survival of mice challenged with Δ*crp* ETEC was 72 h and 20% of the mice survived (Fig. 9D). Thus, whilst the full extent to which CRP co-ordinates the ETEC virulence programme remains to be determined, CRP is clearly central to the pathogenic response.

Discussion

A complex hierarchy of salt and glucose-dependent regulation controls toxin expression

We propose that toxin expression in ETEC can be controlled by osmo-metabolic flux. This is relevant to conditions in the small intestine (osmolarity equivalent to 300 mM NaCl) disease symp-toms (the extrusion of cations and cAMP into the gut lumen) and

treatment (the ingestion of solutions containing glucose and salt) [7–11,32]. A molecular model, describing how the different signals are integrated, is illustrated in Fig. 10. Two gene regulatory proteins, CRP and H-NS, are central to our model. Hence, H-NS directly represses the expression of *eltAB*, *estA1* and *estA2* (pathways "a" and "b" in Fig. 10). For *estA2* and *eltAB* this repression can be relieved, in an H-NS dependent manner, by increased osmolarity. At P*estA2* CRP directly activates transcrip-tion by a Class I mechanism (pathway "c"). H-NS can interfere with this process by competing with CRP for binding at P*estA2* (pathway "d"). Finally, CRP can indirectly repress expression of *eltAB* via an unknown pathway that is influenced by H-NS ("e"). Both pathways "c" and "e" are sensitive to glucose availability because of their dependence on CRP. We speculate that pathway
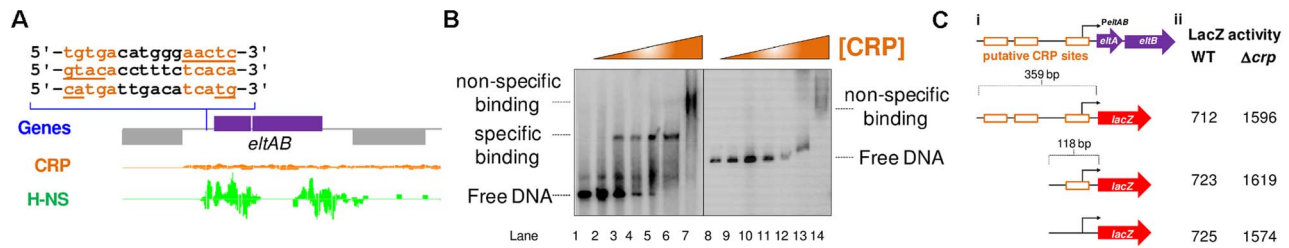
Fig. 6. The *eltAB* promoter is indirectly repressed by CRP. A) The Panel shows ChIP-seq data for CRP and H-NS binding at the *eltAB* locus. The sequence of 3 putative CRP binding sites proposed by Bodero and Munson (2009) are shown. The CRP and H-NS binding profiles are plots of sequence read counts at each position of the genome on both the top (above the central line) and bottom (below the central line) strand of the DNA. The y-axis scale for H-NS binding is 1,785 reads on each strand and for CRP binding is 14,000 reads on each strand. B) Results of an Electrophoretic Mobility Shift Assay to measure binding of CRP to the 93 bp P*estA2* fragment (Lanes 1–7) or the 359 bp P*eltAB* fragment (Lanes 8–14). Specific and non-specific binding of CRP is indicated to the left and right of the gel. CRP was added at a concentration of 0.2–7.0 μM. C) Panel (i) shows different P*eltAB::lacZ* fusions. The *lacZ* gene is shown as a red arrow and the *eltAB* operon is shown in purple. P*eltAB* is illustrated using a bent arrow and the putative CRP binding sites are shown as open orange boxes. In panel (ii) the values are β-galactsidase activity measurements taken in M182 or the Δ*crp* derivative. Assays were done in LB medium.
doi:10.1371/journal.ppat.1004605.g006

"e" may include H-NS since the effects of salt and sugar on *eltAB* expression were epistatic (Fig. 8). Our model for H-NS repression of *eltAB* is consistent with previous work [26]. However, our conclusion that *eltAB* is indirectly repressed by CRP disagrees with a previous study [27]. Even so, we were able to faithfully reproduce most of the observations previously described by Bodero and Munson [27]. We note that Bodero and Munson previously suggested that CRP may bind targets at P*eltAB* with a 7, rather than 6, base pair spacer between the two CRP half sites. Such CRP targets have never been described amongst hundreds of known CRP regulated promoters. Furthermore, we found no such CRP sites in our ChIP-seq analysis. Given that these DNA sequences can be deleted, without negating the effect of CRP on P*eltAB* activity, the regulatory effect of CRP must be indirect.

