

# Rothamsted Repository Download

## A - Papers appearing in refereed journals

Alhammadi, M. M., Hothersall, J., Lloyd, G. S., Titman, S., Guest, T., Browning, D. F., Grainger, D. C., Busby, S. J. W. and Haycocks, J. R. J. 2025. Genome-wide mapping of cAMP receptor protein binding in enteroaggregative *Escherichia coli* reveals targeting of virulence-associated genes. *Microbiology*. 171 (8), p. 001592.  
<https://doi.org/10.1099/mic.0.001592>

The publisher's version can be accessed at:

- <https://doi.org/10.1099/mic.0.001592>

The output can be accessed at: <https://repository.rothamsted.ac.uk/item/99475/genome-wide-mapping-of-camp-receptor-protein-binding-in-enteroaggregative-escherichia-coli-reveals-targeting-of-virulence-associated-genes>.

© 18 August 2025, Please contact [library@rothamsted.ac.uk](mailto:library@rothamsted.ac.uk) for copyright queries.

# Genome-wide mapping of cAMP receptor protein binding in enteroaggregative *Escherichia coli* reveals targeting of virulence-associated genes

Munirah M. Alhammadadi<sup>†</sup>, Joanne Hothersall, Georgina S. Lloyd, Sophie V. Titman<sup>‡</sup>, Thomas Guest<sup>§</sup>, Douglas F. Browning<sup>#</sup>, David C. Grainger, Stephen J. W. Busby and James R. J. Haycocks<sup>\*†¶</sup>

## Abstract

Bacterial pathogens employ a diverse array of virulence factors to colonize and subsequently elicit disease in their host. These factors are often subject to extensive regulation at the transcriptional level to ensure that their expression is timely. Although many pathogens use bespoke transcription factors that primarily target virulence genes, global transcription factors also sometimes play a role in controlling these genes. Enteroaggregative *Escherichia coli* (EAEC) is a significant cause of watery and mucoid diarrhoea globally. The organism colonizes the small intestine before producing toxins that elicit disease, using a multitude of virulence factors that are encoded both chromosomally and on virulence plasmids. In this work, we have studied the cAMP receptor protein (CRP), a well-characterized bacterial global transcription factor, focusing on its role in the pathogenicity of the prototype EAEC strain 042. We show that, although most functional CRP binding sites on the chromosome are conserved between *E. coli* K-12 and 042, CRP has been co-opted to couple the expression of some virulence genes to the nutritional state of the cell. We report novel mechanisms for CRP-dependent regulation of genes whose products contribute to the maturation of a bacterial antibiotic, export of a polysaccharide capsule and production of a putative adhesin.

## DATA AVAILABILITY

ChIP-seq data are available from ArrayExpress under accession code E-MTAB-14748.

## INTRODUCTION

Enteroaggregative *Escherichia coli* (EAEC) is a common cause of diarrhoea in adults and children in both industrialized and developing countries [1–3]. The organism binds to the brush border of the small intestine and forms a distinctive ‘stacked brick’ structure [4]. Various proteins, including adhesins and toxins, have been identified as key virulence factors, though it is clear that there is enormous variation from one EAEC strain to another [5, 6]. These factors are encoded by chromosomal segments that are absent from harmless laboratory strains of *E. coli*, such as the MG1655 K-12 strain, which many researchers use as a reference strain [7].

The expression of EAEC virulence genes is tightly regulated, and many of these genes are located on large virulence plasmids. Hence, the prototype EAEC 042 strain, which is often adopted as a paradigm [7], contains the pAA2 plasmid that encodes a transcription activator, AggR, which is the ‘master’ regulator of virulence [8, 9]. Transcription initiation at many promoters that

Received 19 May 2025; Accepted 11 July 2025; Published 18 August 2025

**Author affiliations:** <sup>†</sup>Institute of Microbiology and Infection, School of Biosciences, College of Life and Environmental Sciences, University of Birmingham, Birmingham, B15 2TT, UK.

**\*Correspondence:** James R. J. Haycocks, james.haycocks@newcastle.ac.uk

**Keywords:** bacterial gene regulation; cAMP receptor protein (CRP); enteroaggregative *Escherichia coli* (EAEC); transcription; virulence.

**Abbreviations:** CRP, cAMP receptor protein; EAEC, enteroaggregative *Escherichia coli*; EMSA, electrophoretic mobility shift assay.

**†Present address:** Department of Biology, College of Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

**‡Present address:** Plant Sciences and the Bioeconomy, Rothamsted Research, Harpenden, AL5 2JQ, UK

**§Present address:** Department of Microbiology, Biocenter, University of Würzburg, Würzburg, D-97074, Germany

**#Present address:** School of Biosciences, College of Health & Life Sciences, Aston University, Birmingham, B4 7ET, UK

**¶Present address:** Newcastle University Biosciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK.

A supplementary figure and four supplementary tables are available with the online version of this article.

001592 © 2025 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License. This article was made open access via a Publish and Read agreement between the Microbiology Society and the corresponding author's institution.

control expression of 042 virulence genes, both on the pAA2 plasmid and the 042 chromosome, is dependent on AggR, which is a member of the well-studied AraC/-XylS family of transcription activators [9, 10]. In previous papers, we reported the organization of several AggR-dependent promoters, and we suggested that AggR activity is regulated by a stochastically triggered feed-forward loop [10, 11]. In parallel research, we studied the expression of the EAEC 042 plasmid-borne *pet* gene that encodes a secreted toxin [known as Pet (plasmid-encoded toxin)] that is important for successful EAEC infections and is harmful to human hosts. We were surprised to find that *pet* gene expression is independent of AggR but, rather, depends on the cAMP receptor protein (CRP) [12, 13]. Recall that *E. coli* and related bacterial species exploit hundreds of different transcription factors to couple gene expression to changes in their environment. Whilst some of these, such as the lactose operon repressor, bind to a small number of DNA targets, others, such as CRP, bind at hundreds of targets and are global regulators of gene transcription [14, 15]. The activity of CRP is controlled by the level of 3'5'-cAMP, which, in general, in enteric bacteria, rises in response to certain stresses, such as starvation [16, 17].

In more recent work, we identified the *pic* gene, located on the EAEC 042 chromosome, as a second virulence determinant whose transcription depends on CRP rather than AggR [18]. This gene encodes a secreted mucinase that was originally labelled as 'protein involved in colonization' [19]. Encouraged by these results, we have now used chromatin immunoprecipitation in combination with high-throughput DNA sequencing (ChIP-seq) to identify the full complement of DNA sites for CRP in EAEC 042. As expected, we found hundreds of targets, and most are common with the MG1655 laboratory strain. However, after triage, we identified a small number of targets that are not found in MG1655. Amongst these, we found three new targets where CRP binding directly affects transcript initiation, and these are reported here. We suggest that this argues for a role for CRP during EAEC colonization and infection.

## METHODS

### Strains, plasmids and oligonucleotides

Strains, plasmids and oligonucleotide primers used are listed in Table S1, available in the online Supplementary Material. ChIP-seq was exploited to identify gene regulatory regions in EAEC 042 that are targets for CRP. Using PCR and primers listed in Table S1, these regions were amplified and flanked with *EcoRI* and *HindIII* to facilitate cloning into the *lac* expression vectors, pRW224 or pRW50, which were then used to measure promoter activities in strains M182 and M182 $\Delta$ *crp*, exactly as in our previous studies [12, 13]. Differences found in M182 $\Delta$ *crp* were complemented by plasmid pDCRP and derivatives. For *in vitro* studies, *EcoRI-HindIII* fragments carrying different gene regulatory regions were cloned into the pSR 'holding' plasmid. For plasmid maintenance, strains harbouring pSR and pDCRP were grown in media supplemented with 100  $\mu$ g ml<sup>-1</sup> ampicillin, while strains harbouring pRW224 and pRW50 were grown in media containing 15  $\mu$ g ml<sup>-1</sup> tetracycline.

### DNA constructs

DNA fragments used for electrophoretic mobility shift assay (EMSA) experiments were generated by PCR using *E. coli* 042 gDNA as template using primers listed in Table S1. The *pmchA* construct was generated by cloning the 619 bp *pmchA*<sub>600</sub> construct *EcoRI/HindIII* into pRW50, which served as an intermediate template for the generation of the shorter *pmchA* construct by site-directed mutagenesis (SDM), using primers *pmchA*<sub>200</sub> SDM F/R. For cloning *EcoRI/HindIII* into pSR, primers *pmchA*<sub>200</sub> F and *pmchA* R were used to amplify the 213 bp *pmchA* fragment. The *pmchA* -37G, -11G and -11A mutations were created by SDM, using the *pmchA* construct cloned into either pRW50 or pSR as a template. Promoter fragments sub-cloned between pRW50 and pSR were amplified with primers *pmchA* F/*pmchA* R and cloned with *EcoRI/HindIII*. To make the *pmchA* -14A mutant, two separate 'arms' were generated (using *pmchA* SOE -14A left arm F/R and *pmchA* SOE -14A right arm F/R primers), which were spliced together by overlap extension PCR to generate the final product. This was then cloned *EcoRI/HindIII* into pRW50.

