Genome-wide mapping of cyclic AMP receptor protein binding in Enteroaggregative *Escherichia coli* reveals targeting of virulence-associated genes

Munirah M. Alhammadi[1,2], Joanne Hothersall[1], Georgina S. Lloyd[1], Sophie V. Titman[1,3], Thomas Guest[1,4], Douglas F. Browning[1,5], David C. Grainger[1], Stephen J.W. Busby[1] and James R.J. Haycocks[1,6]*.

[1] Institute of Microbiology and Infection, School of Biosciences, College of Life and Environmental Sciences, University of Birmingham, Birmingham, B15 2TT, UK.

[2] Current address: Department of Biology, College of Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia.

[3] Current address: Plant Sciences and the Bioeconomy, Rothamstead Research, Harpenden, AL5 2JQ, UK.

[4] Current address: Department of Microbiology, Biocenter, University of Würzburg, Würzburg, D-97074, Germany.

[5] Current address: School of Biosciences, College of Health & Life Sciences, Aston University, Birmingham, B4 7ET, UK.

[6] Current address: Newcastle University Biosciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK.

*For correspondence:

Email: james.haycocks@newcastle.ac.uk

Keywords: CRP, enteroaggregative *Escherichia coli* (EAEC), bacterial gene regulation, transcription, virulence.

ABSTRACT

Bacterial pathogens use a wide array of virulence factors to colonise 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 colonises 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-characterised bacterial global transcription factor, focusing on its role in 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 adhesion, production of a bacterial antibiotic, and export of a polysaccharide capsule.

INTRODUCTION

Enteroaggregative *Escherichia coli* (EAEC) is a common cause of diarrhoea in adults and children in both industrialised 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 *Escherichia 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, that is often adopted as a paradigm (7), contains the pAA2 plasmid that encodes a transcription activator, AggR, which has been found to be the 'master' regulator of virulence (8,9). Transcript initiation at many promoters that control expression of 042 virulence genes, both on the pAA2 plasmid and on 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 organisation 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 cyclic AMP receptor protein (CRP) (12,13). Recall that *Escherichia 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' cyclic adenosine monophosphate (cAMP) which, in general, in enteric bacteria, rises in response to certain stresses, such as starvation (16,17).

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

80 METHODS AND MATERIALS

81 *Strains, plasmids and oligonucleotides.*

82 Strains, plasmids and oligonucleotide primers used are listed in Table S1. ChIP-seq was exploited to
83 identify gene regulatory regions in EAEC 042 that are targets for CRP. Using PCR and primers listed
84 in Table S1, these regions were amplified and flanked with *Eco*RI and *Hind*III to facilitate cloning into
85 the *lac* expression vectors, pRW224 or pRW50, which were then used to measure promoter activities
86 in strains M182 and M182Δ*crp*, exactly as in our previous studies (12,13). Differences found in
87 M182Δ*crp* were complemented by plasmid pDCRP and derivatives. For *in vitro* studies, *EcoR*I-*Hind*III
88 fragments carrying different gene regulatory regions were cloned into the pSR 'holding' plasmid. For
89 the purposes of plasmid maintenance, strains harbouring pSR and pDCRP were grown in media
90 supplemented with 100 µg/ml ampicillin, for strains harbouring pRW224 and pRW50 media contained
91 15 µg/ml tetracycline.

92 *DNA constructs*

93 DNA fragments used for EMSA experiments were generated by PCR using *Escherichia coli* 042 gDNA
94 as template using primers listed in Table S1. The p*mchA* construct was generated by cloning the 619 bp
95 p*mchA*_600 construct *Eco*RI/*Hind*III into pRW50, which served as an intermediate template for

96 generation of the shorter p*mchA* construct by site-directed mutagenesis (SDM), using primers
97 p*mchA*_200 SDM F/R. For cloning *Eco*RI/*Hind*III into pSR, primers p*mchA*_200 F and p*mchA* R were
98 used to amplify the 213 bp p*mchA* fragment. The p*mchA* p37G, p11G and p11A mutations were created
99 by SDM, using the p*mchA* construct cloned into either pRW50 or pSR as a template. Promoter
100 fragments sub-cloned between pRW50 and pSR were amplified with primers p*mchA* F/p*mchA* R and
101 cloned *Eco*RI/*Hind*III. To make the p*mchA* 14A mutant, two separate "arms" were generated (using
102 p*mchA* SOE 14A left arm F/R and p*mchA* SOE 14A right arm F/R primers), which were spliced together
103 by overlap extension PCR to generate the final product. This was then cloned *Eco*RI/*Hind*III into
104 pRW50.

105 Mutations to the 042 p0536 promoter construct were introduced using SDM, except for the p0536 38T
106 36A construct, which was generated by PCR using overlapping oligos, then cloned *Eco*RI/*Hind*III into
107 pRW50. The p*kpsM*_201 DNA fragment was generated using p*kpsM*_201 F/R primers. p*kpsM*_440 was
108 constructed using primers p*kpsMII*_440 F/R, with mutations being introduced using p*kpsM* P1- F and
109 p*kpsM* P2- R.

110 All plasmid constructs were confirmed by Sanger DNA base sequencing.

111 *ChIP-seq*

112 Chromatin immunoprecipitation experiments were carried out in duplicate, using the method of
113 Middlemiss *et al.*(20) with modifications. EAEC 042 overnight cultures were sub-cultured into 40 ml
114 of fresh LB media and grown to mid-log phase (OD_{600}= 0.4-0.5), and were crosslinked in 1 % *v/v*
115 formaldehyde for 20 minutes. The reaction was quenched by adding glycine to 0.5 M, and cells were
116 washed in an equal volume of cold 1x TBS. Lysis was carried out by re-suspending cell pellets in 1 ml
117 immunoprecipitation buffer (consisting of 50 mM HEPES-KOH pH 7.5, 150 mM NaCl, 1 mM EDTA,
118 1 % Triton-X 100, 0.1 % sodium deoxycholate, with 1x EDTA-free protease inhibitor cocktail (Roche)
119 added per 50 ml) containing 2 mg mL^{-1} lysozyme at 37°C for 30 minutes. Samples were chilled on ice,
120 prior to sonication using a Bioruptor plus instrument (using the high intensity setting, for 30 cycles, 30
121 secs on, 30 secs off). Lysates were clarified by centrifugation and immunoprecipitation was carried out

122 by adding anti-CRP antibody (clone 1D8D9, purchased from Biolegend), with overnight incubation at
123 4ºC on a rotating wheel. Immunocomplexes were immobilised by adding 15 µl each of washed Protein
124 A and Protein G Dynabeads (Invitrogen) for 1 hour at 4ºC on a rotating wheel. Samples were then
125 washed 6 times with immunoprecipitation buffer, once with high salt immunoprecipitation buffer
126 (containing 500 mM NaCl), once with ChIP wash buffer (20 mM Tris HCl pH 8.0, 250 mM LiCl, 0.5 %
127 IGEPAL CA-630, 0.5 % sodium deoxycholate and 1 mM EDTA), and once with TE buffer. DNA was
128 eluted in buffer containing 50 mM Tris-HCl pH 7.5, 10 mM EDTA, and 1 % SDS. To decrosslink, 5 µl
129 of 10 mg mL^{-1} proteinase K (Sigma Aldrich) was added to the eluted DNA, and samples were incubated
130 at 42ºC for 1 hour, followed by incubation at 65ºC for 5 hours. DNA was purified using a Qiagen PC
131 purification kit, and libraries were prepared using an NEBnext Ultra II DNA library preparation kit.
132 Library size distributions were checked using an Agilent Tapestation instrument (using High-sensitivity
133 DNA Screentapes), and were quantified by qPCR using an NEBnext library quantification kit. Libraries
134 were pooled, and sequenced by Azenta Life Sciences (Next Generation Sequencing Services) on a
135 MiSeq instrument (paired end, 2 x 150 bp configuration). Sequencing data have been submitted to
136 ArrayExpress and is available under accession code E-MTAB-14748.