Oral Rehydration Therapy is likely to impact on toxin expression

Our model for regulation of ST and LT expression is pertinent to both ETEC mediated disease and its treatment. ST and LT trigger the extrusion of H_2O, cations, and cAMP (the cofactor for CRP) from the small intestine into the gut lumen [4–9]. Furthermore, solutions of salt and glucose are consumed by patients to reverse this process [10,11]. We speculate that, during

infection, extrusion of electrolytes and cAMP into the gut lumen could create a positive feedback loop to drive toxin expression. Importantly, our model also suggests that ORT may provide benefits beyond stimulating rehydration of the patient. The concentration of glucose used in ORT is ~10-fold higher than is required to repress *estA2* expression. Hence, even if 90% of glucose present in ORT solutions is absorbed before reaching the site of infection, sufficient glucose should be present to down regulate toxin expression. Furthermore, even though salt is able to induce expression of *estA2* and *eltAB*, the effect is only observed at concentrations far higher than those found in ORT solutions.

Differential regulation of *estA1* and *estA2* by CRP

Our observation that *estA1* and *estA2* are oppositely regulated by CRP is intriguing given the similarities between the promoter sequences of these genes. Differential regulation is dependent on the promoter -35 element (Fig. 5). At Class I CRP regulated promoters an αCTD protomer sits between CRP and domain 4 of the RNA polymerase σ subunit, which is bound to the promoter -35 element [12]. Thus, one possible explanation is that changes in the -35 element result in subtle repositioning of σ. This could result in unproductive interactions between αCTD and σ when CRP is present.
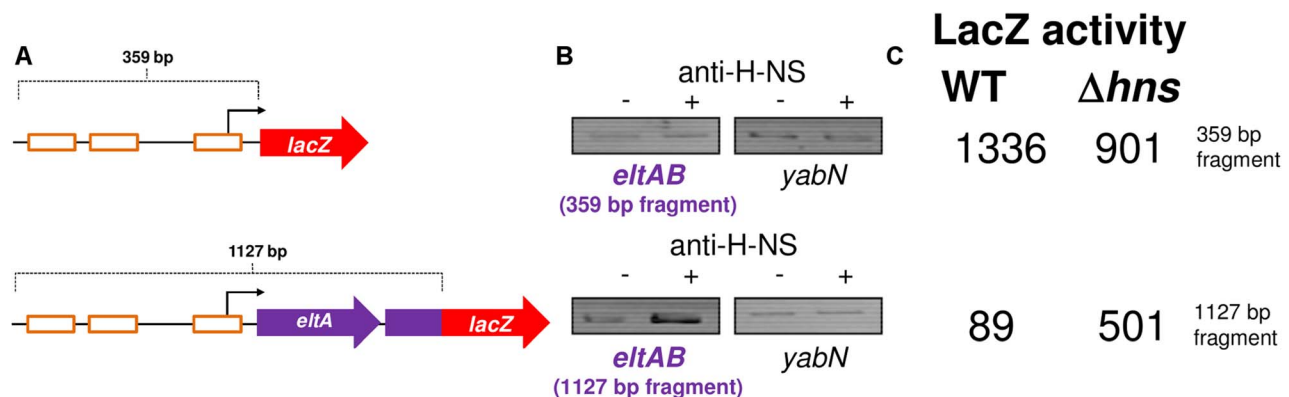
Fig. 7. The *eltAB* promoter is directly repressed by H-NS. A) The panel shows two P*eltAB::lacZ* fusions. B) The results of a ChIP-PCR analysis used to measure binding of H-NS to the two different P*eltAB* derivatives shown in Panel A. C) The values are β-galactsidase activity measurements for lysates of M182, or the Δ*hns* derivative, carrying the different P*estA2::lacZ* fusions. Assays were done in LB medium.
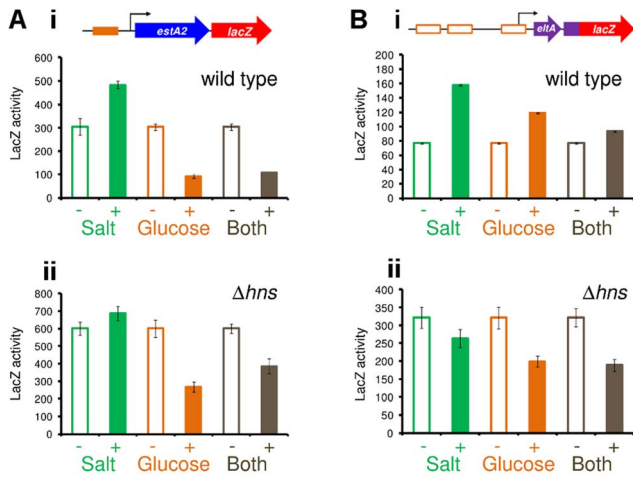doi:10.1371/journal.ppat.1004605.g007

Fig. 8. H-NS and CRP integrate signals of osmolarity and metabolism to control expression of LT and ST. The figure shows β-galactosidase activity measurements for lysates obtained from cultures of M182 (i) or the Δhns derivative (ii) containing A) PestA2 or B) PeltAB fused to lacZ in plasmid pRW50. Cultures were grown in the presence and absence of 2% glucose and/or salt (60 mM NaCl and 20 mM KCl). Assays were done in M9 minimal medium so that the glucose and salt concentrations could be more accurately controlled. doi:10.1371/journal.ppat.1004605.g008