Mutations to the 042 p0536 promoter construct were introduced using SDM, except for the p0536 -38T -36A construct, which was generated by PCR using overlapping oligos, and then cloned *EcoRI/HindIII* into pRW50. The *pkpsM*<sub>201</sub> DNA fragment was generated using *pkpsM*<sub>201</sub> F/R primers. *pkpsM*<sub>440</sub> was constructed using primers *pkpsMII*<sub>440</sub> F/R, with mutations being introduced using *pkpsM* P1- F and *pkpsM* P2- R.

All plasmid constructs were confirmed by Sanger DNA base sequencing.

### ChIP-seq

Chromatin immunoprecipitation experiments were carried out in duplicate, using the method of Middlemiss *et al.* [20] with modifications. EAEC 042 overnight cultures were sub-cultured into 40 ml of fresh Luria-Bertani (LB) media and grown to mid-log phase ( $OD_{600}$  = 0.4–0.5) and were crosslinked in 1% v/v formaldehyde for 20 min. The reaction was quenched by adding glycine to 0.5 M, and cells were washed in an equal volume of cold 1  $\times$  TBS. Lysis was carried out by re-suspending cell pellets in 1 ml immunoprecipitation buffer [consisting of 50 mM HEPES-KOH pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton-X 100 and 0.1% sodium deoxycholate, with 1  $\times$  EDTA-free protease inhibitor cocktail (Roche) added per 50 ml] containing 2 mg ml<sup>-1</sup> lysozyme at 37 °C for 30 min. Samples were chilled on ice prior to sonication using a Bioruptor Plus instrument (using

the high-intensity setting, for 30 cycles, 30 s on, 30 s off). Lysates were clarified by centrifugation and immunoprecipitation was carried out by adding anti-CRP antibody (clone 1D8D9, purchased from Biolegend), with overnight incubation at 4 °C on a rotating wheel. Immunocomplexes were immobilized by adding 15 µl each of washed Protein A and Protein G Dynabeads (Invitrogen) for 1 h at 4 °C on a rotating wheel. Samples were then washed six times with immunoprecipitation buffer, once with high salt immunoprecipitation buffer (containing 500 mM NaCl), once with ChIP wash buffer (20 mM Tris HCl pH 8.0, 250 mM LiCl, 0.5% IGEPAL CA-630, 0.5% sodium deoxycholate and 1 mM EDTA) and once with TE buffer. DNA was eluted in buffer containing 50 mM Tris-HCl pH 7.5, 10 mM EDTA and 1% SDS. To decrosslink, 5 µl of 10 mg ml<sup>-1</sup> proteinase K (Sigma-Aldrich) was added to the eluted DNA, and samples were incubated at 42 °C for 1 h, followed by incubation at 65 °C for 5 h. DNA was purified using a QIAGEN PCR Purification Kit, and libraries were prepared using an NEBnext Ultra II DNA Library Preparation Kit. Library size distributions were checked using an Agilent TapeStation instrument (using High-Sensitivity DNA ScreenTapes) and were quantified by quantitative PCR using an NEBNext Library Quantification Kit. Libraries were pooled in an equimolar ratio and sequenced by Azenta Life Sciences (next-generation sequencing services) on a MiSeq instrument (paired-end, 2×150 bp configuration). Sequencing data have been submitted to ArrayExpress and are available under accession code E-MTAB-14748.

## Bioinformatics

Reads were aligned to the EAEC 042 chromosome and pAA2 plasmid (references genome accession numbers NC\_017626.1 and NC\_017627.1, respectively) using Bowtie2 [21]. Following alignment of Illumina reads to the reference genomes, MACS2 software [22] was used to identify peaks enriched in immunoprecipitated samples, compared to the mock controls (a cut-off value of  $q=0.05$  was used), and called peaks were visually inspected using the Artemis genome browser. The R package ChIPseeker was used to identify the nearest ORF to peak centres [23]. To identify a binding motif, 150 bp of the DNA sequence flanking each peak centre was extracted and submitted to the MEME suite using the MEME tool [24].

## β-Galactosidase assays

β-Galactosidase assays were done using the method of Miller [25]. Overnight cultures of strain M182 (and the *Δcrp* derivative) were sub-cultured into fresh LB (with appropriate antibiotics) and grown to mid-log phase ( $OD_{600}=0.3-0.6$ ) aerobically with shaking at 37 °C.

## Electrophoretic mobility shift assays

DNA fragments were generated by excision from a verified plasmid construct, using *EcoRI/AatII* and *HindIII* restriction enzymes. Each fragment was end-labelled using <sup>32</sup>P-γ-ATP (purchased from Hartmann Analytic GmbH) and T4 polynucleotide kinase (NEB). EMSA reactions were carried out as previously described [18]. For each reaction, 0.5 ng of the <sup>32</sup>P-radiolabelled DNA fragment was incubated with or without CRP in buffer containing 20 mM HEPES (pH 8.0), 5 mM MgCl<sub>2</sub>, 50 mM potassium glutamate, 1 mM DTT, 0.2 mM cAMP, 5% (v/v) glycerol, 25 µg ml<sup>-1</sup> herring-sperm DNA and 0.5 mg ml<sup>-1</sup> bovine serum albumin. Reactions were incubated at 37 °C for 30 min before being run on a 6% polyacrylamide gel that was visualized using a Bio-Rad PMI Personal Molecular Imager.

## Potassium permanganate footprinting

Potassium permanganate footprinting was carried out as previously described [18]. DNA carrying the promoter was purified from *HindIII/AatII*-digested pSR, containing the cloned *EcoRI/HindIII* fragment, and labelled at the *HindIII* end with γ<sup>32</sup>P-ATP. Each reaction (20 µl) contained c. 3 nM labelled DNA fragment in 20 mM HEPES (pH 8.0), 5 mM MgCl<sub>2</sub>, 50 mM potassium glutamate, 1 mM DTT and 0.5 mg ml<sup>-1</sup> bovine serum albumin and 0.2 mM cAMP, 100 nM CRP and 50 nM RNA polymerase as required. Footprinting reactions were incubated at 37 °C for 30 min and then treated with 200 mM potassium permanganate for 4 min. Reactions were then stopped with 50 µl of stop solution (3 M ammonium acetate, 0.1 M EDTA pH 8.0, 1.5 M β-mercaptoethanol). Following phenol–chloroform extraction and ethanol precipitation, samples were resuspended in 40 µl 1 M piperidine and incubated for 30 min at 90 °C. Samples were again purified by phenol–chloroform extraction and ethanol precipitation and then resuspended in 8 µl loading buffer (95% formamide, 20 mM EDTA pH 8.0, 0.025% bromophenol blue and 0.025% xylene cyanol FF). Samples were analysed by 6% denaturing gel electrophoresis. Gels were calibrated with a Maxam–Gilbert ‘GA’ sequencing reaction of the labelled fragment and visualised with an Amersham Typhoon 5 Biomolecular Imager.

## In vitro transcription

*In vitro* transcription reactions were carried out using the method of Kolb *et al.* [26]. Template DNA (used at a concentration of 16 µg ml<sup>-1</sup>) was incubated with or without CRP in buffer containing 40 mM Tris-acetate pH 7.9, 10 mM MgCl<sub>2</sub>, 1 mM DTT, 100 mM KCl, 100 µg ml<sup>-1</sup> BSA, 200 µM ATP/CTP/GTP, 10 µM and 5 µCi α<sup>32</sup>P-UTP for 10 min at 37 °C. Reactions containing CRP were supplemented with 0.2 mM cAMP. To start the reaction, σ<sup>70</sup>-containing RNA polymerase holoenzyme was added

at a concentration of 50 nM for 10 min at 37°C. Reactions were stopped by the addition of stop solution (95% formamide, 10 mM EDTA pH 8.0, 0.025% bromophenol blue and 0.025% xylene cyanol FF). Reaction products were separated on a 6% denaturing acrylamide gel and were visualized using a Bio-Rad PMI Personal Molecular Imager.

## Proteins

CRP was purified using the Ghosaini method [27]. RNA polymerase holoenzyme was purchased from NEB.

