

RESEARCH ARTICLE

Unusually Situated Binding Sites for Bacterial Transcription Factors Can Have Hidden Functionality

James R. J. Haycocks, David C. Grainger*

Institute of Microbiology and Infection, School of Biosciences, College of Life and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham, United Kingdom

* d.grainger@bham.ac.uk



Abstract

A commonly accepted paradigm of molecular biology is that transcription factors control gene expression by binding sites at the 5' end of a gene. However, there is growing evidence that transcription factor targets can occur within genes or between convergent genes. In this work, we have investigated one such target for the cyclic AMP receptor protein (CRP) of enterotoxigenic *Escherichia coli*. We show that CRP binds between two convergent genes. When bound, CRP regulates transcription of a small open reading frame, which we term *aatS*, embedded within one of the adjacent genes. Our work demonstrates that non-canonical sites of transcription factor binding can have hidden functionality.

OPEN ACCESS

Citation: Haycocks JRJ, Grainger DC (2016) Unusually Situated Binding Sites for Bacterial Transcription Factors Can Have Hidden Functionality. PLoS ONE 11(6): e0157016. doi:10.1371/journal.pone.0157016

Editor: Akira Ishihama, Hosei University, JAPAN

Received: April 13, 2016

Accepted: May 23, 2016

Published: June 3, 2016

Copyright: © 2016 Haycocks, Grainger. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by the Wellcome Trust Award Number WT085092MA to DCG and BBSRC Award Number BB/N005961/1 to DCG. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

A long standing view is that promoters, and transcription factor binding sites, should locate to the 5' end of an annotated gene [1]. Unexpectedly, recent genome-wide studies of RNA polymerase distribution, and transcription factor binding, have shown that the situation is more complex [2]. Notably, transcription factor binding sites are frequently found within genes or between convergent genes. For example, RutR (a regulator of pyrimidine catabolism) binds predominantly at sites within genes [3]. Similarly, LeuO (a regulator of leucine biosynthesis) binds to numerous targets between convergent genes [4]. A major challenge is to understand if, and how, such transcription factor binding sites contribute to gene regulation.

In this work we have focused on the the cAMP receptor protein (CRP). Widely conserved in bacteria, CRP is global regulator of transcription that responds to cAMP levels [1]. When bound to DNA, CRP regulates transcription by one of two distinct mechanisms. For example, at class I promoters, CRP binds upstream of the promoter -35 element and interacts with the C-terminal domain of the RNA polymerase α -subunit [5–7]. Conversely, at class II promoters, CRP binds close to the promoter -35 element and interacts with the N-terminal domain of the RNA polymerase α -subunit [8, 9]. In both cases, CRP stimulates transcription [5–9]. Recently, we have shown that CRP binds numerous sites that are not close to the 5' end of an annotated gene [10, 11]. In this work we have focused on one such atypical target; a CRP site between two

convergent genes. Our data show that CRP activates expression of a small open reading frame (ORF) completely embedded within one of these convergent genes. Activation occurs by a standard class II CRP-dependent mechanism. Thus, although the genomic context of the CRP binding site is unusual, gene regulation proceeds via a well-defined mechanism.

Methods and Materials

Strains, plasmids and oligonucleotides

Strains and plasmids used are listed in [Table 1](#). Oligonucleotides used are listed in [Table 2](#).

β-galactosidase assays

β-galactosidase assays were done according to the method of Miller [12]. Cells were grown in M9 minimal media supplemented with 1% fructose to stationary phase as specified in figure legends. Since *PaatS* was 2-fold more active in stationary phase cultures compared to log phase cultures the cells from the former phase of growth were used in all experiments. Values shown are the mean of three independent experiments. Error bars represent the standard deviation of three independent experiments. To calculate *P* we used a paired, two-tailed, student's *t* test.

Table 1. Strains and plasmids used in this study.

Strain or plasmid	Description	Source
Strains		
<i>Escherichia coli</i> M182	<i>Δlac galK galU strA</i>	[39]
<i>Escherichia coli</i> M182 Δ <i>crp</i>		[39]
Plasmids		
pRW50	Broad-host-range <i>lac</i> fusion vector for cloning promoters on <i>EcoRI</i> – <i>HindIII</i> fragments: contains the RK2 origin of replication and encodes TcR.	[40]
pSR	pBR322-derived plasmid containing an <i>EcoRI</i> – <i>HindIII</i> fragment upstream of the <i>loop</i> transcription terminator	[17]
pRW225	pRW50-derived plasmid in which the ribosome binding site upstream of <i>lacZ</i> has been deleted.	[23]
pDCRP	<i>crp</i> gene preceding its native promoter (located on <i>EcoRI</i> – <i>Sall</i> flanked fragment). pBR322 derived. Encodes AmpR. ColE1 origin.	[41]

doi:10.1371/journal.pone.0157016.t001

Table 2. Oligonucleotides used in this study.