H-NS prevents CRP regulation of select target genes

Our data indicate that several strong CRP binding sites in the H10407 genome are occluded by H-NS. This strongly suggests that the CRP regulon has evolved to incorporate additional environmental signals through the action of H-NS. The repressive effect of H-NS on transcription has been widely described [23]. H-NS represses transcription predominantly by occluding the binding of RNAP or by trapping RNAP at promoters [20]. Recently, it was shown that H-NS occludes many binding sites for the CRP homologue, FNR, in *E. coli* [21]. Thus, occlusion of transcription factor binding sites appears to be a major function of H-NS, especially for CRP family proteins. Note that, in order to exclude CRP from target promoters, sites of H-NS nucleation and CRP binding need not overlap precisely. For example, at both *estA1* and *estA2*, maximal H-NS binding is observed within the coding sequence of the gene (Fig. 2C). Despite this, H-NS oligomerisation across adjacent DNA is sufficient to prevent CRP binding.

Conclusions

In summary, our model provides a framework for better understanding ETEC mediated disease and its treatment. Moreover, our catalogue of CRP and H-NS binding targets provide a useful community resource for further studies of all *E. coli* strains. In particular, our ChIP-seq data for CRP report >50 targets not identified previously in *E. coli* K-12 and 8 ETEC-specific targets. Finally, our data show how very small changes in the organisation
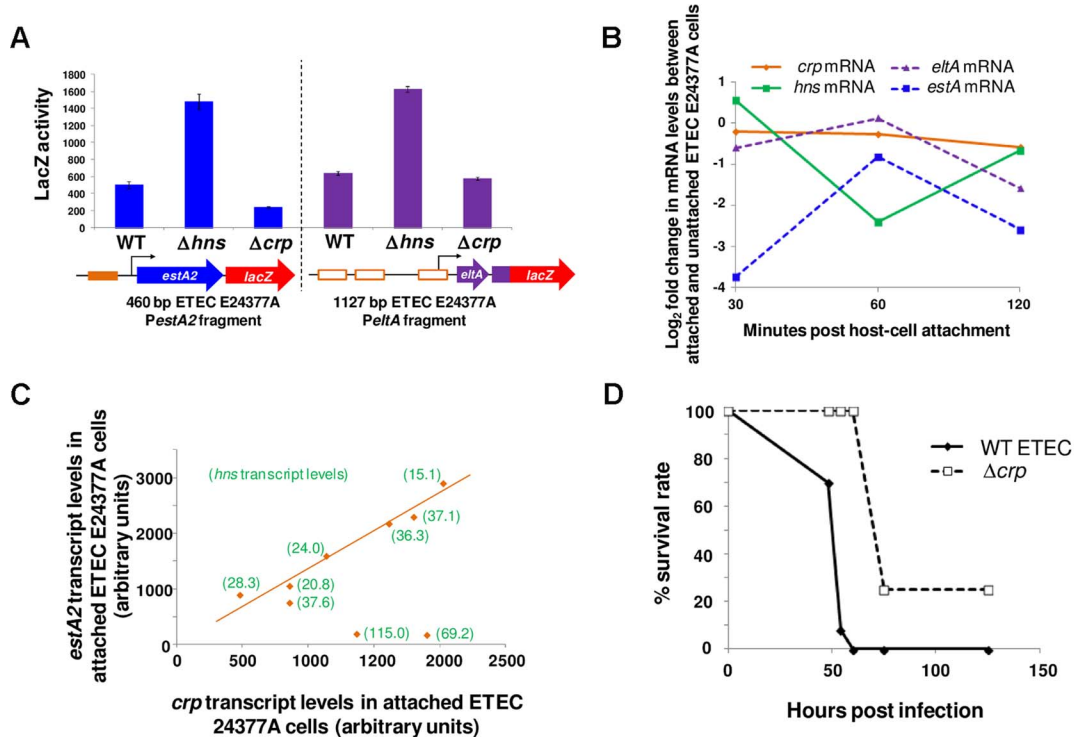