## RESULTS AND DISCUSSION

### ChIP analysis of CRP binding targets in EAEC 042

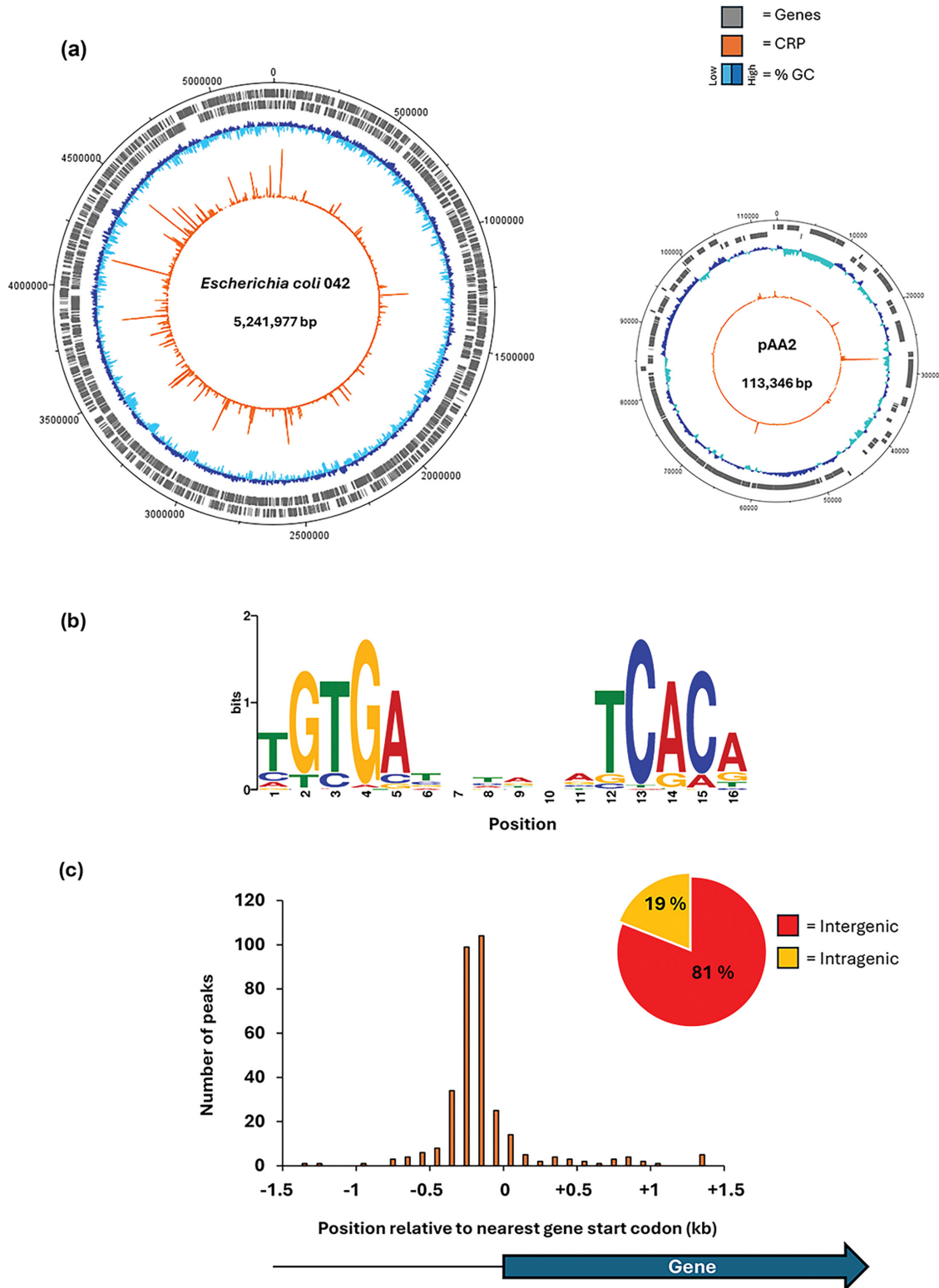
RegulonDB, the leading transcription regulation database for *E. coli*, lists over 300 transcription units where CRP is involved in regulation [28]. In most cases, CRP binds to one or more specific target sequences at the regulatory region of the unit and alone or in combination with other transcription factors, either activates or represses transcript initiation. In previous work, working with laboratory strains of *E. coli*, ChIP was applied directly to measure CRP binding, independent of any regulatory function [14]. These studies confirmed hundreds of targets but, surprisingly, also identified targets outside of gene regulatory regions where CRP bound, apparently, with no ensuing effects on transcription [29]. Prompted by our observation that the promoters of the EAEC *pet* and *pic* genes are dependent on CRP for transcript initiation, we applied ChIP to map sites at which CRP is bound in EAEC 042 growing in mid-log phase in LB media. Fig. 1a illustrates the distribution of the 322 sites on the EAEC 042 chromosome and the 10 sites on the virulence plasmid: these are listed in Table S2, along with their location with respect to the nearest coding region. We used MEME to deduce a CRP-binding logo from DNA base sequences centred on each peak (Fig. 1b). As expected, this motif consists of two copies of the 5-bp element 5'-TGTGA-3', organized as an inverted repeat, separated by 6 bp. Note that this corresponds with previously derived logos for CRP [30]. Inspection of the location of these targets showed that, as in previous work, nearly one in five of the DNA sites for CRP is at an intragenic location and is unlikely to be directly associated with any transcriptional regulation (Fig. 1c) [14, 31].

In order to minimize false positives, we applied a cutoff to our data for targets assigned a score (during peak calling) of 100 or more: this reduced the number of targets to 150 on the EAEC chromosome and 5 on the plasmid. Inspection of these revealed that 9 are located in chromosome segments that are absent from the reference MG1655 genome, whilst 117 are located to regulatory regions that have been annotated in MG1655, with CRP identified as a regulatory factor. For each of these 117 targets, we compared the MG1655 base sequence from position -80 to +20 with respect to the transcript start (according to RegulonDB) with the EAEC 042 sequence (see Table S3). Identical sequences were found at 55 targets and only 3 targets showed more than 3 base differences. We found single base insertions, upstream of the promoter -35 region in the *serC* and *pckA* regulatory regions.

### CRP interactions at uncommon EAEC 042 targets

Table 1 lists the nine EAEC 042 chromosomal targets not found in MG1655, together with the five targets on the plasmid. To validate these targets, we used EMSAs using purified CRP and <sup>32</sup>P-end labelled fragments covering each regulatory region (generated by PCR; see Table S1). We did not test the 2670 and *set1A* targets as they appeared to be outside of any regulatory region. Similarly, with the plasmid-borne targets, we ignored RS26580 and the two *repA* targets as they appear to be connected with core plasmid functions, and the DNA site for CRP in the *pet* regulatory region has been already exhaustively characterized [13]. Fig. 2 shows the bandshift experiment with the remaining eight targets. The data show a strong CRP interaction with six targets, but not with 3975 or *virK*. Inspection of the chromosomal context of each of the strong binding sites suggested that the EAEC 042 *kpsM*, *mchA*, 0536 and 0225 targets were most likely to be associated with a promoter-active regulatory region. Hence, DNA fragments carrying each region were cloned into plasmid pRW50, a low copy number *lac* expression vector. Each recombinant plasmid was used to transform the *E. coli* K-12  $\Delta$ *lac* strains, M182 and M182 $\Delta$ *crp*, and  $\beta$ -galactosidase activities were measured. These activities are taken as a measure of promoter activity, and comparison of levels in the *crp*<sup>+</sup> and  $\Delta$ *crp* strains gives an indication of any CRP-dependent regulation. As a control, an assay with pRW50 carrying the semi-synthetic CC(-41.5) promoter, whose activity is tightly coupled to CRP [32], is included. The data, illustrated in Fig. 3, show that the *pkpsM*\_201, *pmchA* and p0536 fragments carry at least one promoter. The observed activity with *mchA* is clearly CRP-dependent, whilst, with the *pkpsM*\_201 and p0536 fragments, activity is higher in the  $\Delta$ *crp* strain, suggesting repression by CRP. These effects were investigated further, and our experiments are described in the next sections. Interestingly, where these genes are found in other EAEC isolates, the associated CRP binding site appears to be conserved for these targets (Table S4 and Fig. S1). Whilst there is no evidence for any promoter activity associated with the 0225 target, we were concerned that CRP might be regulating expression of the divergent gene, 0224, and so the cloned fragment was inverted in pRW50. However, measured expression levels were very low (data not shown), and we conclude that the intergenic region between 0224 and 0225 may be bereft of promoter activity (at least in our experimental conditions).





**Fig. 1.** Genome-wide distribution of CRP as determined by ChIP-seq. (a) The circular plots show binding of CRP across the EAEC 042 chromosome and plasmid pAA2. Genes are shown as grey blocks, the G+C content is shown as a dark blue and cyan plot and CRP binding is shown as an orange plot. (b) The binding motif obtained from 332 CRP peak sequences using MEME suite [24]. (c) Histogram showing positions of CRP binding motifs relative to gene start codons. Four peaks were excluded that fall outside the range of the x-axis. The pie chart shows the proportion of CRP binding sites that are located within genes or in intergenic regions.

**Table 1.** Uncommon targets for CRP in EAEC strain 042

Peak centre*	Score†	Associated gene‡	Distance to translation start§	Binding site sequence ¶	Associated gene product**
<b>Targets located on the chromosome</b>					
3356318	6783	3143	+38.5	TGTGATCTACAACACG	Hypothetical protein
3459052	894	<i>kpsM</i>	−492.5	TGTGATTATATCACA	Polysialic acid transport permease protein
3421201	615	<i>mchA</i>	−79.5	TGTGAGCTTATGCACA	Microcin activation protein
461029	312	0414	+72.5	TGTGCGCAAGATCACA	Conserved hypothetical protein
4217753	250	3975	−33.5	CGTGTATACGATAACA	Hypothetical protein
2833992	209	2670	+230.5	TGTGATACACAGCAAC	Conserved hypothetical protein
4926886	174	<i>set1A</i>	+48.5	ACTGACGGTTTCCCA	Enterotoxin 1
608383	168	0536	−120.5	TGTGATGTAAAGCGCA	Putative adhesin
263739	111	0225	−80.5	TGTGAGCCGCATCACA	Putative type VI secretion system protein
<b>Target located on plasmid pAA2</b>					
27988	2975	<i>pet</i>	−117.5	CGAGAGCATTTGTCACA	Serine protease autotransporter toxin Pet
62001	907	RS26580	−247.5	TGTGACCATATATCA	TraS
18257	387	<i>virK</i>	+235.5	TGTGGTGACTGGTACA	Virulence protein
112767	321	<i>repA</i>	−84.5	ACAGATCTTCGTCACA	Replication protein A
108769	293	<i>repA</i>	−84.5	TGTGACGAAGATCTGT	Replication protein A

Experimentally determined DNA sites for CRP that are not found in *E. coli* K-12.