Oligonucleotide name	Sequence (5'-3')	Source
Oligonucleotides used to generate <i>PaatS</i> derivatives		
<i>aatS1</i> For	GGCTGCGAATTCATAAAGTGATAAAAAATCACATAAAATTTTATTAAAAGGATATAACCTTCATATCACTTGTAATTAATTTGTGTCA	This study
<i>aatS1</i> Rev	GCCCGAAGCTTCATGGATATACTTCTTAAGTATTATAAACAAGGTGGAACAGTTGTTATGGTAACCCATGACACAAATTTAATTACAA	This study
<i>aatS2</i> Rev	GCCCGAAGCTTCATGGATATTGAAGAAAAGTATTATAAACAAGGTGGAACAGTTGTTATGGTAACCCATGACACAAATTTAATTACAA	This study
Oligonucleotides used for primer extension experiments		
D49724	GGTTGGACGCCCGGCATAGTTTTTCAGCAGGTCGTTG	[13]

doi:10.1371/journal.pone.0157016.t002

Primer extension

Primer extension was done as described by Lloyd *et al.* [13]. Briefly, RNA was extracted from M182 cells (or the Δcrp derivative) harbouring pRW50 containing the *aatS1* promoter fragment. Extractions were done using a Qiagen RNeasy mini kit according to the manufacturer's instructions and residual DNA was removed using a Turbo DNA-free kit (Ambion). RNA integrity was then checked by visualisation following agarose gel electrophoresis. The ratio of absorbance at 260 nm and 280 nm was used to assess the purity of the RNA. Primer extension products were analysed on a 6% denaturing polyacrylamide gel alongside size standards generated by manual sequencing of M13 phage DNA. Full gel images are provided in the supporting information (S1 Fig).

Proteins

CRP and σ^{70} proteins were purified as previously described [14, 15]. RNA polymerase core enzyme was purchased from Cambio, and was incubated at 37°C with a ten-fold molar excess of σ^{70} before use to generate RNA polymerase holoenzyme.

DNase I footprinting

DNA templates were generated by excision of *AatII-HindIII* fragments from maxipreps of pSR plasmid containing the 134 bp *aatS1* promoter fragment. The resulting 212 bp *AatII-HindIII* fragment was labelled at the *HindIII* end using γ -dATP and T4 polynucleotide kinase. Footprinting reactions were done as previously described [15] in buffer containing 120 mM KCl, 100 μ M EDTA, 20 mM Tris pH 7, 10 mM $MgCl_2$ and \sim 10 nM of *PaatS* DNA. Reactions also containing 12.5 μ g ml⁻¹ Herring Sperm DNA as a competitive inhibitor. Footprints were analysed on denaturing 6% polyacrylamide gels alongside calibrating Maxam-Gilbert G/A ladders [16]. Full gel images are provided in the supporting information (S1 Fig).

Multi-round *in vitro* transcription assays

In vitro transcription reactions were carried out as described by Savery *et al.*, [14] using the method of Kolb *et al.* [17]. Supercoiled DNA was extracted from M182 cultures harbouring pSR containing the *aatS1* fragment. Reaction buffer contained 20 mM Tris pH 7.9, 200 mM GTP/ATP/CTP, 10 mM UTP, 5 μ Ci (α -³²P) UTP, 500 mM DTT, 5 mM $MgCl_2$, 100 μ g ml⁻¹ BSA and 0.2 mM cAMP. Template DNA (at a final concentration of 16 μ g ml⁻¹) was incubated with CRP at 37°C for 5 minutes in reaction buffer, prior to the addition of RNA polymerase holoenzyme to start the reaction. Full gel images are provided in the supporting information (S1 Fig).

Bioinformatics

NCBI ORF finder was used to search for putative genes [18]. NCBI BLAST was used to search for *aatS* homologues. A list of DUF1602-containing proteins was obtained using the InterPro tool [19], the list is correct as of 14/01/16.

Results

Identification of promoter-like DNA sequence elements between two convergent genes

The starting point for this work was our previous analysis of global CRP binding in enterotoxigenic *Escherichia coli* (ETEC) [11]. This work predicted a CRP target between the convergent