Fig. 9. Modulation of estA2 and eltA transcription during attachment of ETEC E24377A to gut epithelial cells. A) The figure shows β-galactosidase activity measurements for lysates obtained from cultures of M182 or the Δhns and Δcrp derivatives containing PestA2 (460 bp fragment) or PeltAB (1126 bp fragment) from ETEC 24377A fused to lacZ in plasmid pRW50. B) The panel shows log2 fold changes in the transcription of crp, hns, eltA and estA in ETEC E24377A cells over a two hour incubation with a Caco-2 intestinal epithelial cell culture (29). The log2 values represent the fold change in transcription between ETEC cells attached and unattached to Caco-2 cells at each time point. C) The panel shows a scatter plot of absolute crp and estA2 mRNA levels in ETEC E24377A attached to Caco-2 intestinal epithelial cells. Each data point represents a different biological replicate. For each data point the absolute level of hns mRNA is shown in parenthesis. D) The panel shows the survival rate of BALB/C mice (n = 30) after intranasal inoculation with wild type ETEC H10407 or the Δcrp derivative. doi:10.1371/journal.ppat.1004605.g009
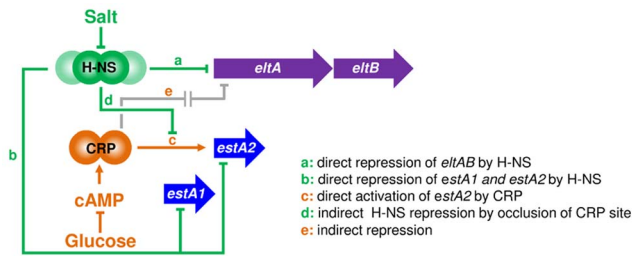
Fig. 10. An osmo-metabolic gene regulatory circuit comprised of CRP and H-NS controls expression of LT and ST. The diagram illustrates the regulatory effects of salt, cAMP and glucose on transcription from the various ST and LT promoter regions. doi:10.1371/journal.ppat.1004605.g010

of gene regulatory regions can have major effects on gene expression, such that transcription responds differently to the same environmental cues.

Materials and Methods

Strains, plasmids and oligonucleotides

ETEC strain H10407 is described by Crossman et al. [1]. The C-terminal crp-3×FLAG tag was introduced into the H10407 chromosome using the recombineering method of Stringer et al. [33]. Wild type E. coli K-12 strains JCB387 and M182 have been described previously [34,35]. The Δhns M182 derivative was generated by P1 transduction of hns::kan from E. coli K12 derivative YN3144 (a gift from Ding Jin). Plasmids pRW50 and pSR are described by Lodge et al. [36] and Kolb et al. [37]. More detailed descriptions of strains and plasmids, along with the sequences of oligonucleotides, are provided in S2 Table.

ChIP-seq

Cultures were grown to mid-log phase in M9 minimal medium with 1% (w/v) fructose at 37°C. Targeted ChIP experiments (Fig. 4 and 6) were done exactly as described by Singh and Grainger [38] using PestA2 or PeltAB fragments cloned in pRW50 carried in strain M182. The ChIP-seq was done as described extensively by Singh et al. [25] using strain H10407. Briefly, H-NS and CRP-3×FLAG were immunoprecipitated using protein A sepharose (GE Healthcare) in combination with 2 μL of anti-H-NS or 2 μl of anti-FLAG respectively. After immunoprecipitation and washing, beads were resupended in 100 μL 1× Quick Blunting Buffer (NEB) with dNTPs (as specified by the manufac-turer) and 2 μL Quick Blunting Enzyme Mix, and incubated for 30 minutes at 24°C with gentle mixing. After being collected by centrifugation, the beads were again washed and the associated DNA was A-tailed by resuspension of beads in 100 μL 1× NEB buffer #2 supplemented with 2 mM dATP and 10 units of Klenow Fragment (3'→5' exo-; NEB). Following incubation for 30 minutes at 37°C, with gentle mixing, the beads were again collected and washed. Illumina adapters (1 μl NEXTflex ChIP-seq barcoded adapters; BioO Scientific) were added to beads resuspended in 100 μL 1× Quick Ligation reaction buffer and 4 μL Quick T4 DNA Ligase (NEB), and incubated for 15 minutes at 24°C with gentle mixing. After washing the beads, the DNA was the eluted into a fresh tube by addition of 100 μL ChIP elution buffer (50 mM Tris–HCl, pH 7.5, 10 mM EDTA, 1% SDS) and incubation at 65°C for 10 minutes. The eluate was collected by centrifugation for one minute at 4000 rpm. Crosslinks were reversed by incubation for 10 minutes at 100°C. Samples were purified by phenol extraction and precipitated with ethanol, 40 μg

glycogen and 8.3 mM sodium acetate. DNA was pelleted for 15 minutes at 4°C at top speed in a microcentrifuge, washed with 70% ethanol, dried and resuspended in 11 μL H_2O. After quantification by PCR each library was amplified, purified and resuspended in 20 μL H_2O. Libraries were the sequenced using a HiSeq 2000 sequencer (Illumina; University at Buffalo Next Generation Sequencing Core Facility). Sequence reads were aligned to non-repetitive sequences in the E. coli H10407 genome using CLC Genomics Workbench and overall coverage was determined using custom Python scripts. Sequence reads have been submitted to the EBI ArrayExpress database and can be accessed using accession number E-MTAB-2917.