\*Peak centre coordinate as defined by MACS2.

†Score assigned by MACS2 during peak calling.

‡Adjacent annotated gene in EAEC 042.

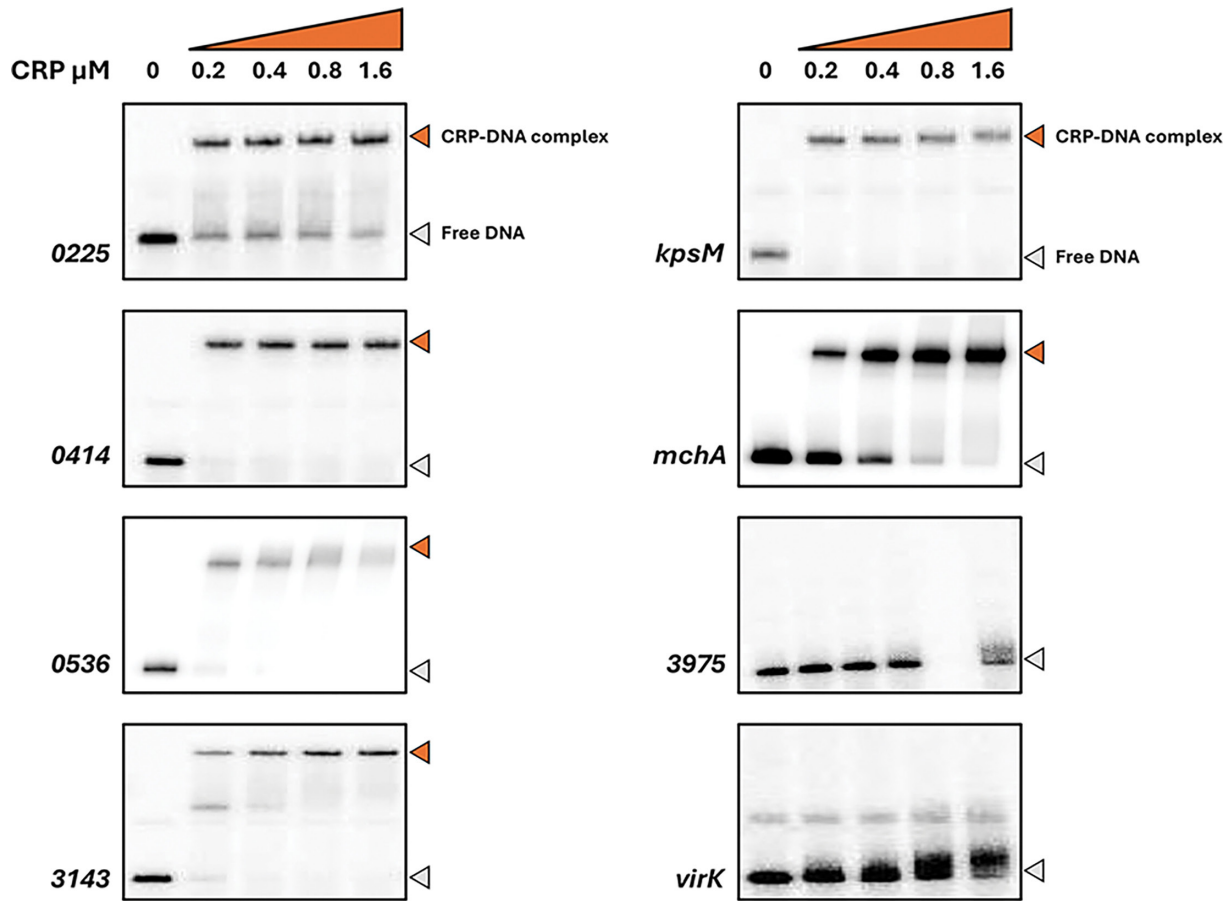
§Distance from DNA site for CRP to translation start site of annotated gene (‘−’ indicates upstream, ‘+’ indicates downstream).

¶Base sequence of DNA site for CRP.

\*\*Product of adjacent gene according to Chaudhuri et al. [7].

### CRP-dependent activation at the *mchA* promoter

The EAEC 042 *mchA* gene encodes an enzyme that transfers a glucose residue to microcin H47, a well-characterized bacterial antibiotic [33–36]. The significance of the glucose residue is that it serves as a bridge to link a siderophore molecule, thereby creating a siderophore-toxin which can play an important role in bacterial colonization and, hence, virulence [37]. Data in Fig. 3 suggest that *mchA* expression may be due to a strong promoter that is dependent on CRP. Recall that most bacterial promoters carry crucial sequence elements located ~10, 14 and 35 bases upstream of the transcript start site (known as the −10, extended −10 and −35 elements, respectively) and that one or more of these is usually defective at promoters dependent on an activator such as CRP [38, 39]. Moreover, CRP-activated promoters fall into two major classes: Class I promoters, the DNA site for CRP is located over 15 bases upstream from the promoter −35 element, whilst for Class II promoters, it overlaps the −35 element [40]. Inspection of the base sequence adjacent to the proposed DNA site for CRP at the *mchA* regulatory region suggested the existence of a Class II CRP-dependent promoter (Fig. 4a). On this basis, we created the −37G and −14A mutations that would be expected to inactivate the DNA site for CRP and the extended −10 element, respectively. DNA fragments carrying these base substitutions were cloned into pRW50: data in Fig. 4b show that CRP-dependent activation is completely stopped by either mutation. Furthermore, a −11G mutation within the predicted −10 hexamer reduced activity and a −11A mutation, to make the base at the −11 position consensus, increased activity (Fig. 4b). To confirm the existence of a Class II CRP-activated promoter, we complemented the M182Δ*crp* strain with plasmid-borne *crp* carrying substitutions in either or both of the major activating regions (known as AR1 and AR2). Data in Fig. 4c show that both AR1 and AR2 are needed for optimal complementation, with AR2 making the larger contribution to activation. Since AR2 is known to play little or no role at Class I CRP-activated promoters [40], taken together, the evidence suggests that EAEC 042 *mchA* expression depends on a single Class II CRP-dependent promoter, and this was confirmed by an *in vitro* transcription assay (Fig. 4d).

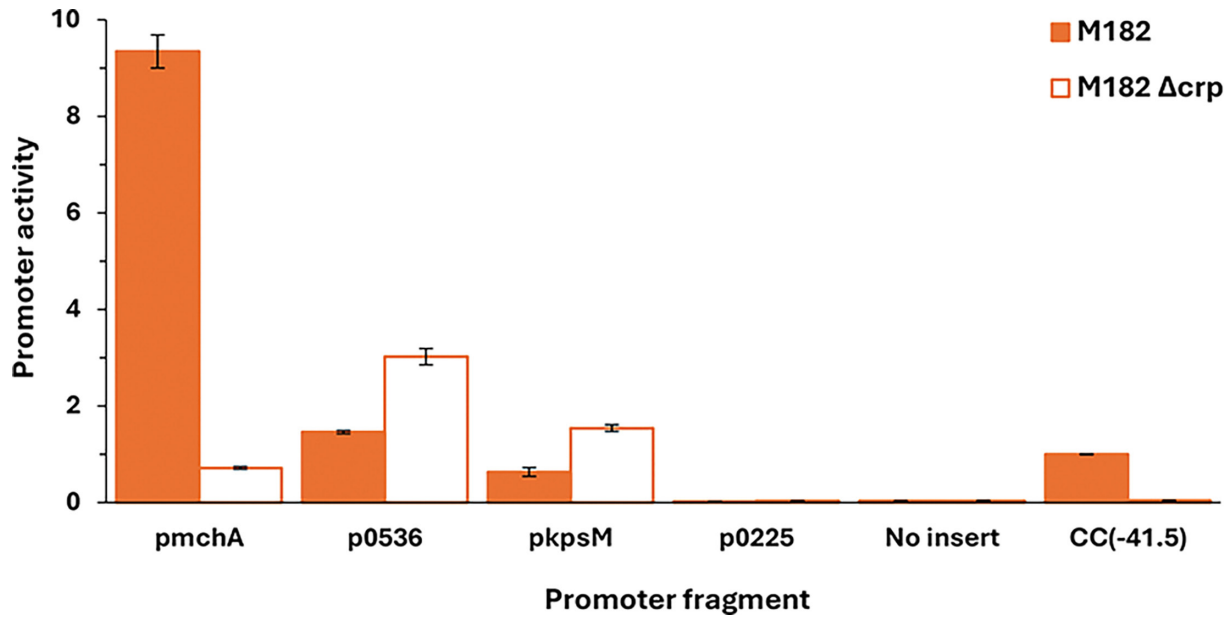


**Fig. 2.** Electrophoretic mobility shift assays of CRP binding to EAEC 042 targets. Eight 042 targets, each ~200bp in length, were excised from pSR as *HindIII*/*EcoRI* or *HindIII*/*AatIII* fragments and radiolabelled with  $\gamma^{32}$ -P ATP. These were incubated in the presence of cAMP, with increasing concentrations of CRP (0–1.6  $\mu$ M) and then separated by electrophoresis on a 6% polyacrylamide gel. In each case, the band in the first lane shows the mobility of free DNA and is indicated by a grey arrow. In subsequent lanes, a reduced mobility band indicates a DNA fragment complexed with CRP, marked with an orange arrow.