137 *Bioinformatics*

138 Reads were aligned to the EAEC 042 chromosome and pAA2 plasmid (references genome accession
139 numbers NC_017626.1 and NC_017627.1 respectively) using Bowtie2 (21). Following alignment of
140 Illumina reads to the reference genomes, MACS2 software (22) was used to identify peaks enriched in
141 immunoprecipitated samples, compared to the mock controls (a cut-off value of q= 0.05 was used), and
142 called peaks were visually inspected using the Artemis genome browser. The R package ChIPseeker
143 was used to identify the nearest open reading frame to peak centres (23). To identify a binding motif,
144 150 bp of the DNA sequence flanking each peak centre was extracted and submitted to MEME suite
145 using the MEME tool (24).

146

147

148 *β-galactosidase assays*

149 β-galactosidase assays were done using the method of Miller (25). Overnight cultures of strain M182
150 (and the Δ*crp* derivative) were sub-cultured into fresh LB (with appropriate antibiotics) and grown to
151 mid log phase (OD_{600}= 0.3-0.6) aerobically with shaking at 37ºC.

152 *Electrophoretic mobility shift assays (EMSAs)*

153 DNA fragments were generated by excision from a verified plasmid construct, using *Eco*RI/*AatII* and
154 *Hind*III restriction enzymes. Each fragment was end-labelled using ^{32}P-γ-ATP (purchased from
155 Hartmann Analytic GmbH) and T4 polynucleotide kinase (NEB). EMSA reactions were carried out as
156 previously described (18). For each reaction, 0.5 ng of the ^{32}P-radiolabelled DNA fragment was
157 incubated with or without CRP in buffer containing 20 mM 4-(2-hydroxyethyl)-1-piperazine ethane
158 sulphonic acid (HEPES) (pH 8.0), 5 mM MgCl2, 50 mM potassium glutamate, 1 mM dithiothreitol
159 (DTT), 0.2 mM cAMP, 5 % (v/v) glycerol, 25 µg mL^{-1} herring-sperm DNA and 0.5 mg mL^{-1} bovine
160 serum albumin. Reactions were incubated at 37ºC for 30 mins before being run on a 6 % polyacrylamide
161 gel that was visualised using a Bio-Rad PMI Personal Molecular Imager.

162 *Potassium permanganate footprinting*

163 Potassium permanganate footprinting was carried out as previously described (18). DNA carrying the
164 promoter was purified from *Hind*III/*Aat*II digested pSR, containing the cloned *Eco*RI/*Hind*III fragment,
165 and labelled at the *Hind*III end with γ^{32}P-ATP. Each reaction (20 µL) contained c. 3 nM labelled DNA
166 fragment in 20 mM HEPES (pH 8.0), 5 mM $MgCl_2$, 50 mM potassium glutamate, 1 mM dithiothreitol
167 (DTT) and 0.5 mg mL^{-1} bovine serum albumin, and 0.2 mM cyclic AMP, 100 nM CRP and 50 nM RNA
168 polymerase as required. Footprinting reactions were incubated at 37°C for 30 min, and then treated with
169 200 mM potassium permanganate for 4 min. Reactions were then stopped with 50 µL of stop solution
170 (3 M ammonium acetate, 0.1 M EDTA pH 8.0, 1.5 M β-mercaptoethanol). Following phenol–
171 chloroform extraction and ethanol precipitation, samples were resuspended in 40 µl 1 M piperidine and
172 incubated for 30 min at 90°C. Samples were again purified by phenol–chloroform extraction and
173 ethanol precipitation then resuspended in 8 µl loading buffer (95% formamide, 20 mM EDTA pH 8.0,

174 0.025 % bromophenol blue, 0.025 % xylene cyanol FF). Samples were analysed by 6% denaturing gel
175 electrophoresis. Gels were calibrated with a Maxam–Gilbert 'GA' sequencing reaction of the labelled
176 fragment and visualised with an Amersham Typhoon 5 biomolecular imager.

177 *In vitro transcription*

178 *In vitro* transcription reactions were carried out using the method of Kolb et al.(26). Template DNA
179 (used at a concentration of 16 μg mL^{-1}) was incubated with or without CRP in buffer containing 40 mM
180 Tris-Acetate pH 7.9, 10 mM $MgCl_2$, 1 mM DTT, 100 mM KCl, 100 μg mL^{-1} BSA, 200 μM
181 ATP/CTP/GTP, 10 μM and 5 μCi α^{32}P-UTP for ten minutes at 37°C. Reactions containing CRP were
182 supplemented with 0.2 mM cAMP. To start the reaction, σ^{70}-containing RNA polymerase holoenzyme
183 was added at a concentration of 50 nM for ten minutes at 37°C. Reactions were stopped by the addition
184 of stop solution (95 % formamide, 10 mM EDTA pH 8.0, 0.025 % bromophenol blue and 0.025 %
185 xylene cyanol FF). Reactions products were separated on a 6 % denaturing acrylamide gel, and were
186 visualised using a Bio-Rad PMI Personal Molecular Imager.

187 *Proteins*

188 CRP was purified using the Ghosaini method (27). RNA polymerase holoenzyme was purchased from
189 NEB.

190 RESULTS AND DISCUSSION

191 *ChIP analysis of CRP binding targets in EAEC 042*

192 RegulonDB, the leading transcription regulation database for *Escherichia coli* lists over 300
193 transcription units where CRP is involved in regulation (28). In most cases, CRP binds to one or more
194 specific target sequences at the regulatory region of the unit, and, alone, or in combination with other
195 transcription factors, either activates or represses transcript initiation. In previous work, working with
196 laboratory strains of *E. coli*, ChIP was applied directly to measure CRP binding, independent of any
197 regulatory function (14). These studies confirmed hundreds of targets but, surprisingly, also identified
198 targets outside of gene regulatory regions where CRP bound, apparently, with no ensuing effects on

199 transcription (29). Prompted, by our observation that the promoters of the EAEC *pet* and *pic* genes are
200 dependent on CRP for transcript initiation, we applied ChIP to map sites at which CRP is bound in
201 EAEC 042 growing in mid-log phase in LB media. Figure 1a illustrates the distribution of the 322 sites
202 on the EAEC 042 chromosome and the 10 sites on the virulence plasmid: these are listed in Table S2,
203 along with their location with respect to the nearest coding region. We used MEME to deduce a CRP-
204 binding logo from DNA base sequences centred on each peak (Figure 1b). As expected, this motif
205 consists of two copies of the 5-base pair element 5`-TGTGA-3`, organised as an inverted repeat,
206 separated by 6 base pairs. Note that this corresponds with previously derived logos for CRP (30).
207 Inspection of the location of these targets showed that, as in previous work, nearly 1 in 5 of the DNA
208 sites for CRP is at an intragenic location and is unlikely directly to be associated with any transcriptional
209 regulation (Figure 1c)(14,31) .