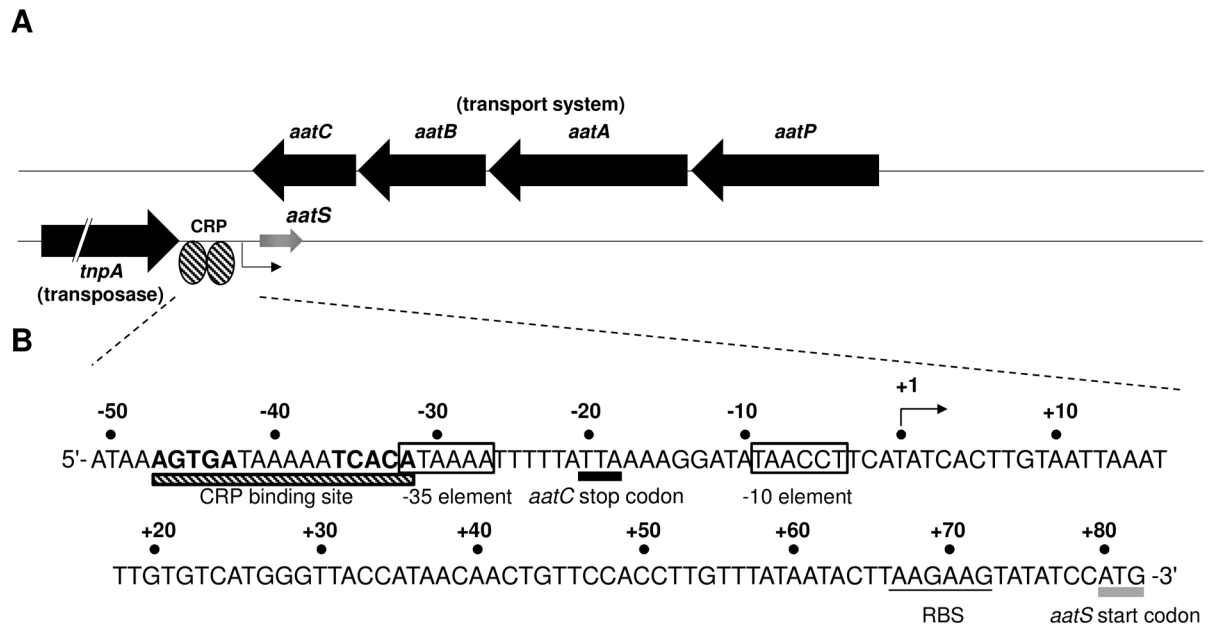


Fig 1. The *aatPABC* operon of ETEC H10407. Schematic of the *aatPABC* operon and adjacent *tnpA* gene. The two DNA strands are shown as black lines. Known genes are shown as black arrows and the predicted *aatS* gene as a grey arrow. Gene names are shown in italic and gene function in parenthesis. The position of a putative CRP binding site is indicated by striped ovals. **A.** Sequence of the *tnpA*-*aatC* intergenic region. The CRP site is highlighted as a striped rectangle with the two half sites highlighted in bold. The start codon of the *aatS* open reading frame is highlighted with a grey rectangle. The transcription start site, as determined by mRNA primer extension is denoted "+1" and indicated by a bent arrow. Distances upstream (-) and downstream (+) of this start site are numbered. The -35 and -10 hexamers are boxed, and the ribosome binding site (RBS) is underlined.

doi:10.1371/journal.pone.0157016.g001

genes *aatC* (encoding a subunit of a transport system) and *tnpA* (encoding a transposase). The genomic context of the CRP site is shown in Fig 1A and the surrounding DNA sequence is shown in Fig 1B. We searched the locus for promoters that may be under the control of CRP. A sequence (5'-TAACCT-3') that resembled a promoter -10 hexamer was found close to the CRP site within the *aatC* coding region (Fig 1B). A 189 bp ORF, which we named *aatS*, and a ribosome binding site (RBS) were also identified (Fig 1B). Note that *aatS* fully overlaps *aatC* but is in the opposite orientation. We considered that the overlap might result from *aatC* being annotated incorrectly. However, on close examination, this seems unlikely; there is no alternative stop codon within the *aatC* open reading frame. Furthermore, the full *aatC* sequence is conserved in numerous bacteria.

PaatS is a CRP-dependent promoter *in vivo*

Our next aim was to determine if the putative promoter upstream of *aatS* (*PaatS*) was functional. Thus, a 134 bp DNA fragment containing the sequence shown in Fig 1B, flanked by *EcoRI* and *HindIII* restriction sites, was generated. The DNA fragment, named *aatS1*, was cloned upstream of *lacZ* in plasmid pRW50 to create a *PaatS::lacZ* fusion. Next, *E. coli* M182 and the Δ *crp* derivative were transformed with the resulting plasmid. Transformants were cultured in liquid media and RNA was isolated as a template for primer extension. A 166 nucleotide primer extension product was observed using RNA from M182. Conversely, little extension product was observed using RNA from the Δ *crp* derivative (compare lanes 5 and 6 in Fig 2A). Thus, the *aatS* transcription start site (marked "+1" in Fig 1B) is located 4 bp downstream of the putative *aatS* promoter -10 hexamer. Furthermore, *PaatS* is poorly active in the