### CRP-dependent repression at the 0536 promoter

Based on sequence homology, the EAEC 042 0536 gene appears likely to encode an adhesin [7]. The data in Fig. 3 suggest that the 0536 gene regulatory region contains one or more promoters and that CRP could down-regulate expression from them. Inspection of the base sequence upstream of the 0536 translation start suggested 4 possible promoter –10 elements. Hence, starting with the cloned p0536 regulatory region fragment, we constructed single base changes that would be expected to inactivate each of the putative –10 elements denoted –11G, –1G, 17G and 25G (Fig. 5a). Assays of the promoter activity of each mutant fragment showed that the –11G mutation alone caused complete loss of activity (Fig. 5b), suggesting that the 0536 regulatory region contains a single major promoter. This was corroborated by a permanganate footprint and an *in vitro* transcription assay (Fig. 5c, d). The transcription assay shows an RNA that initiates 7 bp downstream from the suggested 5'-TAGATT-3' –10 element, and the permanganate shows the expected open complex unwinding starting at the upstream end of the –10 element. Both the footprint and the transcript are suppressed by the inclusion of CRP in the assay (Fig. 5c, d). Taken together, the data argue that the EAEC 042 0536 gene regulatory region contains a single promoter whose activity is down-regulated by CRP. Inspection of the promoter base sequence shows that the most likely –35 element is 5'-TTTACA-3', located 17-bp upstream of the –10 element, which corresponds to the optimal spacing. This is sandwiched between the two flanking 5-bp elements that comprise the DNA site for CRP, and this suggests a simple mechanism for the repressive action of CRP. To confirm this, the –38T –36A double base substitution was constructed in the cloned p0536 regulatory region fragment (Fig. 5a). These substitutions, which create a consensus upstream 5-bp element in the DNA site for CRP, result in stronger repression of the 0536 promoter by CRP (Fig. 5e).





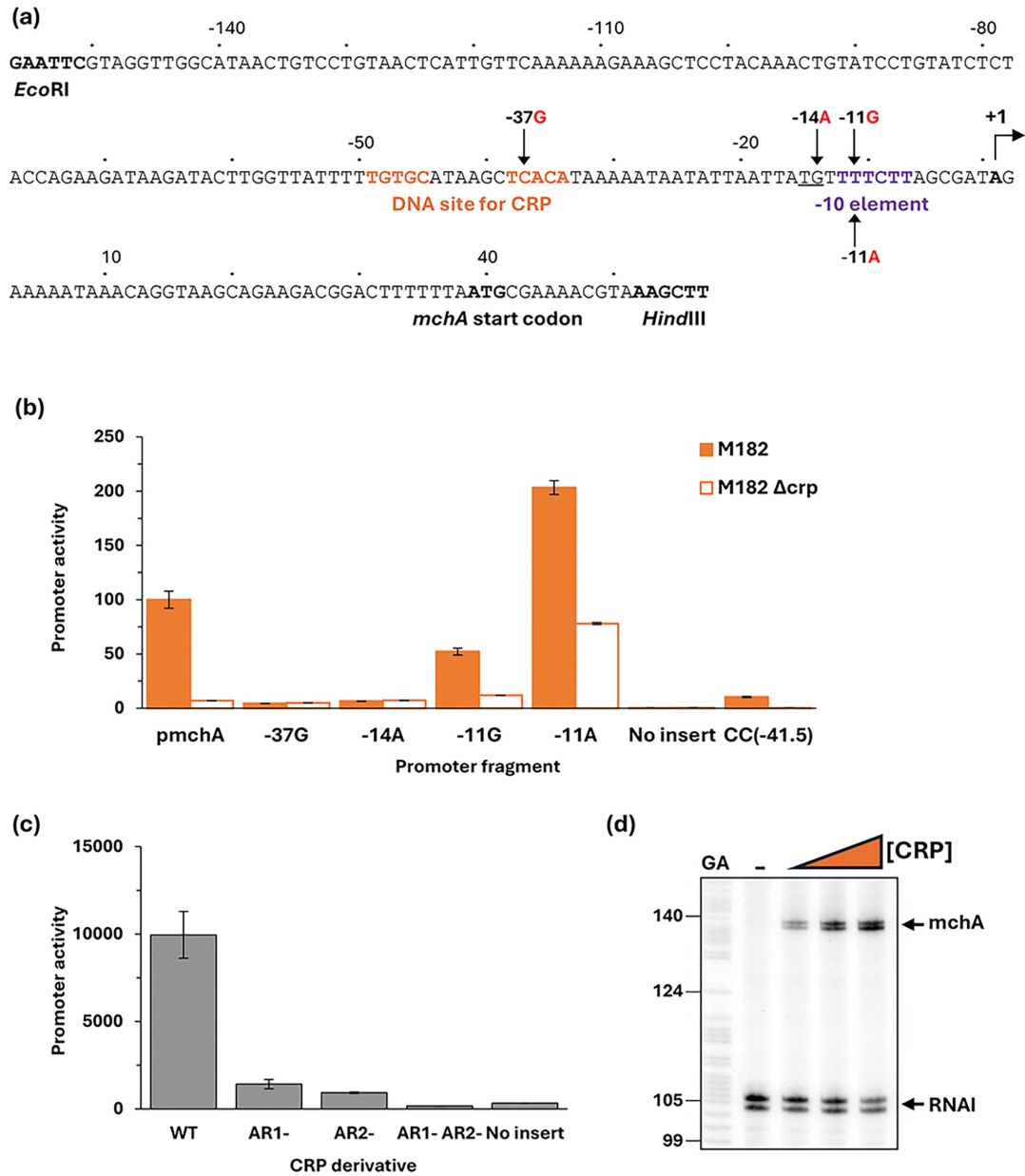
**Fig. 3.** CRP regulation at EAEC 042 promoters. The figure shows  $\beta$ -galactosidase activity measured in *E. coli* K12  $\Delta$ lac M182 and M182  $\Delta$ crp strains carrying a lacZ reporter plasmid pRW50 containing a target promoter region: pmchA, p0536, pkpsM<sub>201</sub> or p0225. Activities are expressed as fold activity compared to the activity of pRW50 carrying the promoter, CC(-41.5), measured in M182 crp<sup>+</sup> cells. pRW50 with no insert was included as a negative control. Data are representative of duplicate independent experiments, averaged from three biological replicates, with SD shown.

### CRP-dependent repression at the kpsM regulatory region

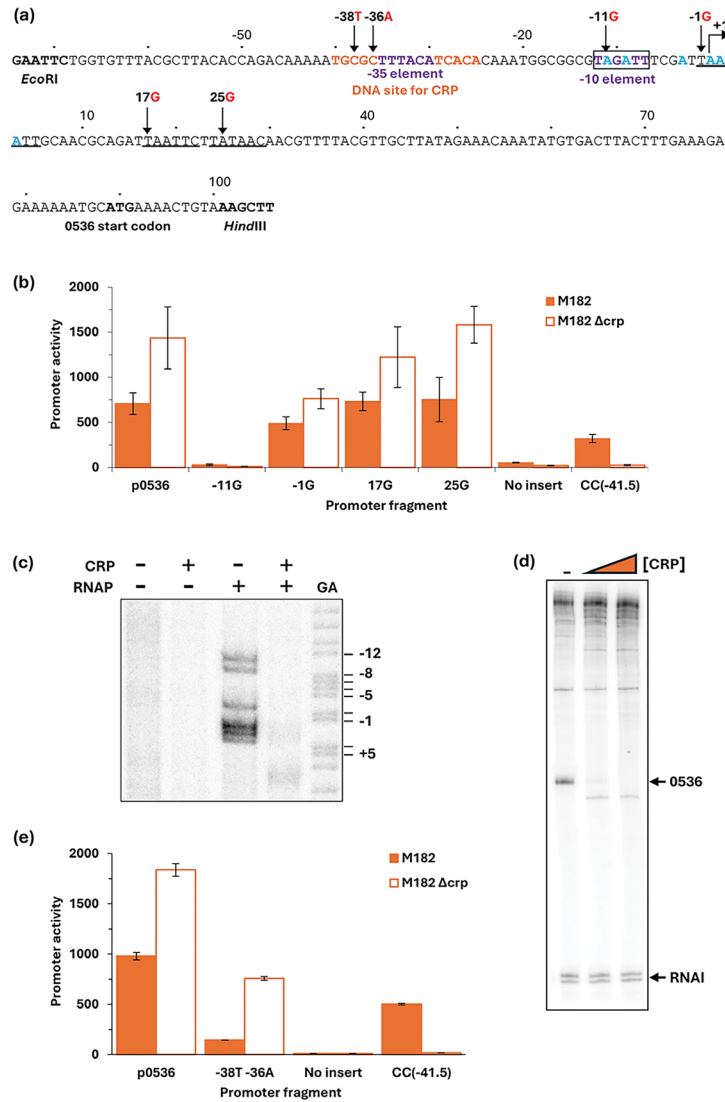
The *E. coli* kps gene products enable the production of a form of capsule known as K antigen [41–43]. Although absent in MG1655, the kps genes are found in many *E. coli* pathovars, and their expression has been studied in some detail [44, 45]. The kpsM gene is unusual in that its promoter is located ~750-bp upstream from its start codon. Inspection of the regulatory region showed that the DNA site for CRP is ~250-bp downstream from this promoter (Fig. 6a, denoted here as P1), which is consistent with the observed repression by CRP (Fig. 3). However, we noted that the two consensus halves of this DNA site for CRP flank a near perfect promoter –10 sequence (5'-TATAAA-3'), suggesting the possible existence of a second promoter (denoted here as P2). To investigate this, we cloned a 440-bp DNA fragment covering both the CRP binding site and the previously identified promoter region (pkpsM<sub>440</sub>, Fig. 6a). We made single base changes in the pkpsM<sub>440</sub> fragment, in which we corrupted the upstream promoter –10 element (P1), the downstream –10 element (P2) or both (P1 P2), and we measured the net promoter activity in our crp<sup>+</sup> and  $\Delta$ crp strains as in Fig. 3. The data, illustrated in Fig. 6b, show that kpsM promoter activity is largely suppressed by the two base changes together. Crucially, the fragment carrying the P1 mutation shows significant activity in the  $\Delta$ crp genetic background, but this activity is greatly suppressed in the crp<sup>+</sup> strain. These data are consistent with a model in which the activity of the dominant upstream promoter (P1) is moderated by CRP, but a second weaker downstream promoter (P2) is sharply repressed by CRP because its –10 hexamer element is embedded in the DNA site for CRP. This was corroborated by an *in vitro* transcription assay (Fig. 6c).