210 In order to minimise false positives, we applied a cut-off to our data for targets assigned a score (during
211 peak calling) of 100 or more: this reduced the number of targets to 150 on the EAEC chromosome and
212 5 on the plasmid. Inspection of these revealed that 9 are located in chromosome segments that are absent
213 from the reference MG1655 genome, whilst 117 located to regulatory regions that have been annotated
214 in MG1655, with CRP identified as a regulatory factor. For each of these 117 targets, we compared the
215 MG1655 base sequence from position -80 to +20 with respect to the transcript start (according to
216 RegulonDB) with the EAEC 042 sequence (see Table S3). Identical sequences were found at 55 targets
217 and only 3 targets showed more than 3 base differences. We found single base insertions, upstream of
218 the promoter -35 region in the *serC* and *pckA* regulatory regions.

219 *CRP interactions at uncommon EAEC 042 targets*

220 Table 1 lists the 9 EAEC 042 chromosomal targets not found in MG1655, together with the 5 targets
221 on the plasmid. To validate these targets, we used electromobility shift assays (EMSA) using purified
222 CRP and ^{32}P-end labelled fragments covering each regulatory region (generated by PCR, see Table S1).
223 We did not test the 2670 and *set1A* targets as they appeared to be outside of any regulatory region.
224 Similarly, with the plasmid-borne targets, we ignored RS26580 and the two *repA* targets as they appear
225 to be connected with core plasmid functions, and the DNA site for CRP in the *pet* regulatory region has

226 been already exhaustively characterised (13). Figure 2 shows the bandshift experiment with the
227 remaining 8 targets. The data show a strong CRP interaction with 6 targets, but not with 3975 or *virK*.
228 Inspection of the chromosomal context of each of the strong binding sites suggested that the EAEC 042
229 *kpsM*, *mchA*, 0536 and 0225 targets were most likely to be associated with a promoter-active regulatory
230 region. Hence DNA fragments carrying each region were cloned into plasmid pRW50, a low copy
231 number *lac* expression vector, each recombinant plasmid was used to transform the *Escherichia coli*
232 K-12 Δ*lac* strains, M182 and M182Δ*crp*, and β-galactosidase activities were measured. These activities
233 are taken as a measure of promoter activity and comparison of levels in the crp^+ and Δ*crp* strains give
234 an indication of any CRP-dependent regulation. As a control, an assay with pRW50 carrying the semi-
235 synthetic CC(-41.5) promoter, whose activity is tightly coupled to CRP (32), is included. The data,
236 illustrated in Figure 3, show that the p*kpsM*_201, p*mchA*, p0536 fragments carry at least one promoter.
237 The observed activity with *mchA* is clearly CRP-dependent, whilst, with the p*kpsM*_201 and p0536
238 fragments, activity is higher in the Δ*crp* strain, suggesting repression by CRP. These effects were
239 investigated further and our experiments are described in the next sections. Whilst there is no evidence
240 for any promoter activity associated with the 0225 target, we were concerned that CRP might be
241 regulating expression of the divergent gene, 0224 and so the cloned fragment was inverted in pRW50.
242 However, measured expression levels were very low (data not shown), and we conclude that the
243 intergenic region between 0224 and 0225 may be bereft of promoter activity (at least in our
244 experimental conditions).

245 *CRP-dependent activation at the mchA promoter*

246 The EAEC 042 *mchA* gene encodes an enzyme that transfers a glucose residue to microcin H47, a well-
247 characterised bacterial antibiotic (33–36). The significance of the glucose residue is that it serves as a
248 bridge to link a siderophore molecule, thereby creating a siderophore-toxin which can play an important
249 role in bacterial colonisation and, hence, virulence (37). Data in Figure 3 suggest that *mchA* expression
250 may be due to a strong promoter that is dependent on CRP. Recall that most bacterial promoters carry
251 crucial sequence elements located ~10, 14 and 35 bases upstream of the transcript start site (known as
252 the -10, extended -10 and -35 elements, respectively), and that one or more of these is usually defective

253 at promoters dependent on an activator such as CRP (38,39). Moreover CRP-activated promoters fall
254 into two major classes: at Class I promoters the DNA site for CRP is located over 15 bases upstream
255 from the promoter -35 element whilst, for Class II promoters, it overlaps the -35 element (40).
256 Inspection of the base sequence adjacent the proposed DNA site for CRP at the *mchA* regulatory region
257 suggested the existence of a Class II CRP-dependent promoter (Figure 4a). On this basis we created the
258 37G and 14A mutations that would be expected to inactivate the DNA site for CRP and the extended -
259 10 element respectively. DNA fragments carrying these base substitutions were cloned into pRW50:
260 data in Figure 4b show that CRP-dependent activation is completely stopped by either mutation.
261 Furthermore, an 11G mutation within the predicted -10 hexamer reduced activity and an 11A mutation,
262 to make the base at the -11 position consensus, increased activity (Figure 4b). To confirm the existence
263 of a Class II CRP-activated promoter, we complemented the M182Δ*crp* strain with plasmid-borne *crp*
264 carrying substitutions in either or both of the major activating regions (known as AR1 and AR2). Data
265 in Figure 4c show that both AR1 and AR2 are needed for optimal complementation, with AR2 making
266 the larger contribution to activation. Since AR2 is known to play little or no role at Class I CRP-
267 activated promoters (40), taken together, the evidence suggests that EAEC 042 *mchA* expression
268 depends on a single Class II CRP-dependent promoter, and this was confirmed by an *in vitro*
269 transcription assay (Figure 4d).

270 *CRP-dependent repression at the 0536 promoter*

271 The EAEC 042 0536 gene appears likely to encode an adhesin (7), data in Figure 3 suggest that the
272 0536 gene regulatory region contains one or more promoters, and that CRP could down-regulate
273 expression from them. Inspection of the base sequence upstream of the 0536 translation start suggested
274 4 possible promoter -10 elements. Hence, starting with the cloned p0536 regulatory region fragment,
275 we constructed single base changes that would be expected to inactivate each of the putative -10
276 elements denoted 11G, 1G, 17G and 25G (Figure 5a). Assays of the promoter activity of each mutant
277 fragment showed that the 11G mutation alone caused complete loss of activity (Figure 5b), suggesting
278 that the 0536 regulatory region contains a single major promoter. This was corroborated by a
279 permanganate footprint and an *in vitro* transcription assay (Figure 5c & 5d). The transcription assay

280 shows an RNA that initiates 7 base pairs downstream from the suggested 5'-TAGATT-3' -10 element,
281 and the permanganate shows the expected open complex unwinding starting at the upstream end of the
282 -10 element. Both the footprint and the transcript are suppressed by the inclusion of CRP in the assay
283 (Figure 5c & 5d). Taken together, the data argue that the EAEC 042 0536 gene regulatory region
284 contains a single promoter whose activity is down-regulated by CRP. Inspection of the promoter base
285 sequence shows that the most likely -35 element is 5`-TTTACA-3`, located 17 base pairs upstream of
286 the -10 element, which corresponds to the optimal spacing. This is sandwiched between the two flanking
287 5-base pair elements that comprise the DNA site for CRP and this suggests a simple mechanism for the
288 repressive action of CRP. To confirm this, the 38T 36A double base substitution was constructed in the
289 cloned p0536 regulatory region fragment (Figure 5a). These substitutions, which create a consensus
290 upstream 5-base pair element in the DNA site for CRP, result in stronger repression of the 0536
291 promoter by CRP (Figure 5e).