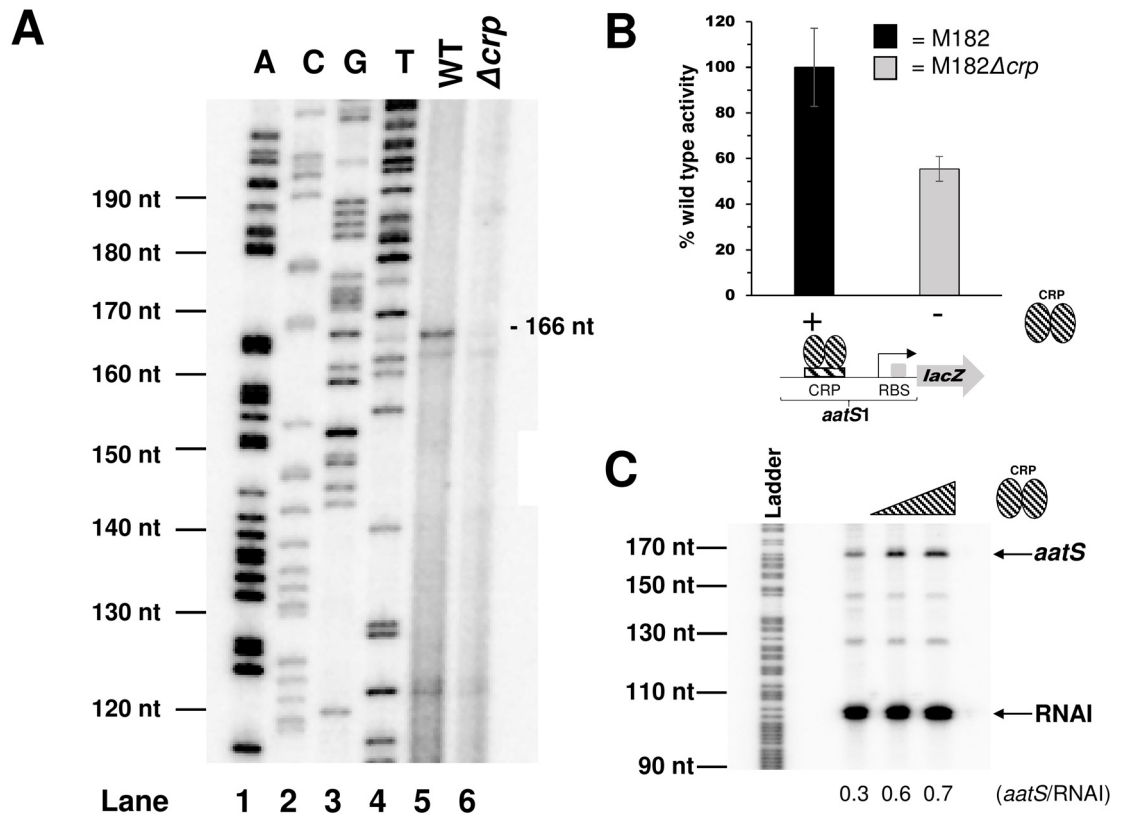


Fig 2. Characterisation of the *PaatS* promoter. **A.** Primer extension analysis of the *aatS* transcript. Lanes 1–4 on the gel are arbitrary size standards, used for calibration, generated by sequencing of M13mp18 phage DNA. Lane 5 shows the primer extension product generated using RNA from wildtype M182 cells carrying the *aatS1::lacZ* fusion. Lane 6 shows the primer extension product generated using RNA from M182 Δcrp cells carrying the *aatS1::lacZ* fusion. The transcription start site is indicated in Fig 1B. **B.** β -galactosidase activity determined using lysates of M182 wildtype or M182 Δcrp cells carrying *PaatS* cloned upstream of *lacZ* in plasmid pRW50. Values shown are percentages of activity observed in strain M182 (92 Miller units). We obtained 7 and 3 Miller units of activity from lysates of M182 or M182 Δcrp carrying promoterless pRW50. Error bars represent the standard deviation of three independent experiments. **C.** Multi-round *in vitro* transcription assay using *PaatS*. The *aatS1* DNA fragment was cloned into pSR upstream of a λ loop terminator. Purified, supercoiled pSR plasmid was incubated with purified CRP at 37°C, and the reaction started by the addition of 400 nM σ^{70} -RNA polymerase holoenzyme. CRP concentrations are; 0 nM, 200 nM, or 400 nM. The 108 nt RNAI transcript from the pSR replication origin, and the 169 nt transcript from *PaatS*, are indicated. The gel is calibrated with an arbitrary G+A DNA sequencing reaction as a size standard.

doi:10.1371/journal.pone.0157016.g002

absence of CRP. Consistent with our primer extension analysis, *PaatS* controlled expression of *lacZ* was also significantly reduced in cells lacking CRP ($P = 0.018$; Fig 2B).

PaatS is a CRP dependent promoter *in vitro*

We next confirmed activation of *PaatS* by CRP using *in vitro* transcription assays. The *aatS1* DNA fragment was cloned upstream of the λ loop transcription terminator in plasmid pSR. Thus, the transcript generated from *PaatS* is expected to be 169 nucleotides (nt) in length. Note that, pSR also encodes RNAI; a 108 nt factor independent transcript derived from the plasmid replication origin. Thus, we observed two bands on denaturing PAGE gels (Fig 2C). As expected, synthesis of RNAI did not require CRP. However, production of the larger 169 nt *aatS* transcript was stimulated by CRP.

CRP binds a site overlapping the *PaatS* -35 element

To confirm CRP binding to the predicted target upstream of *PaatS* we used DNase I footprinting. The result of the experiment is shown in Fig 3. Lane 1 shows the banding pattern resulting from DNase I cleavage of *aatS1* in the absence of CRP. Alterations to this pattern are evident in lanes 2–6 as increasing concentrations of CRP are added. Thus, CRP induced the appearance of three hypersensitive bands (starred) and protected the flanking DNA regions (underlined). The footprint aligns precisely with the predicted DNA target for CRP. The CRP site is 39.5 bases upstream of the *PaatS* transcription start site (Fig 1B). However, we note that the *aatS* mRNA