## CONCLUSIONS

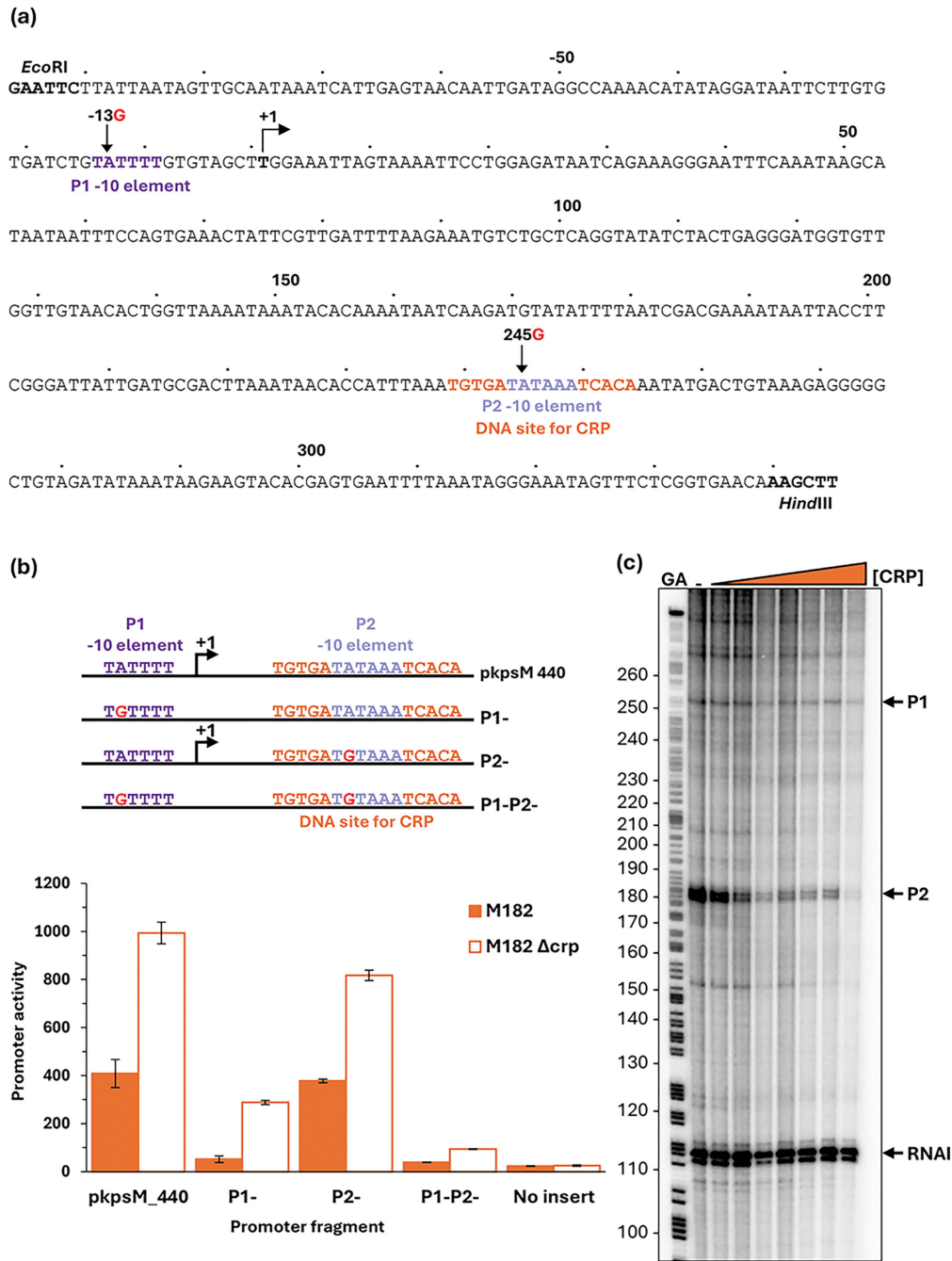
The starting point of this work was a ChIP-seq experiment, designed to identify the different DNA sites for CRP on the chromosome and virulence plasmid of the enteroaggregative *E. coli* strain 042. Arguably, CRP is the most widely studied bacterial transcription factor. It was discovered originally as the essential activator of *E. coli* lactose operon expression [46], and its activity was found to be triggered by the binding of cAMP [47–50]. Subsequently, it was shown to regulate hundreds of *E. coli* transcription units, and its biological role appears to be to reprogramme gene transcription in response to nutritional stresses [30, 51–53]. Our research, reported here, was prompted by the observation that transcription of two key EAEC 042 virulence genes (*pet* and *pic*) directly depends on CRP [13, 18], rather than AggR, which is regarded as the master regulator of virulence in EAEC [9]. We wanted to discover if the transcription of any other EAEC virulence genes was likewise activated by CRP. Our experiments showed that, as expected, CRP binds to hundreds of gene regulatory regions and that the vast majority of binding targets are common with the MG1655 *E. coli* reference strain. However, after triage, we found CRP-dependent regulation at three genes that are not found in MG1655. In one case, mchA, CRP is an activator, whilst in the other two cases, 0536 and kpsM, CRP downregulates gene expression. We propose that, whilst EAEC virulence is primarily controlled by the activity of AggR, CRP activity, which



**Fig. 4.** CRP-dependent activation of the EAEC 042 *mchA* promoter. (a) Sequence of the *pmchA* promoter fragment, amplified to incorporate *EcoRI* and *HindIII* restriction sites. Numbering is relative to the predicted transcript start site, noted by +1. Locations of the -10 element and CRP binding site are shown in bold purple and orange, respectively; the extended -10 is underlined, and substitutions are labelled accordingly, for example, -11G indicates the base at position -11 is changed to a G. (b) This panel shows mutational analysis of the *mchA* promoter.  $\beta$ -Galactosidase activity was measured in *E. coli* K12  $\Delta$ *lac* strains M182 and M182  $\Delta$ crp carrying a reporter plasmid pRW50 containing either the starting *mchA* promoter (*pmchA*) cloned upstream of *lacZ* or the *mchA* promoter with a substitution in the CRP binding site (-37G), in the extended -10 (-14A) or the -10 element (-11G or -11A). Promoter activities are expressed as a percentage relative to the starting *mchA* promoter activity in M182 *crp*<sup>+</sup> cells. pRW50 with no insert was included as a negative control. Data are from duplicate independent experiments, averaged from three biological replicates, with SD shown. (c) Complementation with CRP derivatives demonstrated Class II activation of the *mchA* promoter.  $\beta$ -Galactosidase activity was measured in M182  $\Delta$ crp cells containing the WT *mchA* promoter *lacZ* fusion (in pRW50) with derivatives of CRP expressed from pDCRP: WT CRP (WT), CRP HL159 mutated in AR1 (AR1-), KE101 in AR2 (AR2-) and HL159 KE101 in both AR1 and AR2 (AR1- AR2-). pD with no CRP insert was included as a negative control.  $\beta$ -Galactosidase activity is expressed as Miller units (nmol of ONPG hydrolysed min<sup>-1</sup> mg<sup>-1</sup> dry cell mass). Data are from duplicate independent experiments, averaged from three biological replicates, with SD shown. (d) A CRP-dependent transcript was detected by multi-round *in vitro* transcription from the *mchA* promoter, cloned in pSR plasmid, with *E. coli* holo-RNA polymerase (50 nM). The 110 nt RNAI transcript is an internal positive control. Both the *mchA* and RNAI transcripts are indicated by arrows. A Maxim-Gilbert GA sequencing reaction was used to calibrate the transcripts. CRP concentrations: lane 1 (0  $\mu$ M), lane 2 (25 nM) and lane 3 (100 nM).