Bioinformatics

ChIP-seq peaks were identified as described previously [25]. We refer to these peaks as "high stringency" peaks. A second round of peak calling was performed in which the sequence read threshold values (i.e. the minimum number of sequence reads at a given genomic position that is required for a peak to be called) was reduced by 20%. We refer to these peaks as "low stringency" peaks. MEME [39] was used to identify enriched sequence motifs in the sequences from 50 bp upstream to 50 bp downstream of the high stringency peak centres. Thus, we identified a motif closely resembling the known CRP consensus site in many of the regions surrounding high stringency ChIP-seq peaks. These CRP site sequences are included in Table 1. Those high stringency peaks for which MEME did not identify a motif were used for a second round of analysis using MEME. This also identified a motif closely resembling the known CRP consensus site. These CRP site sequences are also included in Table 1. We used MEME to identify enriched sequence motifs in the low stringency peak list. This also identified a motif closely resembling the known CRP consensus site. These CRP site sequences are also included in Table 1. "High-confidence" ChIP-seq peaks listed in Table 1 include all the high stringency peaks but only those low stringency peaks for which we identified a motif using MEME. A complete list of all peaks, including low stringency peaks for which a motif was not identified by MEME, is provided in S3 Table. In order to assess the location of CRP sites with respect to TSSs we used the targets listed in Table 1. For each target the predicted sequence from MEME was used in a BLAST search against the E. coli K-12 MG1655 genome. All but 11 CRP sites in ETEC had a single perfect match in the E. coli K-12 chromosome. For each perfect match the distance from the centre of the CRP site to all transcription start sites was calculated. Transcription start site coordinates are from Kim et al. [40] and Cho et al. [41]. Distances between −200 and +100 were selected and all other distances were discarded. Distances were then grouped in bins of 5 bp each and the most common distance bins were identified. Note that, because the position of the CRP site was transposed onto the E. coli K-12 genome, the distance between CRP sites and TSSs

The PWM describing CRP binding sites was generated using the PREDetector software package and our previous list of 68 CRP binding sites in the E. coli K-12 genome [15,42]. Subsequent bioinformatic screens of plasmids p666 and p948 were done by importing the relevant genbank files into PREDetector and running a binding site search with a cut-off of 7 using settings that did not exclude CRP sites within genes. The "score" for each site predicted by PREDetector increases if a closer match to the PWM is found. To generate the chromosome and plasmid maps shown in Fig. 1 we used DNA plotter software [43].

Data shown in Fig. 9B–C were extracted from the publically available datasets of Kansal et al. [29] that measure changes in the

ETEC E24377A transcriptome upon contact with Caco-2 intestinal epithelial cells. The data are hosted under the GEO accession code GSE40427. For each assay condition (planktonic and attached ETEC cells) we extracted the signal intensity for microarray probe sets A1527 (*crp*), UTI189_C1433 (*hns*), D4754 (*eltA*) and D4048 (*estA*). The average signal intensity was calculated and the fold change in transcription in attached compared to planctonic ETEC cells was determined for each time point. The data in Fig. 9C show a comparison of absolute signal intensities for probe sets A1527 (*crp*) and D4048 (*estA*) compared for each of the two replicates obtained at 30, 60 or 120 minutes after attachment to host cells. Signal intensities obtained after 30 minutes growth in LB medium (three replicates) are also included in this analysis.

Proteins

The CRP and σ^{70} purification was done exactly as described previously [44,45]. RNA polymerase core enzyme was purchased from Epicenter. RNA polymerase holoenzyme was generated by incubating the core enzyme with an equimolar concentration of σ^{70} at room temperature for 20 minutes prior to use. H-NS was overexpressed in T7 express cells from plasmid pJ414hns. After overexpressing H-NS, cells were collected from the culture by centrifugation and resuspended in buffer A (20 mM Tris-HCl pH 7.2, 1 mM EDTA and 10% (v/v) glycerol) containing 100 mg/ml PMSF. Cells were lysed by sonication and the sample was cleared by centrifugation. The supernatent was loaded directly onto a Heparin column (Amersham) pre-equilibrated with buffer A. A linear NaCl gradient was applied and H-NS was found to elute at approximately 500 mM NaCl. The peak fractions were pooled and diluted 3-fold with buffer A. The sample was then loaded onto an S-FF column (Amersham) pre-equilibrated with Buffer A. A NaCl gradient was applied and H-NS eluted at approximately 550 mM NaCl. The H-NS containing fractions were then dialysed against a buffer containing 20 mM Tris HCl (pH 7.2), 300 mM KCl and 10% Glycerol (v/v)for storage at $-80°C$.

DNAse I footprinting and Electrophoretic Mobility Shift Assays

DNA fragments for DNAse I footprinting or EMSA assays were excised from pSR by sequential digestion with *Hin*dIII and then *Aat*II. After digestion, fragments were labelled at the *Hin*dIII end using [γ-^{32}P]-ATP and T4 polynucleotide kinase. DNAse I footprints and EMSA experiments were then done as described by Grainger *et al.* [45] except that cAMP was added to reactions at a concentration of 0.2 mM. Radio-labelled DNA fragments were used at a final concentration of ~10 nM. Note that all *in vitro* DNA binding reactions contained a vast excess (12.5 μg ml^{-1}) of Herring sperm DNA as a non-specific competitor. Footprints were analysed on a 6% DNA sequencing gel (molecular dynamics). The results of all footprints and EMSA experiments were visualized using a Fuji phosphor screen and Bio-Rad Molecular Imager FX.