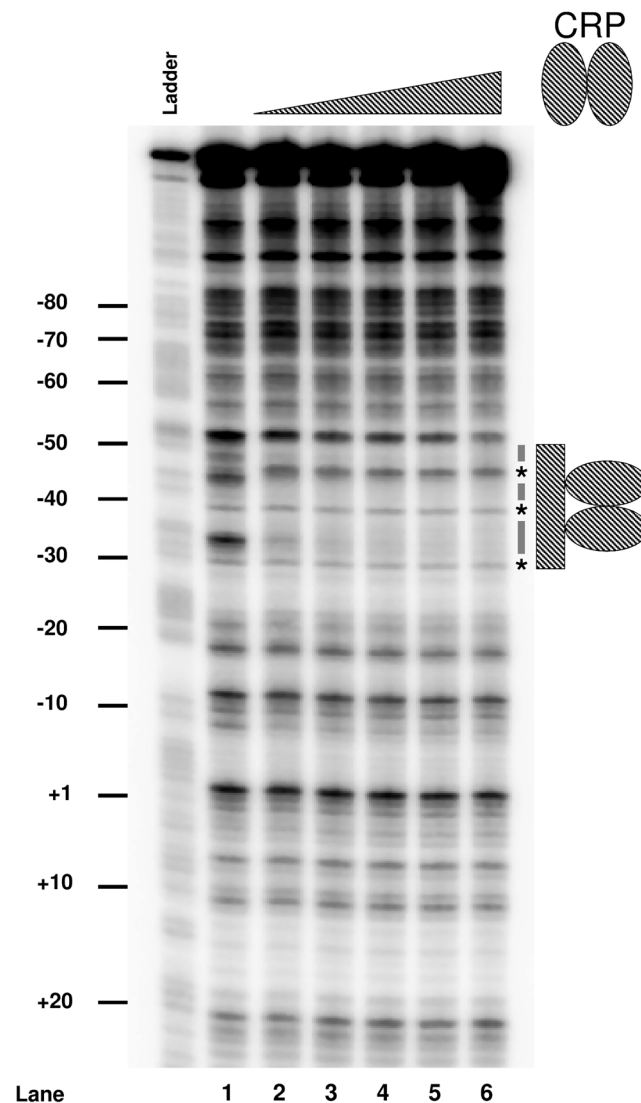


Fig 3. Binding of CRP to the *PaatS* region. DNase I footprint analysis of the *PaatS* region. The lane labelled 'G+A' is a Maxim/Gilbert G+A sequencing reaction. Lane 1 shows the cleavage pattern obtained from *aatS1* DNA digested with DNase I in the absence of CRP. Lanes 2–6 show DNase I cleavage patterns generated in the presence of increasing concentrations of CRP (0.35 μ M, 0.7 μ M, 1.05 μ M, 1.4 μ M and 2.1 μ M). The predicted CRP site is indicated by a hashed grey bar. DNA protection is indicated by black dashes, hypersensitive bends are highlighted by stars.

doi:10.1371/journal.pone.0157016.g003

start is 4 bp, rather than the usual 7 bp, downstream of the promoter -10 element. Hence, CRP is well positioned to interact with RNA polymerase via a standard class II mechanism.

The *aatS* coding region is preceded by a functional ribosome binding site

The predicted *aatS* ORF is located 82 bp downstream of the *aatS* transcription start site, within the coding region of *aatC*. We noticed that the 5' end of the *aatS* mRNA contains a sequence, 5' -UUAAGAAGU-3', that resembles a RBS (5' -UAAGGAGGU-3') [20, 21]. In addition, the predicted *aatS* RBS is located 6 bp upstream of the *aatS* start codon, a position suitable for translation initiation [22]. To determine if the RBS was functional, we created a translational *aatS1::lacZ* fusion and explored the effects of mutating the RBS on *lacZ* expression. Thus, we generated a derivative of the *aatS1* fragment, called *aatS2*, where the sequence of the RBS was altered to 5' -UUUUCUUCA-3' (Fig 4). Both *aatS1* and *aatS2* were translationally fused to *lacZ* by cloning into pRW225 [23]. Resulting plasmids were used to transform M182 and the Δcrp derivative. LacZ activities were then determined in lysates of transformants grown in