**Fig. 5.** CRP-dependent repression of the EAEC 042 0536 promoter. (a) Sequence of the p0536\_100 promoter fragment, amplified to incorporate *EcoRI* and *HindIII* restriction sites. Numbering is relative to the predicted transcript start site, denoted by +1. Locations of the -10 element and predicted -35 element are shown in purple (with a box around the -10); the CRP binding site is shown in orange. Other potential -10 elements are underlined, and substitutions are labelled accordingly, for example, -11G indicates the base at position -11 is changed to a G. (b) Mutational analysis was used to locate the 0536 promoter -10 element.  $\beta$ -Galactosidase activity was measured in *E. coli* K12  $\Delta$ lac strains M182 and M182  $\Delta$ crp carrying a reporter plasmid pRW50 containing either the WT 0536 promoter (p0536\_100) cloned upstream of *lacZ* or the 0536 promoter with a mutation in one of four possible -10 elements: -11G (p0536\_105), -1G (p0536\_106), +17G (p0536\_103) or +25G (p0536\_104). pRW50 with the CC(-41.5) promoter was included as a positive control.  $\beta$ -Galactosidase activity is expressed as Miller units (nmol of ONPG hydrolysed min<sup>-1</sup> mg<sup>-1</sup> dry cell mass). Data are from duplicate independent experiments, averaged from three biological replicates, with SD shown. (c) Unwinding of the open complex at the -10 element was confirmed with potassium permanganate footprinting of an end-labelled *AatII*/*HindIII* 0536 promoter fragment excised from pSR, incubated in the presence of cAMP (200  $\mu$ M), with (+) and without (-) CRP (100 nM) and RNA polymerase (50 nM). A Maxim-Gilbert GA sequencing reaction was used to calibrate the gel. The locations of cleavage sites are shown in blue in panel (a). (d) A transcript, produced in the absence of CRP, was detected by multi-round *in vitro* transcription from the p0536\_100 promoter, cloned in pSR plasmid, with *E. coli* holo-RNA polymerase (50 nM). The 110 nt RNAI transcript is an internal positive control. Both the 0536 and RNAI transcripts are indicated by black arrows. A Maxim-Gilbert GA sequencing reaction was used to calibrate the transcripts. CRP concentrations: lane 1 (0  $\mu$ M), lane 2 (25 nM) and lane 3 (100 nM). (e) Substitutions to create a consensus CRP binding site resulted in stronger CRP repression of the 0536 promoter.  $\beta$ -Galactosidase activity was measured in *E. coli* K12  $\Delta$ lac M182 and M182  $\Delta$ crp strains carrying a reporter plasmid pRW50 containing either the WT 0536 promoter (p0536\_100) cloned upstream of *lacZ* or the 0536 promoter with a double base substitution -38T -36A (p0536\_107) in the CRP binding site to make it consensus. pRW50 with the CC(-41.5) promoter was included as a positive control.  $\beta$ -Galactosidase activity is expressed as Miller units (nmol of ONPG hydrolysed min<sup>-1</sup> mg<sup>-1</sup> dry cell mass). Data are from duplicate independent experiments, averaged from three biological replicates, with SD shown.



**Fig. 6.** CRP-dependent repression of the EAEC 042 *kpsM* promoter. (a) Sequence of the *pkpsM\_440* promoter fragment, amplified to incorporate *EcoRI* and *HindIII* restriction sites. Numbering is relative to the predicted transcript start site from the P1 promoter, denoted by +1. Locations of the P1 and P2 promoter -10 elements are coloured purple and blue, respectively; the CRP binding site is shown in orange; and substitutions labelled accordingly, -13G indicates the base at position -13 is changed to a G. (b) Mutational analysis confirmed the presence of two *kpsM* promoters.  $\beta$ -Galactosidase activity was measured in *E. coli* K12  $\Delta lac$  M182 and M182  $\Delta crp$  strains carrying a reporter plasmid pRW224 containing either the starting *kpsM* fragment (*pkpsM\_440*) cloned upstream of *lacZ* or the *kpsM* fragment with a -13G mutation in the P1 -10 element (P1-), 245G in P2 (P2-) or -13G 245G in both (P1-P2-). pRW224 with no insert was included as a negative control.  $\beta$ -Galactosidase activity is expressed as Miller units (nmol of ONPG hydrolysed min<sup>-1</sup> mg<sup>-1</sup> dry cell mass). Data are from duplicate independent experiments, averaged from three biological replicates, with SD shown. (c) *kpsM* P1 and P2 transcripts were detected by multi-round *in vitro* transcription from the *pkpsM\_440* DNA construct, cloned into the pSR plasmid, with *E. coli* holo-RNA polymerase (123 nM). The 110 nt RNAI transcript is an internal positive control. *kpsM* P1, P2 and RNAI transcripts are indicated by black arrows. A Maxim-Gilbert GA sequencing reaction was used to calibrate the transcripts. CRP concentrations: lane 1 (0  $\mu$ M), lane 2 (0.1  $\mu$ M), lane 3 (0.2  $\mu$ M), lane 4 (0.4  $\mu$ M), lane 5 (0.6  $\mu$ M), lane 6 (0.8  $\mu$ M), lane 7 (1  $\mu$ M) and lane 8 (2  $\mu$ M).

increases in response to stress [17], drives expression of a mucinase, the plasmid-encoded toxin and the microcin-siderophore, all of which are useful during the colonization stage of infection and, subsequently, at the exit stage of infection. In addition, CRP suppresses the expression of an adhesin and the production of capsule. By exploiting CRP, EAEC aligns the infection with nutrient availability. From a mechanistic point of view, the two CRP-suppressed promoters are remarkable. Most bacterial repressors act by blocking the access of RNA polymerase to the target promoter. It is significant that, at the EAEC 0536 and *kpsM* P2 promoters, evolution has driven the DNA site for CRP to include the promoter −35 and −10 hexamer elements, respectively. Recall that, at most bacterial promoters, these are the two most important elements for promoter function [30], and yet, this arrangement is uncommon at CRP-repressed promoters.

Although most textbooks consider the *E. coli* CRP protein to be a gene regulatory transcription factor, at many targets, its binding has little or no direct consequence for gene expression [54]. This was first discovered following the application of chromatin immunoprecipitation, which reports factor binding independent of any function, and this study with the EAEC strain 042 tells the same story as the original chromatin immunoprecipitation study on *E. coli* K-12 [14, 29]. A plausible explanation for this is that the original function of ancestral CRP had little or nothing to do with transcription regulation but, rather, was concerned with the management of bacterial chromosomes [54]. It has been argued that gene regulatory functions evolved later, and this may explain the variety of architectures found at different CRP-regulated promoters [54, 55]. Similar observations have been made with other bacterial transcription factors [56–58], and it is possible to argue that these function-less bound factors can, given certain constraints, evolve to participate in functional gene regulatory regions. Given this flexibility, it is easy to imagine that incoming DNA segments, which encode genetic determinants that are ‘useful’ to a bacterial host, could become subject to regulation by CRP. Our data suggest that whilst CRP-dependent function may have resulted at some loci, there are still others where CRP binding has no functional outcome. However, we can discern how CRP imposes a layer of regulation on the expression of virulence determinants, and this may, in the future, prompt new strategies for interfering with virulence gene expression.

#### Funding information

This work was supported by the Biotechnology and Biological Sciences Research Council (BB/R017689/1 and BB/W00285X/1) and Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R898), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. D.C.G. is grateful for funding from a Wellcome Trust Investigator Award (212193/Z/18/Z).

#### Conflicts of interest

The authors declare that there are no conflicts of interest.

#### References

- Wilson A, Evans J, Chart H, Cheasty T, Wheeler JG, et al. Characterisation of strains of enteroaggregative *Escherichia coli* isolated during the infectious intestinal disease study in England. *Eur J Epidemiol* 2001;17:1125–1130.
- Okeke IN, Lamikanra A, Czczulin J, Dubovsky F, Kaper JB, et al. Heterogeneous virulence of enteroaggregative *Escherichia coli* strains isolated from children in Southwest Nigeria. *J Infect Dis* 2000;181:252–260.
- França FLS, Wells TJ, Browning DF, Nogueira RT, Sarges FS, et al. Genotypic and phenotypic characterisation of enteroaggregative *Escherichia coli* from children in Rio de Janeiro, Brazil. *PLoS One* 2013;8:e69971.
- Nataro JP, Kaper JB, Robins-Browne R, Prado V, Vial P, et al. Patterns of adherence of diarrheagenic *Escherichia coli* to HEp-2 cells. *Pediatr Infect Dis J* 1987;6:829–831.
- Boisen N, Østerlund MT, Joensen KG, Santiago AE, Mandomando I, et al. Redefining enteroaggregative *Escherichia coli* (EAEC): genomic characterization of epidemiological EAEC strains. *PLoS Negl Trop Dis* 2020;14:e0008613.
- Nataro JP, Deng Y, Cookson S, Cravioto A, Savarino SJ, et al. Heterogeneity of enteroaggregative *Escherichia coli* virulence demonstrated in volunteers. *J Infect Dis* 1995;171:465–468.
- Chaudhuri RR, Sebaihia M, Hobman JL, Webber MA, Leyton DL, et al. Complete genome sequence and comparative metabolic profiling of the prototypical enteroaggregative *Escherichia coli* strain 042. *PLoS One* 2010;5:e8801.
- Hüttner M, Prieto A, Espelt J, Bernabeu M, Juárez A. Stringent response and AggR-dependent virulence regulation in the enteroaggregative *Escherichia coli* strain 042. *Front Microbiol* 2018;9:717.
- Morin N, Santiago AE, Ernst RK, Guillot SJ, Nataro JP. Characterization of the AggR regulon in enteroaggregative *Escherichia coli*. *Infect Immun* 2013;81:122–132.
- Yasir M, Icke C, Abdelwahab R, Haycocks JR, Godfrey RE, et al. Organization and architecture of AggR-dependent promoters from enteroaggregative *Escherichia coli*. *Mol Microbiol* 2019;111:534–551.
- Abdelwahab R, Yasir M, Godfrey RE, Christie GS, Element SJ, et al. Antimicrobial resistance and gene regulation in enteroaggregative *Escherichia coli* from Egyptian children with diarrhoea: similarities and differences. *Virulence* 2020;12:57–74.
- Rossiter AE, Godfrey RE, Connolly JA, Busby SJW, Henderson IR, et al. Expression of different bacterial cytotoxins is controlled by two global transcription factors, CRP and Fis, that co-operate in a shared-recruitment mechanism. *Biochem J* 2015;466:323–335.
- Rossiter AE, Browning DF, Leyton DL, Johnson MD, Godfrey RE, et al. 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* 2011;81:179–191.
- 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 USA* 2005;102:17693–17698.
- Pérez-Rueda E, Collado-Vides J. The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. *Nucleic Acids Res* 2000;28:1838–1847.
- Botsford JL, Harman JG. Cyclic AMP in prokaryotes. *Microbiol Rev* 1992;56:100–122.
- Görke B, Stülke J. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Microbiol* 2008;6:613–624.
- Alhammadi MM, Godfrey RE, Ingram JO, Singh G, Bathurst CL, et al. Novel organisation and regulation of the *pic* promoter from