Primer extension assays

Transcript start sites were mapped by primer extension, as described in Lloyd *et al.* [46] using RNA purified from strains carrying the 92 bp PestA1 or 93 bp PestA2 fragment cloned in pRW50. The 5′ end-labelled primer D49724, which anneals downstream of the *Hin*dIII site in pRW50, was used in all experiments. Primer extension products were analysed on denaturing 6% polyacrylamide gels, calibrated with size standards,

and visualized using a Fuji phosphor screen and Bio-Rad Molecular Imager FX.

In vitro transcription assays

The *in vitro* transcription experiments were performed as described previously Savery *et al.* [35] using the system of Kolb *et al.* [38]. A Qiagen maxiprep kit was used to purify supercoiled pSR plasmid carrying the different promoter inserts. This template (~16 μg ml−1) was pre-incubated with purified CRP in buffer containing 0.2 mM cAMP, 20 mM Tris pH 7.9, 5 mM MgCl$_2$, 500 μM DTT, 50 mM KCl, 100 μg ml^{-1} BSA, 200 μM ATP, 200 μM GTP, 200 μM CTP, 10 μM UTP with 5 μCi [α-32P]-UTP. The reaction was started by adding purified *E. coli* RNA polymerase. Labelled RNA products were analysed on a denaturing polyacrylamide gel.

β-galactosidase assays and addition of glucose and salt to growth medium

β-Galactosidase assays were done using the protocol of Miller [47]. All assay values are the mean of three independent experiments with a standard deviation <10% of the mean. Cells were grown aerobically at 37°C to mid-log phase in LB medium unless stated otherwise. For all experiments investigating the effects of glucose and salt M9 minimal medium was used so that the glucose and salt concentrations could be controlled more accurately. The amount of glucose is shown as percentage *w/v*. The addition of "salt" refers to a 3:1 molar ration of NaCl to KCl. We have arbitrarily described 30 mM NaCl and 10 mM KCl as being a "1%" salt solution.

Intranasal mouse infection model assays

Strains of ETEC were grown in Luria Broth (LB) to an OD$_{600}$ of 1.0. Groups of 10 mice (8–10 week old BALB/c) were infected intranasally with approximately 1×10^9 colony forming units of bacteria in 100 μl of inoculums according to Byrd *et al.* [30]. Mice were monitored daily for 6 days post-infection for weight and morbidity.

Ethics statement

The protocol 12-02-015IBT "Oral Immunization of Mice with Enterotoxigenic: *E coli* (ETEC)" has been approved by the Noble Life Sciences IACUC committee. All animal care and use procedures adhere to the guidelines set by the Public Health Service Policy, U.S. Dept. of Agriculture (USDA) and the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health.

Supporting Information

S1 Fig Binding of CRP to predicted targets *in vitro*. A) The data show binding of CRP to a target from each of the bins shown in Fig. 2. B) CRP binding to remaining targets scoring >10. CRP was used at concentrations of 0, 175, 350 or 700 nM. The "score" describes how well the predicted target matches the PWM.
(PDF)

S2 Fig Promoter DNA fragments used in this work. A) ETEC H10407 PestA2 containing DNA fragments. B) ETEC H10407 derived DNA fragments containing PestA1. C) ETEC H10407 PeltAB containing sequences. D) DNA fragments containing sequences upstream of the *estA2* and *eltAB* genes of ETEC E24377A.
(PDF)

S3 Fig A) Mutations in P*estA2* UP-element alter the migration of the promoter DNA on an agarose gel. The DNA sequences used are shown in part (i) and the mobility of the fragments, on an agarose gel, are shown in part (ii). Note that each sample has been loaded in duplicate. B) Mutating the UP-element renders P*estA2* uninducible by CRP. Part (i) shows LacZ activity data for the different promoter fragments cloned in pRW50. The pRW50 derivatives were used to transform M182 or the Δ*crp* derivative. Part (ii) shows the result of in vitro transcription assays using the different promoter fragments, cloned in pSR, as a template.
(PDF)

S4 Fig Activity of different promoter::*lacZ* fusions in the presence of increasing glucose and salt concentrations. The figure shows β-galactosidase activity measurements for lysates obtained from cultures of M182 carrying the A) *estA1* B) *estA2* or C) *eltAB* promoters cloned in pRW50. Panel D) shows β-galactosidase activity values for lysates of M182 and M182Δ*hns* cells, carrying the *estA2* promoter, or a derivative lacking the CRP site, cloned in pRW50. Cells were grown in the presence or absence of 2% glucose. Assays were done in M9 minimal medium so that the glucose and salt concentrations could be more accurately controlled.
(PDF)

S1 Table Putative CRP binding targets on ETEC plasmids p948 and p666 identified by PREDetector.
(DOCX)

S2 Table Strains, plasmids and oligonucleotides.
(DOCX)

S3 Table All CRP binding sites on the ETEC H10407 chromosome identified by ChIP-seq.
(DOCX)