292 *CRP-dependent repression at the kpsM regulatory region*

293 The *E. coli kps* gene products enable the production of a form of capsule known as K antigen (41–43).
294 Although absent in MG1655, the *kps* genes are found in many *E. coli* pathovars, and their expression
295 has been studied in some detail (44,45). The *kpsM* gene is unusual in that its promoter is located ~750
296 base pairs upstream from its start codon. Inspection of the regulatory region showed that the DNA site
297 for CRP is ~250 base pairs downstream from this promoter (Figure 6a, denoted here as P1), which is
298 consistent with the observed repression by CRP (Figure 3). However, we noted that the two consensus
299 halves of this DNA site for CRP flank a near perfect promoter -10 sequence (5`-TATAAA-3`),
300 suggesting the possible existence of a second promoter (denoted here as P2). To investigate this, we
301 cloned a 440 bp DNA fragment covering both the CRP binding site and the previously identified
302 promoter region (p*kpsM*_440, Figure 6a). We made single base changes in the p*kpsM*_440 fragment, in
303 which we corrupted the upstream promoter -10 element ($P1^-$), the downstream -10 element ($P2^-$), or
304 both ($P1^-P2^-$), and we measured the net promoter activity in our crp^+ and Δ*crp* strains as in Figure 3.
305 The data, illustrated in Figure 6b show that *kpsM* promoter activity is largely suppressed by the two
306 base changes together. Crucially, the fragment carrying the $P1^-$ mutation shows significant activity in

307 the Δ*crp* genetic background, but this activity is greatly suppressed in the *crp*$^{+}$ strain. These data are
308 consistent with a model in which the activity of the dominant upstream promoter (P1) is moderated by
309 CRP, but a second weaker downstream promoter (P2) is sharply repressed by CRP because its -10
310 hexamer element is embedded in the DNA site for CRP. This was corroborated by an *in vitro*
311 transcription assay (Figure 6c).

312 CONCLUSIONS

313 The starting point of this work was a ChIP-seq experiment, designed to identify the different DNA sites
314 for CRP on the chromosome and virulence plasmid of enteroaggregative *E. coli* strain 042. Arguably,
315 CRP is the most widely studied bacterial transcription factor. It was discovered originally as the
316 essential activator of *E. coli* lactose operon expression (46) and its activity was found to be triggered
317 by the binding of cyclic AMP (47–50). Subsequently it was shown to regulate 100s of *E. coli*
318 transcription units and its biological role appears to be to reprogramme gene transcription in response
319 to nutritional stresses (30,51–53). Our research, reported here, was prompted by the observation that
320 transcription of two key EAEC 042 virulence genes (*pet* and *pic*) directly depends on CRP (13,18),
321 rather than AggR, which is regarded as the master regulator of virulence in EAEC (9) . We wanted to
322 discover if the transcription of any other EAEC virulence genes was likewise activated by CRP. Our
323 experiments showed that, as expected, CRP binds to hundreds of gene regulatory regions and that the
324 vast majority of binding targets are common with the MG1655 *E. coli* reference strain. However, after
325 triage, we found CRP-dependent regulation at three genes that are not found in MG1655. In one case,
326 *mchA*, CRP is an activator, whilst in the other two cases, 0536 and *kpsM*, CRP downregulates gene
327 expression. We propose that, whilst EAEC virulence is primarily controlled by the activity of AggR,
328 CRP activity, which increases in response to stress (17), drives expression of a mucinase, the plasmid-
329 encoded toxin and the microcin-siderophore, all of which are useful during the colonisation stage of
330 infection, and, subsequently, at the exit stage of infection. In addition, CRP suppresses expression of an
331 adhesin and production of capsule. By exploiting CRP, EAEC aligns the infection with nutrient
332 availability. From a mechanistic point of view, the two CRP-suppressed promoters are remarkable.
333 Most bacterial repressors act by blocking the access of RNA polymerase to the target promoter. It is

334 significant that, at the EAEC 0536 and *kpsM* P2 promoters, evolution has driven the DNA site for CRP
335 to include the promoter -35 and -10 hexamer elements respectively. Recall that, at most bacterial
336 promoters, these are the two most important elements for promoter function (30), and yet this
337 arrangement is uncommon at CRP-repressed promoters.

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

355

356

357

358

359

REFERENCES

1. Wilson A, Evans J, Chart H, Cheasty T, Wheeler JG, Tompkins D, et al. Characterisation of strains of enteroaggregative Escherichia coli isolated during the infectious intestinal disease study in England. Eur J Epidemiol. 2001;17(12):1125–30.

2. Okeke IN, Lamikanra A, Czeczulin J, Dubovsky F, Kaper JB, Nataro JP. Heterogeneous virulence of enteroaggregative Escherichia coli strains isolated from children in Southwest Nigeria. J Infect Dis. 2000 Jan;181(1):252–60.

3. França FLS, Wells TJ, Browning DF, Nogueira RT, Sarges FS, Pereira AC, et al. Genotypic and phenotypic characterisation of enteroaggregative Escherichia coli from children in Rio de Janeiro, Brazil. PLoS One. 2013;8(7):e69971.

4. Nataro JP, Kaper JB, Robins-Browne R, Prado V, Vial P, Levine MM. Patterns of adherence of diarrheagenic Escherichia coli to HEp-2 cells. Pediatr Infect Dis J. 1987 Sep;6(9):829–31.

5. Boisen N, Østerlund MT, Joensen KG, Santiago AE, Mandomando I, Cravioto A, et al. Redefining enteroaggregative Escherichia coli (EAEC): Genomic characterization of epidemiological EAEC strains. PLoS Negl Trop Dis. 2020 Sep;14(9):e0008613.

6. Nataro JP, Deng Y, Cookson S, Cravioto A, Savarino SJ, Guers LD, et al. Heterogeneity of enteroaggregative Escherichia coli virulence demonstrated in volunteers. J Infect Dis. 1995 Feb;171(2):465–8.

7. Chaudhuri RR, Sebaihia M, Hobman JL, Webber MA, Leyton DL, Goldberg MD, et al. Complete Genome Sequence and Comparative Metabolic Profiling of the Prototypical Enteroaggregative Escherichia coli Strain 042. PLoS One. 2010 Jan 20;5(1):e8801.

8. Hüttener 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.

9. Morin N, Santiago AE, Ernst RK, Guillot SJ, Nataro JP. Characterization of the AggR regulon in enteroaggregative Escherichia coli. Infect Immun. 2013 Jan;81(1):122–32.

10. Yasir M, Icke C, Abdelwahab R, Haycocks JR, Godfrey RE, Sazinas P, et al. Organization and architecture of AggR-dependent promoters from enteroaggregative Escherichia coli. Mol Microbiol. 2019 Feb;111(2):534–51.

11. Abdelwahab R, Yasir M, Godfrey RE, Christie GS, Element SJ, Saville F, et al. Antimicrobial resistance and gene regulation in Enteroaggregative Escherichia coli from Egyptian children with diarrhoea: Similarities and differences. Virulence. 2020;12(1):57–74.

12. Rossiter AE, Godfrey RE, Connolly JA, Busby SJW, Henderson IR, Browning DF. 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 Mar 1;466(2):323–35.

13. Rossiter AE, Browning DF, Leyton DL, Johnson MD, Godfrey RE, Wardius CA, 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 Jul;81(1):179–91.

399 14. Grainger DC, Hurd D, Harrison M, Holdstock J, Busby SJW. Studies of the distribution of
400 Escherichia coli cAMP-receptor protein and RNA polymerase along the E. coli chromosome.
401 Proc Natl Acad Sci U S A. 2005 Dec 6;102(49):17693–8.

402 15. Pérez-Rueda E, Collado-Vides J. The repertoire of DNA-binding transcriptional regulators in
403 Escherichia coli K-12. Nucleic Acids Res. 2000 Apr 15;28(8):1838–47.