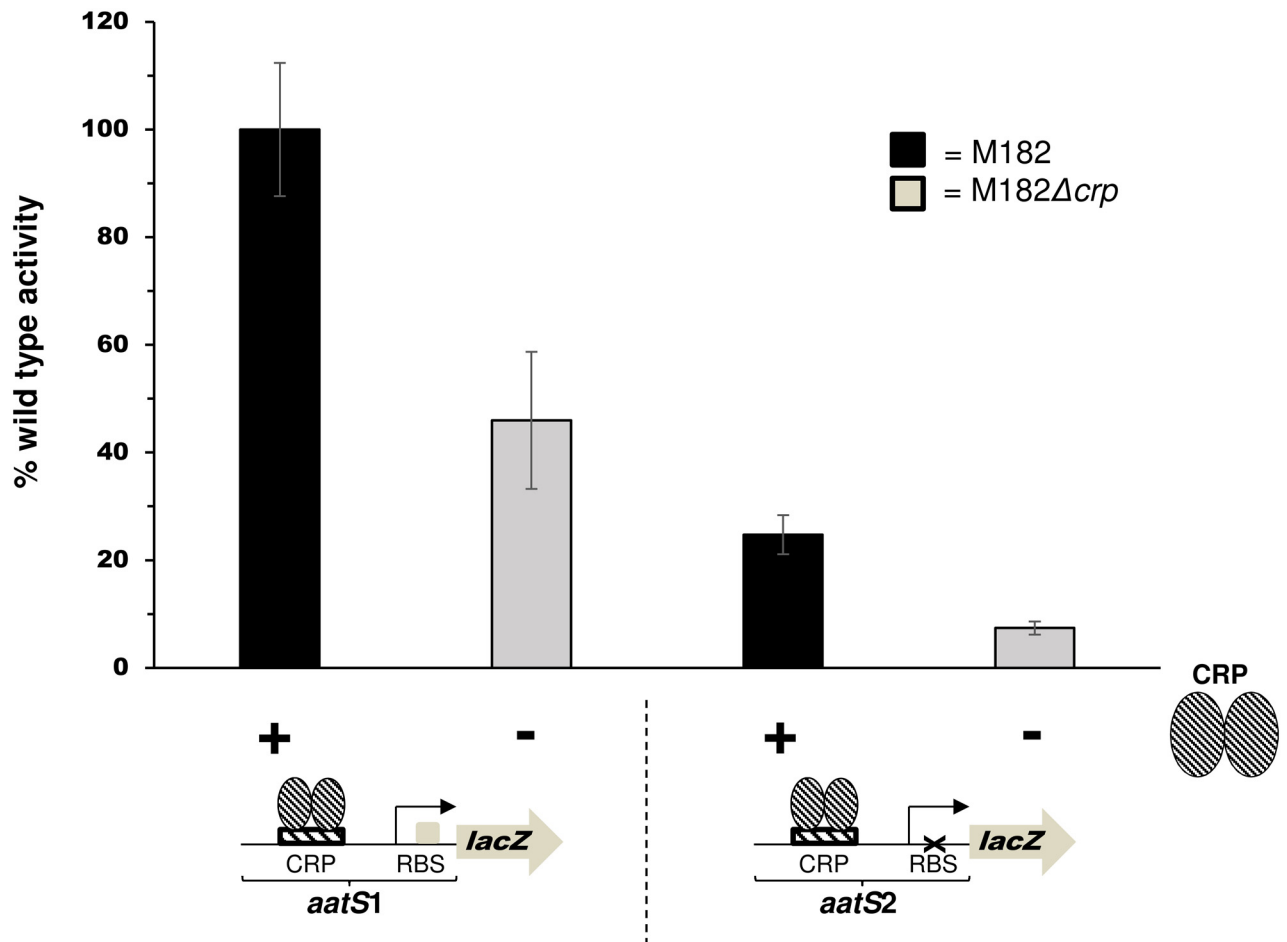


Fig 4. The *aatS* mRNA contains a functional ribosome binding site. The graph shows activity of different *aatS::lacZ* translational fusions. The wildtype ribosome binding site (5'-AAGAAG-3') in the *aatS1* fragment was mutated to (5'-TTCTTC-3') in *aatS2*. LacZ activities were determined using the lysates of stationary phase M182 or M182 Δcrp . In M182 cells *crp* was supplied in trans by plasmid pCRP that encodes *crp* under the control of its own promoter. Values shown are percentages of activity observed in strain M182 (5 Miller units). We obtained 0.25 and 0.26 Miller units of activity from lysates of M182 or M182 Δcrp , carrying promoterless pRW225, respectively. Error bars represent the standard deviation of three independent experiments.

doi:10.1371/journal.pone.0157016.g004

liquid culture (Fig 4). LacZ expression was significantly reduced when the *aatS* RBS was mutated ($P = 0.011$). Residual LacZ expression, driven by the *aatS2::lacZ* fusion in M182Δ*crp* cells, was not significantly different from background LacZ activity ($P = 0.096$). Background LacZ activity was obtained using M182Δ*crp* transformed with pRW225 having no promoter insert.

AatS contains a conserved domain of unknown function

We cannot exclude the possibility that *PaatS* drives transcription of a small or antisense RNA. However, the presence of a functional RBS, appropriately positioned upstream of the *aatS* open reading frame, is consistent with the transcript being an mRNA. If this is the case, translation would result in the production of a 62 amino acid AatS peptide. To better understand AatS we used BLAST to identify 38 microbial proteins with a significantly similar sequence (E-value <10) in the NCBI blastp database. The search also reveals that AatS shares 67% identity with a conserved domain of unknown function (DUF1602, E-value: $4.07e^{-7}$). Genes encoding DUF1602 are found in diverse species in all kingdoms of life (Table 3). Strikingly, as in the case of *aatS*, genes encoding DUF1602 are often genetically associated with, and occasionally overlap, genes encoding transport systems. We conclude that *aatS* may encode a small protein that could be an ancillary subunit of the Aat transport system in ETEC.

Discussion

In this study we have characterised the function of a predicted CRP binding site located between two convergent genes. We show that CRP activates transcription of *aatS*; a small ORF embedded within the much larger *aatC* gene. The *aatPABC* operon encodes a type I secretion system (the Aat system) found in many pathogenic bacteria [24, 25]. The presence of a functional RBS, correctly positioned upstream of *aatS*, suggests that the gene encodes a small protein rather than a regulatory RNA. Interestingly, many small proteins in bacteria localise to the membrane and function as accessory factors in transport systems [26]. Consistent with such a function, *aatS* is genetically associated with genes encoding transport systems in many bacteria. Furthermore, a potential transmembrane helix is predicted between residues 4 and 21 of AatS [27].

Documented instances of overlapping, protein-encoding, genes in bacteria are rare [28]. We are aware of only two examples; *rpmH* is encoded within *rnpA* in *Thermus thermophilus* and *setAB* resides inside the *pic* gene of *E. coli* 042 and *Shigella flexneri* [29, 30]. However, we speculate that further overlapping transcription units may become evident as unusual transcription factor targets are examined in detail. For instance, in a study of 154 *Mycobacterium tuberculosis* transcription factors, 75% of binding targets were located inside genes [31]. Similarly, a study of 116 transcription factors in *E. coli*, identified many intragenic binding events [32]. Presumably, some of these targets will control production of unannotated transcripts. Thus, in the case

Table 3. Phylogentic distribution of DUF1602-containing proteins.

Kingdom/Species	Number of DUF1602-containing proteins
Archaea (all Euryarchaeota)	5
Bacteria:Actinobacteria:Chlamydiae:Cyanobacteria:Firmicutes: Proteobacteria:	146202218100
EukaryotesFungi:Metazoa:Chlorophyta:Streptophyta:	59141143
Unclassified sequences	4

doi:10.1371/journal.pone.0157016.t003

we have examined, the position of the CRP site is only surprising on first inspection. Detailed investigation of the *aat* locus reveals that CRP acts via a well-established mechanism and it is the position of *aatS*, embedded within *aatC*, which confounds the situation. In summary, whilst some bacterial transcription factors bind primarily in expected locations [33–35] many have unusually situated targets [3,10,11,36,37]. Surprising binding sites are often ignored [36] or dismissed as artefacts [38]. This work demonstrates that careful genetic and biochemical analysis can identify regulatory function for such targets.

Supporting Information

S1 Fig. Raw gel images. Panels A-C show raw gel images from Figs 2A, 2C and 3 respectively. (PDF)

Author Contributions

Conceived and designed the experiments: DCG. Performed the experiments: JRJH. Analyzed the data: DCG JRJH. Contributed reagents/materials/analysis tools: DCG JRJH. Wrote the paper: DCG JRJH.