References

1. Crossman LC, Chaudhuri RR, Beatson SA, Wells TJ, Desvaux M, et al. (2010) A commensal gone bad: complete genome sequence of the prototypical enterotoxigenic *Escherichia coli* strain H10407. J Bacteriol 192:5822–5831.
2. Sack RB (2011) The discovery of cholera - like enterotoxins produced by *Escherichia coli* causing secretory diarrhoea in humans. Indian J Med Res 133: 171–80
3. Gupta SK, Keck J, Ram PK, Crump JA, Miller MA, et al. (2008) Analysis of Data Gaps Pertaining to Enterotoxigenic Escherichia coli Infections in Low and Medium Human Development Index Countries, 1984–2005. Epidemiology and Infection 136:721–738.
4. de Haan L, Hirst TR (2004) Cholera toxin: a paradigm for multi-functional engagement of cellular mechanisms. Mol Membr Biol 21:77–92.
5. Zhang RG, Scott DL, Westbrook ML, Nance S, Spangler BD, et al. (1995) The three-dimensional crystal structure of cholera toxin. J Mol Biol 251:563–73.
6. Taxt A, Aasland R, Sommerfelt H, Nataro J, Puntervoll P (2010) Heat-stable enterotoxin of enterotoxigenic *Escherichia coli* as a vaccine target. Infect Immun 78:1824–31.
7. Yamamoto T, Tamura T, Yokota T (1984) Primary structure of heat-labile enterotoxin produced by Escherichia coli pathogenic for humans. J Biol Chem 259:5037–44.
8. de Haan L, Verweij WR, Feil IK, Holtrop M, Hol WG, et al. (1998) Role of GM1 binding in the mucosal immunogenicity and adjuvant activity of the *Escherichia coli* heat-labile enterotoxin and its B subunit. Immunology 94:424–430.
9. Saslowsky DE, te Welscher YM, Chinnapen DJ, Wagner JS, Wan J, et al. (2013) Ganglioside GM1-mediated transcytosis of cholera toxin bypasses the retrograde pathway and depends on the structure of the ceramide domain. J Biol Chem 288:25804–9.
10. Nalin DR, Cash RA, Islam R, Molla M, Phillips RA (1968) Oral maintenance therapy for cholera in adults. Lancet 2:370–3.
11. Guerrant RL, Carneiro-Filho BA, Dillingham RA (2003) Cholera, diarrhea, and oral rehydration therapy: triumph and indictment. Clin Infect Dis 37:398–405.
12. Busby S, Ebright RH (1999) Transcription activation by catabolite activator protein (CAP). J Mol Biol 293:199–213.
13. Parkinson G, Wilson C, Gunasekera A, Ebright YW, Ebright RH, et al. (1996) Structure of the CAP-DNA complex at 2.5 angstroms resolution: a complete picture of the protein-DNA interface. J Mol Biol 260:395–408.
14. Zheng D, Constantinidou C, Hobman JL, Minchin SD (2004) Identification of the CRP regulon using in vitro and in vivo transcriptional profiling. Nucleic Acids Res 32:5874–93.
15. Grainger DC, Hurd D, Harrison M, Holdstock J, Busby SJ (2005) Studies of the distribution of *Escherichia coli* cAMP-receptor protein and RNA polymerase along the *E. coli* chromosome. Proc Natl Acad Sci U S A 102:17693–8.
16. Rossiter AE, Browning DF, Leyton DL, Johnson MD, Godfrey RE, et al. (2011) Transcription of the plasmid-encoded toxin gene from enteroaggregative *Escherichia coli* is regulated by a novel co-activation mechanism involving CRP and Fis. Mol Microbiol 81:179–91.
17. Navarre WW, Porwollik S, Wang Y, McClelland M, Rosen H, et al. (2006) Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. Science 313:236–8.
18. Liu Y, Chen H, Kenney LJ, Yan J (2010) A divalent switch drives H-NS/DNA-binding conformations between stiffening and bridging modes. Genes Dev 24:339–44.
19. Lim CJ, Lee SY, Kenney LJ, Yan J (2012) Nucleoprotein filament formation is the structural basis for bacterial protein H-NS gene silencing. Sci Rep 2:509.
20. Myers KS, Yan H, Ong IM, Chung D, Liang K, et al. (2013) Genome-scale analysis of *Escherichia coli* FNR reveals complex features of transcription factor binding. PLoS Genet 9(6):e1003565.
21. Dame RT, Wyman C, Wurm R, Wagner R, Goosen N (2002) Structural basis for H-NS-mediated trapping of RNA polymerase in the open initiation complex at the *rrnB* P1. J Biol Chem 277:2146–50.
22. Dorman CJ (2007) H-NS, the genome sentinel. Nat Rev Microbiol 5: 157–61.
23. Atlung T, Ingmer H (1997) H-NS: a modulator of environmentally regulated gene expression. Mol Microbiol 24: 7–17.
24. Steinsland H1, Valentiner-Branth P, Perch M, Dias F, Fischer TK, Aaby P, Mølbak K, Sommerfelt H (2002) Enterotoxigenic *Escherichia coli* infections and diarrhea in a cohort of young children in Guinea-Bissau. J Infect Dis 186: 1740–7.
25. Singh SS, Singh N, Bonocora RP, Fitzgerald DM, Wade JT, et al. (2014) Widespread suppression of intragenic transcription initiation by H-NS. Genes Dev 28: 214–219.
26. Yang J, Tauschek M, Strugnell R, Robins-Browne RM (2005) The H-NS protein represses transcription of the *eltAB* operon, which encodes heat-labile enterotoxin in enterotoxigenic *Escherichia coli*, by binding to regions downstream of the promoter. Microbiology 151: 1199–1208.
27. Bodero MD, Munson GP (2009) Cyclic AMP receptor protein-dependent repression of heat-labile enterotoxin. Infect Immun 77:791–8.
28. Sahl JW, Rasko DA (2012) Analysis of global transcriptional profiles of enterotoxigenic Escherichia coli isolate E24377A. Infect Immun 80: 1232–1242.
29. Kansal R, Rasko DA, Sahl JW, Munson GP, Roy K (2013) Transcriptional modulation of enterotoxigenic *Escherichia coli* virulence genes in response to epithelial cell interactions. Infect Immun 81: 259–70.
30. Byrd W, Mog SR, Cassels FJ (2003) Pathogenicity and immune response measured in mice following intranasal challenge with enterotoxigenic *Escherichia coli* strains H10407 and B7A. Infect Immun 71: 13–21.
31. Byrd W, Boedeker EC (2013) Attenuated *Escherichia coli* strains expressing the colonization factor antigen I (CFA/I) and a detoxified heat-labile enterotoxin (LThK63) enhance clearance of ETEC from the lungs of mice and protect mice from intestinal ETEC colonization and LT-induced fluid accumulation. Vet Immunol Immunopathol 152: 57–67.
32. Gupta S, Chowdhury R (1997) Bile affects production of virulence factors and motility of *Vibrio cholerae*. Infect Immun 65: 1131–1134.
33. Stringer AM, Singh N, Yermakova A, Petrone BL, Amarasinghe JJ, et al. (2012) FRUIT, a scar-free system for targeted chromosomal mutagenesis, epitope tagging, and promoter replacement in *Escherichia coli* and *Salmonella enterica*. PLoS One 7: e44841.