404 16. Botsford JL, Harman JG. Cyclic AMP in prokaryotes. Microbiol Rev. 1992 Mar;56(1):100–22.

405 17. Görke B, Stülke J. Carbon catabolite repression in bacteria: many ways to make the most out of
406 nutrients. Nat Rev Microbiol. 2008 Aug;6(8):613–24.

407 18. Alhammadi MM, Godfrey RE, Ingram JO, Singh G, Bathurst CL, Busby SJW, et al. Novel
408 organisation and regulation of the pic promoter from enteroaggregative and uropathogenic
409 Escherichia coli. Virulence. 2022 Dec;13(1):1393–406.

410 19. Harrington SM, Sheikh J, Henderson IR, Ruiz-Perez F, Cohen PS, Nataro JP. The Pic Protease of
411 Enteroaggregative Escherichia coli Promotes Intestinal Colonization and Growth in the Presence
412 of Mucin. Infection and Immunity. 2009 Jun;77(6):2465–73.

413 20. Middlemiss AD, Haycocks JRJ, Stringer AM, Piddock LJV, Wade JT, Grainger DC. Mapping
414 direct and indirect MarA/SoxS/Rob/RamA regulons in Salmonella Typhimurium reveals
415 repression of csgD and biofilm formation. Microbiology (Reading). 2023 May;169(5):001330.

416 21. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Mar
417 4;9(4):357–9.

418 22. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based Analysis
419 of ChIP-Seq (MACS). Genome Biology. 2008 Sep 17;9(9):R137.

420 23. Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation,
421 comparison and visualization. Bioinformatics. 2015 Jul 15;31(14):2382–3.

422 24. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. Nucleic Acids Res. 2015 Jul
423 1;43(W1):W39-49.

424 25. Miller JH. Experiments in Molecular Genetics. Cold Spring Harbor Laboratory; 1972. 494 p.

425 26. Kolb A, Kotlarz D, Kusano S, Ishihama A. Selectivity of the Escherichia coli RNA polymerase E
426 sigma 38 for overlapping promoters and ability to support CRP activation. Nucleic Acids Res.
427 1995 Mar 11;23(5):819–26.

428 27. Ghosaini LR, Brown AM, Sturtevant JM. Scanning calorimetric study of the thermal unfolding of
429 catabolite activator protein from Escherichia coli in the absence and presence of cyclic
430 mononucleotides. Biochemistry. 1988 Jul 12;27(14):5257–61.

431 28. Salgado H, Gama-Castro S, Lara P, Mejia-Almonte C, Alarcón-Carranza G, López-Almazo AG,
432 et al. RegulonDB v12.0: a comprehensive resource of transcriptional regulation in E. coli K-12.
433 Nucleic Acids Res. 2024 Jan 5;52(D1):D255–64.

434 29. Hollands K, Busby SJW, Lloyd GS. New targets for the cyclic AMP receptor protein in the
435 Escherichia coli K-12 genome. FEMS Microbiol Lett. 2007 Sep;274(1):89–94.

436 30. Kolb A, Busby S, Buc H, Garges S, Adhya S. Transcriptional regulation by cAMP and its
437 receptor protein. Annu Rev Biochem. 1993;62:749–95.

438 31. Haycocks JRJ, Sharma P, Stringer AM, Wade JT, Grainger DC. The molecular basis for control
439 of ETEC enterotoxin expression in response to environment and host. PLoS Pathog. 2015
440 Jan;11(1):e1004605.

441 32. Gaston K, Bell A, Kolb A, Buc H, Busby S. Stringent spacing requirements for transcription
442 activation by CRP. Cell. 1990 Aug 24;62(4):733–43.

443 33. Vassiliadis G, Destoumieux-Garzón D, Lombard C, Rebuffat S, Peduzzi J. Isolation and
444 characterization of two members of the siderophore-microcin family, microcins M and H47.
445 Antimicrob Agents Chemother. 2010 Jan;54(1):288–97.

446 34. Rodríguez E, Gaggero C, Laviña M. The structural gene for microcin H47 encodes a peptide
447 precursor with antibiotic activity. Antimicrob Agents Chemother. 1999 Sep;43(9):2176–82.

448 35. Azpiroz MF, Laviña M. Involvement of Enterobactin Synthesis Pathway in Production of
449 Microcin H47. Antimicrob Agents Chemother. 2004 Apr;48(4):1235–41.

450 36. Patzer SI, Baquero MR, Bravo D, Moreno F, Hantke K. The colicin G, H and X determinants
451 encode microcins M and H47, which might utilize the catecholate siderophore receptors FepA,
452 Cir, Fiu and IroN. Microbiology (Reading). 2003 Sep;149(Pt 9):2557–70.

453 37. Sassone-Corsi M, Nuccio SP, Liu H, Hernandez D, Vu CT, Takahashi AA, et al. Microcins
454 mediate competition among Enterobacteriaceae in the inflamed gut. Nature. 2016 Dec
455 8;540(7632):280–3.

456 38. Busby SJW, Browning DF. Transcription activation in Escherichia coli and Salmonella. EcoSal
457 Plus. 2024 Dec 12;12(1):eesp00392020.

458 39. Feklístov A, Sharon BD, Darst SA, Gross CA. Bacterial sigma factors: a historical, structural, and
459 genomic perspective. Annu Rev Microbiol. 2014;68:357–76.

460 40. Busby S, Ebright RH. Transcription activation by catabolite activator protein (CAP). J Mol Biol.
461 1999 Oct 22;293(2):199–213.

462 41. Roberts IS. The biochemistry and genetics of capsular polysaccharide production in bacteria.
463 Annu Rev Microbiol. 1996;50:285–315.

464 42. Cross AS, Gemski P, Sadoff JC, Orskov F, Orskov I. The importance of the K1 capsule in
465 invasive infections caused by Escherichia coli. J Infect Dis. 1984 Feb;149(2):184–93.

466 43. Arredondo-Alonso S, Blundell-Hunter G, Fu Z, Gladstone RA, Fillol-Salom A, Loraine J, et al.
467 Evolutionary and functional history of the Escherichia coli K1 capsule. Nat Commun. 2023 Jun
468 15;14:3294.

469 44. Aldawood E, Roberts IS. Regulation of Escherichia coli Group 2 Capsule Gene Expression: A
470 Mini Review and Update. Front Microbiol. 2022;13:858767.

471 45. Jia J, King JE, Goldrick MC, Aldawood E, Roberts IS. Three tandem promoters, together with
472 IHF, regulate growth phase dependent expression of the Escherichia coli kps capsule gene cluster.
473 Sci Rep. 2017 Dec 20;7(1):17924.

474 46. Zubay G, Schwartz D, Beckwith J. Mechanism of Activation of Catabolite-Sensitive Genes: A
475 Positive Control System*. Proceedings of the National Academy of Sciences. 1970
476 May;66(1):104–10.

47. Crombrugghe B de, Chen B, Anderson WB, Gottesman ME, Perlman RL, Pastan I. Role of Cyclic Adenosine 3′,5′-Monophosphate and the Cyclic Adenosine 3′,5′-Monophosphate Receptor Protein in the Initiation of lac Transcription. Journal of Biological Chemistry. 1971 Dec 1;246(23):7343–8.

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 Nov 25;11(22):7833–52.

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. Journal of Molecular Biology. 2000 Dec 15;304(5):847–59.

50. Popovych N, Tzeng SR, Tonelli M, Ebright RH, Kalodimos CG. Structural basis for cAMP-mediated allosteric control of the catabolite activator protein. Proc Natl Acad Sci U S A. 2009 Apr 28;106(17):6927–32.