References

1. Browning DF, Busby SJW. The regulation of bacterial transcription initiation. *Nat Rev Microbiol.* 2004; 2: 57–64. PMID: [15035009](#)
2. Wade JT, Grainger DC. Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. *Nat Rev Microbiol.* 2014; 12: 647–653. doi: [10.1038/nrmicro3316](#) PMID: [25069631](#)
3. Shimada T, Ishihama A, Busby SJW, Grainger DC. The *Escherichia coli* RutR transcription factor binds at targets within genes as well as intergenic regions. *Nucleic Acids Res.* 2008; 36: 3950–3955. doi: [10.1093/nar/gkn339](#) PMID: [18515344](#)
4. Shimada T, Bridier A, Briandet R, Ishihama A. New roles of LeuO in transcription regulation of *E. coli* genome: antagonistic interplay with the universal silencer H-NS. *Mol Microbiol.* 2011; 82: 378–397. doi: [10.1111/j.1365-2958.2011.07818.x](#) PMID: [21883529](#)
5. Blatter EE, Ross W, Tang H, Gourse RL, Ebright RH. Domain organization of RNA polymerase α subunit: C-terminal 85 amino acids constitute a domain capable of dimerization and DNA binding. *Cell.* 1994; 78: 889–896. PMID: [8087855](#)
6. Gaston K, Bell A, Kolb A, Buc H, Busby S. Stringent spacing requirements for transcription activation by CRP. *Cell.* 1990; 62: 733–743. PMID: [2167178](#)
7. Savery NJ, Lloyd GS, Busby SJW, Thomas MS, Ebright RH, Gourse RL. 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. *J Bacteriol.* 2002; 184: 2273–2280. PMID: [11914359](#)
8. Niu W, Kim Y, Tau G, Heyduk T, Ebright RH. Transcription activation at class II CAP-dependent promoters: two interactions between CAP and RNA polymerase. *Cell.* 1996; 87: 1123–1134. PMID: [8978616](#)
9. Rhodius VA, West DM, Webster CL, Busby SJW, Savery NJ. Transcription activation at class II CRP-dependent promoters: the role of different activating regions. *Nucleic Acids Res.* 1997; 25: 326–332. PMID: [9016561](#)
10. Grainger DC, Hurd D, Harrison M, Holdstock J, Busby SJW. Studies of the distribution of *Escherichia coli* cAMP-receptor protein and RNA polymerase along the *E. coli* chromosome. *Proc Natl Acad Sci.* 2005; 102: 17693–17698. PMID: [16301522](#)
11. Haycocks JRJ, Sharma P, Stringer AM, Wade JT, Grainger DC. The molecular basis for control of ETEC enterotoxin expression is response to environment and host. *PLoS Pathog.* 2015; 11(1): e1004605. doi: [10.1371/journal.ppat.1004605](#) PMID: [25569153](#)
12. Miller JH. *Experiments in molecular genetics.* Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press. 1972.
13. Lloyd GS, Hollands K, Godfrey RE, Busby SJW. Transcription initiation in the *Escherichia coli* K-12 *mal*-*malX* intergenic region and the role of the cyclic AMP receptor. *FEMS Microbiol Lett.* 2008; 288: 250–257. doi: [10.1111/j.1574-6968.2008.01365.x](#) PMID: [19054084](#)