- enteroaggregative and uropathogenic *Escherichia coli*. *Virulence* 2022;13:1393–1406.
19. Harrington SM, Sheikh J, Henderson IR, Ruiz-Perez F, Cohen PS, et al. The Pic protease of enteroaggregative *Escherichia coli* promotes intestinal colonization and growth in the presence of mucin. *Infect Immun* 2009;77:2465–2473.
20. Middlemiss AD, Haycocks JRJ, Stringer AM, Piddock LJV, Wade JT, et al. Mapping direct and indirect MarA/SoxS/Rob/RamA regulons in *Salmonella* Typhimurium reveals repression of *csgD* and biofilm formation. *Microbiology* 2023;169:001330.
21. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357–359.
22. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 2008;9:R137.
23. Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 2015;31:2382–2383.
24. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME suite. *Nucleic Acids Res* 2015;43:W39–49.
25. Miller JH. *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory, 1972, p. 494.
26. 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.
27. Ghosaini LR, Brown AM, Sturtevant JM. Scanning calorimetric study of the thermal unfolding of catabolite activator protein from *Escherichia coli* in the absence and presence of cyclic mononucleotides. *Biochemistry* 1988;27:5257–5261.
28. Salgado H, Gama-Castro S, Lara P, Mejia-Almonte C, Alarcón-Carranza G, et al. RegulonDB v12.0: a comprehensive resource of transcriptional regulation in *E. coli* K-12. *Nucleic Acids Res* 2024;52:D255–D264.
29. Hollands K, Busby SJW, Lloyd GS. New targets for the cyclic AMP receptor protein in the *Escherichia coli* K-12 genome. *FEMS Microbiol Lett* 2007;274:89–94.
30. Kolb A, Busby S, Buc H, Garges S, Adhya S. Transcriptional regulation by cAMP and its receptor protein. *Annu Rev Biochem* 1993;62:749–795.
31. Haycocks JRJ, Sharma P, Stringer AM, Wade JT, Grainger DC. The molecular basis for control of ETEC enterotoxin expression in response to environment and host. *PLoS Pathog* 2015;11:e1004605.
32. Gaston K, Bell A, Kolb A, Buc H, Busby S. Stringent spacing requirements for transcription activation by CRP. *Cell* 1990;62:733–743.
33. Vassiliadis G, Destoumieux-Garzón D, Lombard C, Rebuffat S, Peduzzi J. Isolation and characterization of two members of the siderophore-microcin family, microcins M and H47. *Antimicrob Agents Chemother* 2010;54:288–297.
34. Rodríguez E, Gaggero C, Laviña M. The structural gene for microcin H47 encodes a peptide precursor with antibiotic activity. *Antimicrob Agents Chemother* 1999;43:2176–2182.
35. Azpiroz MF, Laviña M. Involvement of enterobactin synthesis pathway in production of microcin H47. *Antimicrob Agents Chemother* 2004;48:1235–1241.
36. Patzer SI, Baquero MR, Bravo D, Moreno F, Hantke K. The colicin G, H and X determinants encode microcins M and H47, which might utilize the catecholate siderophore receptors FepA, Cir, Fiu and IroN. *Microbiology* 2003;149:2557–2570.
37. Sassone-Corsi M, Nuccio S-P, Liu H, Hernandez D, Vu CT, et al. Microcins mediate competition among *Enterobacteriaceae* in the inflamed gut. *Nature* 2016;540:280–283.
38. Busby SJW, Browning DF. Transcription activation in *Escherichia coli* and *Salmonella*. *EcoSal Plus* 2024;eesp00392020.
39. Feklistov A, Sharon BD, Darst SA, Gross CA. Bacterial sigma factors: a historical, structural, and genomic perspective. *Annu Rev Microbiol* 2014;68:357–376.
40. Busby S, Ebright RH. Transcription activation by catabolite activator protein (CAP). *J Mol Biol* 1999;293:199–213.
41. Roberts IS. The biochemistry and genetics of capsular polysaccharide production in bacteria. *Annu Rev Microbiol* 1996;50:285–315.
42. Cross AS, Genski P, Sadoff JC, Orskov F, Orskov I. The importance of the K1 capsule in invasive infections caused by *Escherichia coli*. *J Infect Dis* 1984;149:184–193.
43. Arredondo-Alonso S, Blundell-Hunter G, Fu Z, Gladstone RA, Fillol-Salom A, et al. Evolutionary and functional history of the *Escherichia coli* K1 capsule. *Nat Commun* 2023;14:3294.
44. Aldawood E, Roberts IS. Regulation of *Escherichia coli* group 2 capsule gene expression: a mini review and update. *Front Microbiol* 2022;13:858767.
45. Jia J, King JE, Goldrick MC, Aldawood E, Roberts IS. Three tandem promoters, together with IHF, regulate growth phase dependent expression of the *Escherichia coli* kps capsule gene cluster. *Sci Rep* 2017;7:17924.
46. Zubay G, Schwartz D, Beckwith J. Mechanism of activation of catabolite-sensitive genes: a positive control system. *Proc Natl Acad Sci USA* 1970;66:104–110.
47. de Crombrughe B, Chen B, Anderson WB, Gottesman ME, Perlman RL, et al. Role of cyclic adenosine 3',5'-monophosphate and the cyclic adenosine 3',5'-monophosphate receptor protein in the initiation of lac transcription. *J Biol Chem* 1971;246:7343–7348.
48. Kolb A, Spassky A, Chapon C, Blazy B, Buc H. On the different binding affinities of CRP at the lac, gal and malT promoter regions. *Nucleic Acids Res* 1983;11:7833–7852.
49. Passner JM, Schultz SC, Steitz TA. Modeling the cAMP-induced allosteric transition using the crystal structure of CAP-cAMP at 2.1 Å resolution. *J Mol Biol* 2000;304:847–859.
50. Popovych N, Tzeng S-R, Tonelli M, Ebright RH, Kalodimos CG. Structural basis for cAMP-mediated allosteric control of the catabolite activator protein. *Proc Natl Acad Sci USA* 2009;106:6927–6932.
51. Gosset G, Zhang Z, Nayyar S, Cuevas WA, Saier MH. Transcriptome analysis of Crp-dependent catabolite control of gene expression in *Escherichia coli*. *J Bacteriol* 2004;186:3516–3524.
52. Zheng D, Constantinidou C, Hobman JL, Minchin SD. Identification of the CRP regulon using *in vitro* and *in vivo* transcriptional profiling. *Nucleic Acids Res* 2004;32:5874–5893.
53. Shimada T, Fujita N, Yamamoto K, Ishihama A. Novel roles of cAMP receptor protein (CRP) in regulation of transport and metabolism of carbon sources. *PLoS One* 2011;6:e20081.
54. Visweswariah SS, Busby SJW. Evolution of bacterial transcription factors: how proteins take on new tasks, but do not always stop doing the old ones. *Trends Microbiol* 2015;23:463–467.
55. Busby SJW. Transcription activation in bacteria: ancient and modern. *Microbiology* 2019;165:386–395.
56. Dorman CJ, Schumacher MA, Bush MJ, Brennan RG, Buttner MJ. When is a transcription factor a NAP? *Curr Opin Microbiol* 2020;55:26–33.
57. Fitzgerald DM, Stringer AM, Smith C, Lapierre P, Wade JT. Genome-wide mapping of the *Escherichia coli* PhoB regulon reveals many transcriptionally inert, intragenic binding sites. *mBio* 2023;14:e0253522.
58. 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.

Edited by: S. Gebhard and M. Van Der Woude