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 Jun;186(11):3516–24.

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(19):5874–93.

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(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 in Microbiology. 2015 Aug 1;23(8):463–7.

55. Busby SJW. Transcription activation in bacteria: ancient and modern. Microbiology. 2019;165(4):386–95.

56. Dorman CJ, Schumacher MA, Bush MJ, Brennan RG, Buttner MJ. When is a transcription factor a NAP? Curr Opin Microbiol. 2020 Jun;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 Jun 27;14(3):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 Jul;36(12):3950–5.

ACKNOWLEDGEMENTS

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. DCG is grateful for funding from a Wellcome Trust Investigator Award (212193/Z/18/Z).

Conflicts of interest

The authors declare that no conflicts of interest exist.

530
531
532

Table 1: Uncommon targets for CRP in EAEC strain 042

Peak centre[1]	Score[2]	Associated Gene[3]	Distance to translation start[4]	Binding Site Sequence[5]	Associated Gene Product[6]
Targets located on chromosome					
3356318	6783	3143	+38.5	**TGTGA**TCTACA**ACACG**	Hypothetical protein
3459052	894	*kpsM*	-492.5	**TGTGA**TTTATA**TCACA**	Polysialic acid transport permease protein
3421201	615	*mchA*	-79.5	**TGTGA**GCTTAT**GCACA**	Microcin activation protein
461029	312	0414	+72.5	**TGTGC**GCAAGA**TCACA**	Conserved hypothetical protein
4217753	250	3975	-33.5	**CGTGT**ATACGA**TAACA**	Hypothetical protein
2833992	209	2670	+230.5	**TGTGA**TACACA**GCAAC**	Conserved hypothetical protein
4926886	174	*set1A*	+48.5	**ACTGA**CGGTTT**TCCCA**	Enterotoxin 1
608383	168	0536	-120.5	**TGTGA**TGTAAA**GCGCA**	Putative adhesin
263739	111	0225	-80.5	**TGTGA**GCCGCA**TCACA**	Putative type VI secretion system protein
Target located on plasmid pAA2					
27988	2975	*pet*	-117.5	**CGAGA**GCATTG**TCACA**	Serine protease autotransporter toxin Pet
62001	907	RS26580	-247.5	**TGTGA**CCATAT**TATCA**	TraS
18257	387	*virK*	+235.5	**TGTGG**TGACTG**GTACA**	Virulence protein
112767	321	*repA*	-84.5	**ACAGA**TCTTCG**TCACA**	Replication protein A
108769	293	*repA*	-84.5	**TGTGA**CGAAGA**TCTGT**	Replication protein A

533
534
535 **Experimentally determined DNA sites for CRP that are not found in *E. coli* K-12**.
536
537 Notes:
538 [1]Peak centre coordinate as defined by MACS2.
539 [2] Score representing the number of reads mapped to each position within the EAEC 042 genome.
540 [3] Adjacent annotated gene in EAEC 042.
541 [4] Distance from DNA site for CRP to translation start site of annotated gene ("-" indicates upstream, "+" indicates downstream).
542 [5] Base sequence of DNA site for CRP.
543 [6] Product of adjacent gene according to Chaudhuri *et al* (7).
544

545

546 **Figure Legends**

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

554 **Fig. 2.** Electrophoretic mobility shift assays of CRP binding to EAEC 042 targets. Eight 042 targets,
555 each approximately 200 bp in length, were excised from pSR as *Hind*III/*Eco*RI or *Hind*III/*Aat*III
556 fragments and radiolabelled with γ32-P ATP. These were incubated, in the presence of cAMP, with
557 increasing concentrations of CRP (0-1.6 μM) and then separated by electrophoresis on a 6%
558 polyacrylamide gel. In each case the band in the first lane shows the mobility of free DNA, and is
559 indicated by a grey arrow. In subsequent lanes, a reduced mobility band indicates a DNA fragment
560 complexed with CRP, marked with an orange arrow.

561 **Fig. 3.** CRP regulation at EAEC 042 promoters. The figure shows β-galactosidase activity measured in
562 *E. coli* K12 Δ*lac* M182 and M182 Δ*crp* strains carrying a *lacZ* reporter plasmid pRW50 containing a
563 target promoter region: p*mchA*; p0536; p*kpsM*_201; or p0225. Activities are expressed as fold activity
564 compared to the activity of pRW50 carrying the promoter, CC(-41.5) measured in M182 crp^+ cells.
565 pRW50 with no insert was included as a negative control. Data is a representative of duplicate
566 independent experiments, averaged from three biological replicates, with standard deviations shown.

567 **Fig. 4.** CRP-dependent activation of the EAEC 042 *mchA* promoter. (a) Sequence of the p*mchA*
568 promoter fragment, amplified to incorporate *Eco*RI and *Hind*III restriction sites. Numbering is relative
569 to the predicted transcript start site noted by +1. Locations of the -10 element and CRP binding site are
570 shown in bold purple and orange respectively; the extended -10 is underlined; and substitutions labelled
571 accordingly, for example 11G indicates the base at position -11 is changed to a G. (b) This panel shows
572 mutational analysis of the *mchA* promoter. β-galactosidase activity was measured in *E. coli* K12 Δ*lac*
573 strains M182 and M182 Δ*crp* carrying a reporter plasmid pRW50 containing either the starting *mchA*
574 promoter (p*mchA*) cloned upstream of *lacZ* or the *mchA* promoter with a substitution in the CRP binding
575 site (37G), in the extended -10 (14A) or the -10 element (11G or 11A). Promoter activities are expressed
576 as a percentage relative to the starting *mchA* promoter activity in M182 crp^+ cells. pRW50 with no insert
577 was included as a negative control. Data is from duplicate independent experiments, averaged from
578 three biological replicates, with standard deviations shown. (c) Complementation with CRP derivatives
579 demonstrated Class II activation of the *mchA* promoter. β-galactosidase activity was measured in M182
580 Δ*crp* cells containing the wild type *mchA* promoter *lacZ* fusion (in pRW50) with derivatives of CRP
581 expressed from pDCRP: wild type CRP (WT); CRP HL159 mutated in AR1 (AR1-); KE101 in AR2
582 (AR2-); and HL159 KE101 in both AR1 and AR2 (AR1- AR2-). pD with no CRP insert was included
583 as a negative control. β-Galactosidase activity is expressed as Miller Units (nmol of ONPG hydrolysed
584 min^{-1} mg^{-1} dry cell mass). Data is from duplicate independent experiments, averaged from three
585 biological replicates, with standard deviations shown. (d) A CRP-dependent transcript was detected by
586 multi round *in vitro* transcription from the *mchA* promoter, cloned in pSR plasmid, with *E. coli* holo-
587 RNA polymerase (50 nM). The 110 nt RNAI transcript is an internal positive control. Both the *mchA*
588 and RNAI transcripts are indicated by arrows. A Maxim-Gilbert GA sequencing reaction was used to
589 calibrate the transcripts. CRP concentrations: lane 1 (0 μM), lane 2 (25 nM), and lane 3 (100 nM).