14. Savery NJ, Lloyd GS, Kainz M, Gaal T, Ross W, Ebright RH, et al. Transcription activation at class II CRP-dependent promoters: identification of determinants in the C-terminal domain of the RNA polymerase α subunit. *EMBO J.* 1998; 17: 3439–3447. PMID: [9628879](#)
15. Grainger DC, Goldberg MD, Lee DJ, Busby SJW. Selective repression by Fis and H-NS at the *Escherichia coli* *dps* promoter. *Mol Microbiol.* 2008; 68: 1366–1377. doi: [10.1111/j.1365-2958.2008.06253.x](#) PMID: [18452510](#)
16. Maxam AM, Gilbert W. A new method for sequencing DNA. *Proc Natl Acad Sci U S A.* 1977; 74: 560–564. PMID: [265521](#)
17. Kolb A, Kotlarz D, Kusano S, Ishihama A. Selectivity of the *Escherichia coli* RNA polymerase $E\sigma^{38}$ for overlapping promoters and ability to support CRP activation. *Nucleic Acids Res.* 1995; 23: 819–826. PMID: [7708498](#)
18. Wheeler DL, Church DM, Federhen S, Lash AE, Madden TL, Pontius JU, et al. Database resources of the National Center for Biotechnology. *Nucleic Acids Res.* 2003; 31: 28–33. PMID: [12519941](#)
19. Mitchell A, Chang H, Daugherty L, Fraser M, Hunter S, Lopez R, et al. The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res.* 2015; 43: D213–D221. doi: [10.1093/nar/gku1243](#) PMID: [25428371](#)
20. Shine J, Dalgarno L. The 3'-Terminal sequence of *Escherichia coli* 16S ribosomal RNA: Complementarity to nonsense triplets and ribosome binding sites. *Proc Natl Acad Sci.* 1974; 71: 1342–1346. PMID: [4598299](#)
21. Shultzaberger RK, Bucheimer RE, Rudd KE, Schneider TD. Anatomy of *Escherichia coli* ribosome binding sites. *J Mol Biol.* 2001; 313: 215–228. PMID: [11601857](#)
22. Chen H, Bjercknes M, Kumar R, Jay E. Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of *Escherichia coli* mRNAs. *Nucleic Acids Res.* 1994; 22: 4953–4957. PMID: [7528374](#)
23. Islam MD, Shaw RK, Frankel G, Pallen MJ, Busby SJW. Translation of a minigene in the 5' leader sequence of the enterohaemorrhagic *Escherichia coli* *LEE1* transcription unit affects expression of the neighbouring downstream gene. *Biochem J.* 2012; 441: 247–253. doi: [10.1042/BJ20110912](#) PMID: [21973189](#)
24. Crossman LL, Chaudhuri RR, Beatson SA, Wells TJ, Desvaux M, Cunningham AF, et al. A Commensal Gone Bad: Complete Genome Sequence of the Prototypical Enterotoxigenic *Escherichia coli* Strain H10407. *J Bacteriol.* 2010; 192: 5822–5831. doi: [10.1128/JB.00710-10](#) PMID: [20802035](#)
25. Nishi J, Sheikh J, Mizuguchi K, Luisi B, Burland V, Boutin A, et al. The export of coat protein from enteroaggregative *Escherichia coli* by a specific ATP-binding cassette transporter system. *J Biol Chem.* 2003; 278: 45680–45689. PMID: [12933818](#)
26. Storz G, Wolf YI, Ramamurthi KS. Small proteins can no longer be ignored. *Annu Rev Biochem.* 2014; 83: 753–77. doi: [10.1146/annurev-biochem-070611-102400](#) PMID: [24606146](#)
27. Tusnady GE, Simon I. The HMMTOP transmembrane topology prediction server. *Bioinformatics.* 2001; 17: 849–850. PMID: [11590105](#)
28. Ellis JC, Brown JW. Genes within genes in bacteria. *Trends Biochem Sci.* 2003; 28: 521–523. PMID: [14559179](#)
29. Feltens R, Gobringer M, Willkomm DK, Urlaub H, Hartmann RK. An unusual mechanism of bacterial gene expression revealed for the RNase P protein of *Thermus* strains. *Proc Natl Acad Sci.* 2003; 100: 5724–5729. PMID: [12719542](#)
30. Behrens M, Sheikh J, Nataro JP. Regulation of the overlapping *pic/set* locus of *Shigella flexneri* and Enteroaggregative *Escherichia coli*. *Infect Immun.* 2002; 70: 2915–2925. PMID: [12010980](#)
31. Minch KJ, Rustad TR, Peterson EJ, Winkler J, Reis DJ, Ma S, et al. The DNA-binding network of *Mycobacterium tuberculosis*. *Nat Commun.* 2014; 6: 5829.
32. Ishihama A, Shimada T, Yamazaki Y. Transcription profile of *Escherichia coli*: genomic SELEX search for regulatory targets of transcription factors. *Nucleic Acids Res.* 2016; 44: 2058–74. doi: [10.1093/nar/gkw051](#) PMID: [26843427](#)
33. Grainger DC, Aiba H, Hurd D, Browning DF, Busby SJW. Transcription factor distribution in *Escherichia coli*: studies with FNR protein. *Nucleic Acids Res.* 2007; 35: 269–278. PMID: [17164287](#)
34. Grainger DC, Overton TW, Reppas N, Wade JT, Tamai E, Hobman JL, Constantinidou C, Struhl K, Church G, Busby SJW. Genomic studies with *Escherichia coli* MeIR protein: applications of chromatin immunoprecipitation and microarrays *J. Bacteriol.* 2004; 186: 6938–6943.
35. Yamamoto K, Ishihama A, Busby SJW, Grainger DC. The *Escherichia coli* K-12 MntR miniregulon includes *dps*, which encodes the major stationary-phase DNA-binding protein. *J. Bacteriol.* 1993; 175: 1477–1480. doi: [10.1128/JB.01230-10](#) PMID: [21239586](#)

36. Partridge JD, Bodenmiller DM, Humphrys MS, Spiro S. NsrR targets in the *Escherichia coli* genome: new insights into DNA sequence requirements for binding and a role for NsrR in the regulation of motility. *Mol Microbiol.* 2009; 73: 680–694. doi: [10.1111/j.1365-2958.2009.06799.x](https://doi.org/10.1111/j.1365-2958.2009.06799.x) PMID: [19656291](https://pubmed.ncbi.nlm.nih.gov/19656291/)
37. Efromovich S, Grainger D, Bodenmiller D, Spiro S. Genome-wide identification of binding sites for the nitric oxide-sensitive transcriptional regulator NsrR. *Methods in Enzymology.* 2008; 437: 211–233. doi: [10.1016/S0076-6879\(07\)37012-2](https://doi.org/10.1016/S0076-6879(07)37012-2) PMID: [18433631](https://pubmed.ncbi.nlm.nih.gov/18433631/)
38. Waldminghaus T, Skarstad K. CHIP on Chip: surprising results are often artifacts. *BMC genomics.* 2010; 11: 414. doi: [10.1186/1471-2164-11-414](https://doi.org/10.1186/1471-2164-11-414) PMID: [20602746](https://pubmed.ncbi.nlm.nih.gov/20602746/)
39. Busby S, Kottlarz D, Buc H. Deletion mutagenesis of the *Escherichia coli* galactose operon promoter region. *J Mol Biol.* 1983; 167: 259–274. PMID: [6408264](https://pubmed.ncbi.nlm.nih.gov/6408264/)
40. Lodge J, Fear J, Busby S, Gunasekaran P, Kamini NR. Broad host range plasmids carrying the *Escherichia coli* lactose and galactose operons. *FEMS Microbiol Lett.* 1992; 95: 271–276.
41. Bell A, Gaston K, Williams R, Chapman K, Kolb A, Buc H, et al. Mutations that alter the ability of the *Escherichia coli* cyclic AMP receptor protein to activate transcription. *Nucleic Acids Res.* 1990; 18: 7243–7250. PMID: [2259621](https://pubmed.ncbi.nlm.nih.gov/2259621/)