34. Page L, Griffiths L, Cole JA (1990) Different physiological roles of two independent pathways for nitrite reduction to ammonia by enteric bacteria. Arch Microbiol 154:349–54.

35. Busby S, Kotlarz D, Buc H (1983) Deletion mutagenesis of the Escherichia coli galactose operon promoter region. J Mol Biol 167:259–274.

36. Lodge J, Fear J, Busby S, Gunasekaran P, Kamini NR (1992) Broad host range plasmids carrying the Escherichia coli lactose and galactose operons. FEMS Microbiol Lett 74:271–6.

37. Kolb A, Kotlarz D, Kusano S, Ishihama A (1995) Selectivity of the Escherichia coli RNA polymerase E sigma 38 for overlapping promoters and ability to support CRP activation. Nucleic Acids Res 23:819–26.

38. Singh SS, Grainger DC (2013) H-NS can facilitate specific DNA-binding by RNA polymerase in AT-rich gene regulatory regions. PLoS Genet 9: e1003589.

39. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, et al. (2009) MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 37(Web Server issue):W202–8.

40. Kim D, Hong JS, Qiu Y, Nagarajan H, Seo JH, et al. (2012) Comparative analysis of regulatory elements between Escherichia coli and Klebsiella pneumoniae by genome-wide transcription start site profiling. PLoS Genet 8(8):e1002867.

41. Cho BK, Kim D, Knight EM, Zengler K, Palsson BØ (2014) Genome-scale reconstruction of the sigma factor network in Escherichia coli: topology and functional states. BMC Biol 12:4.

42. Hiard S, Maree R, Colson S, Hoskisson PA, Titgemeyer F, et al. (2007) PREDetector: A new tool to identify regulatory elements in bacterial genomes. Biochem Biophys Res Commun 357(4):861–4.

43. Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J (2009) DNAPlotter: circular and linear interactive genome visualization. Bioinformatics 25: 119–20.

44. Savery NJ, Lloyd GS, Kainz M, Gaal T, Ross W, et al. (1998) Transcription activation at Class II CRP-dependent promoters: identification of determinants in the C-terminal domain of the RNA polymerase alpha subunit. EMBO J 17:3439–3447.

45. Grainger DC, Goldberg MD, Lee DJ, Busby SJ (2008) Selective repression by Fis and H-NS at the Escherichia coli dps promoter. Mol Microbiol 68:1366–1377.

46. Lloyd GS, Hollands K, Godfrey RE, Busby SJ (2008) Transcription initiation in the Escherichia coli K-12 malI-malX intergenic region and the role of the cyclic AMP receptor protein. FEMS Microbiol Lett 288: 250–7.

47. Miller J (1972) Experiments in Molecular Genetics. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.