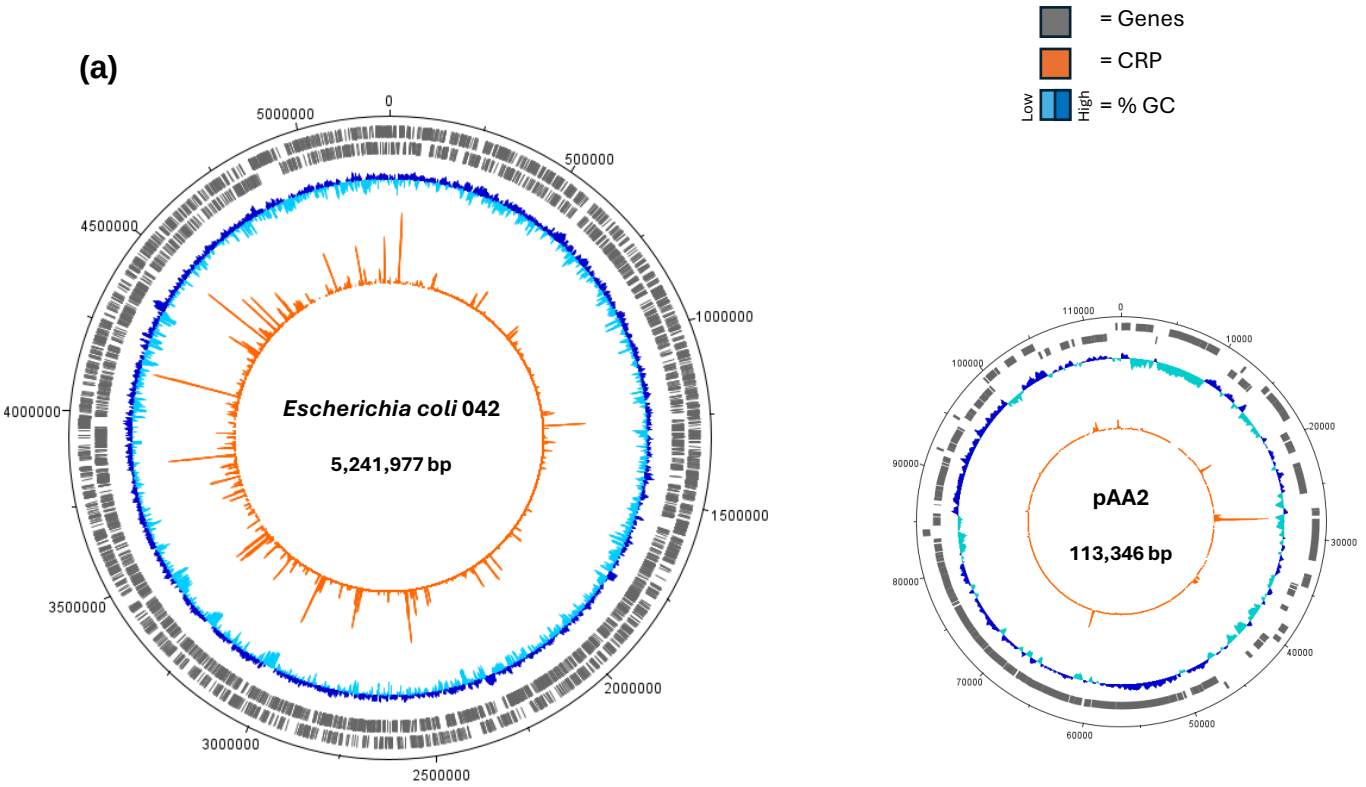
590 **Fig. 5.** CRP-dependent repression of the EAEC 042 0536 promoter. (a) Sequence of the p0536_100
591 promoter fragment, amplified to incorporate *Eco*RI and *Hind*III restriction sites. Numbering is relative
592 to the predicted transcript start site denoted by +1. Locations of the -10 element and predicted -35
593 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 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. β-galactosidase activity was measured in *E. coli* K12 Δ*lac* strains M182 and M182 Δ*crp* carrying a reporter plasmid pRW50 containing either the wild type 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. β-galactosidase activity is expressed as Miller Units (nmol of ONPG hydrolysed min^{-1} mg^{-1} dry cell mass). Data is from duplicate independent experiments, averaged from three biological replicates, with standard deviations shown. (c) Unwinding of the open complex at the -10 element was confirmed with potassium permanganate footprinting of an end-labelled *Aat*II/*Hin*dIII 0536 promoter fragment excised from pSR incubated, in the presence of cAMP (200 μM), 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 μM), 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 p0536 promoter. β-galactosidase activity was measured in *E. coli* K12 Δ*lac* M182 and M182 Δ*crp* strains carrying a reporter plasmid pRW50 containing either the wild type 0536 promoter (p0536_100) cloned upstream of *lacZ* or the 0536 promoter with 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. β-galactosidase activity is expressed as Miller Units (nmol of ONPG hydrolysed min^{-1} mg^{-1} dry cell mass). Data is from duplicate independent experiments, averaged from three biological replicates, with standard deviations shown.

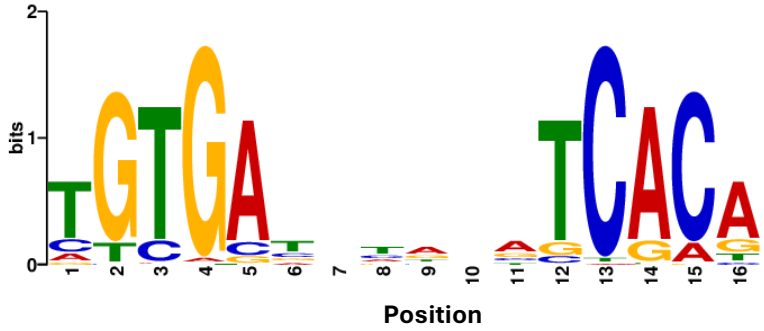
Fig. 6. CRP-dependent repression of the EAEC 042 *kpsM* promoter. (a) Sequence of the p*kpsM*_440 promoter fragment, amplified to incorporate *Eco*RI and *Hind*III restriction sites. Numbering is relative to the predicted transcript start site from the P1 promoter noted by +1. Locations of the P1 and 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. β-galactosidase activity was measured in *E. coli* K12 Δ*lac* M182 and M182 Δ*crp* strains carrying a reporter plasmid pRW224 containing either the starting *kpsM* fragment (p*kpsM*_440) cloned upstream of *lacZ*, or the p*kpsM* fragment with a 13G mutation in the P1 -10 element (P1-), or 245G in P2 (P2-), or 13G 245G in both (P1-P2-). pRW224 with no insert was included as a negative control. β-galactosidase activity is expressed as Miller Units (nmol of ONPG hydrolysed min^{-1} mg^{-1} dry cell mass). Data is from duplicate independent experiments, averaged from three biological replicates, with standard deviations shown. (c) *kpsM* P1 and P2 transcripts were detected by multi round *in vitro* transcription from the p*kpsM*_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 μM), lane 2 (0.1 μM), lane 3 (0.2 μM), lane 4 (0.4 μM), lane 5 (0.6 μM), lane 6 (0.8 μM), lane 7 (1 μM), and lane 8 (2 μM).

Figure 1

(a)


= Genes
= CRP
Low
High
= % GC
Escherichia coli 042
5,241,977 bp
0
500000
1000000
1500000
2000000
2500000
3000000
3500000
4000000
4500000
5000000
pAA2
113,346 bp
0
10000
20000
30000
40000
50000
60000
70000
80000
90000
100000
110000

(b)


bits
0
1
2
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Position

(c)

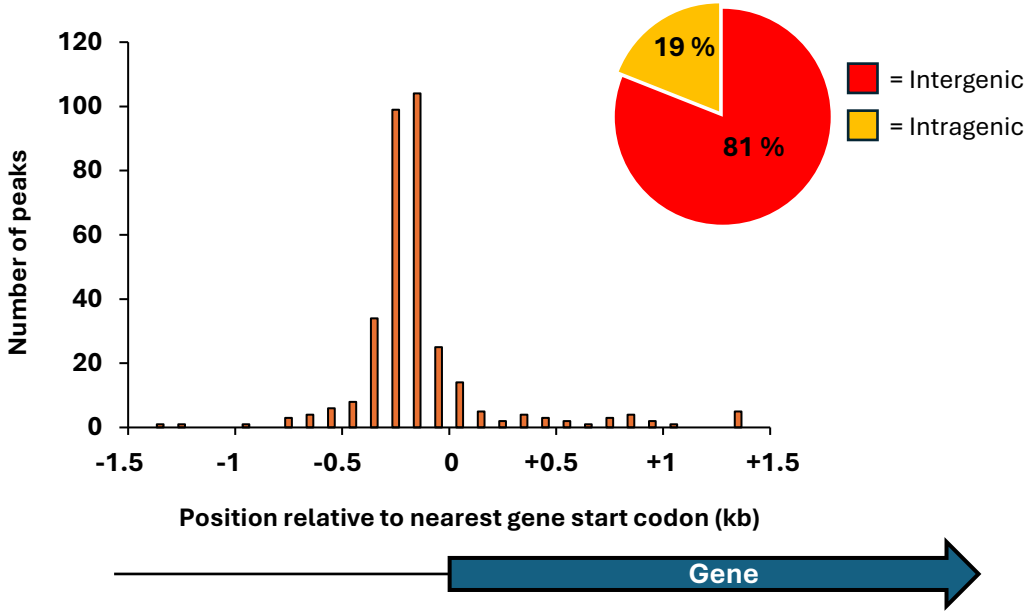

Number of peaks
0
20
40
60
80
100
120
-1.5
-1
-0.5
0
+0.5
+1
+1.5
Position relative to nearest gene start codon (kb)
19 %
81 %
= Intergenic
= Intragenic
Gene

Figure 2

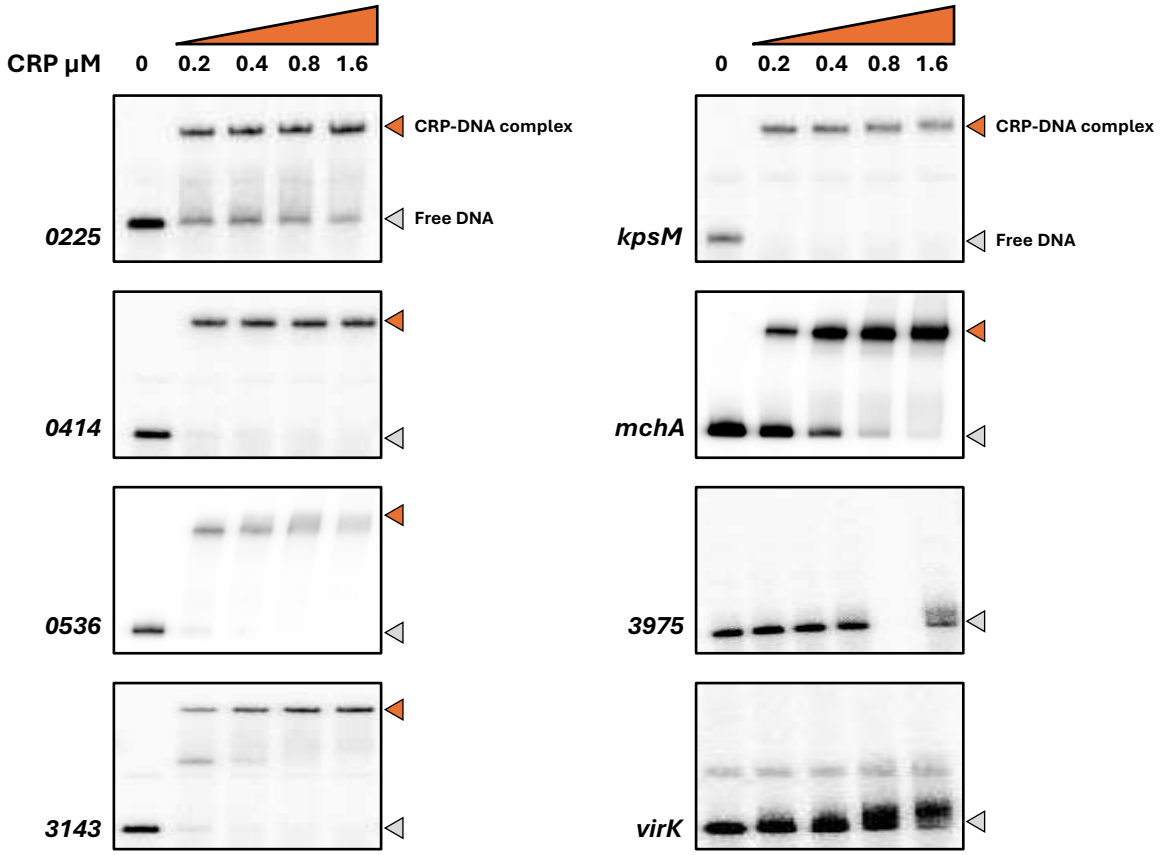

CRP μM
0
0.2
0.4
0.8
1.6
CRP-DNA complex
Free DNA
0225
0414
0536
3143
kpsM
mchA
3975
virK

Figure 3

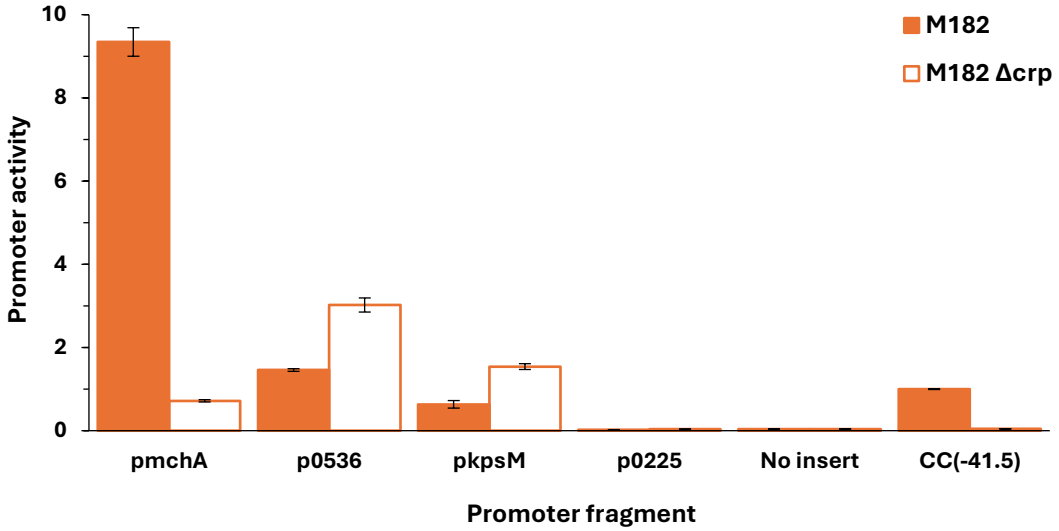

M182
M182 Δcrp
Promoter activity
10
8
6
4
2
0
pmchA
p0536
pkpsM
p0225
No insert
CC(-41.5)
Promoter fragment

Figure 4

(a)

-140 -110 -80

GAATTCGTAGGTTGGCATAACTGTCCTGTAACTCATTGTTCAAAAAAGAAAGCTCCTACAAACTGTATCCTGTATCTCT

***Eco*RI**

37G 14A 11A 11G +1

-50 -20

ACCAGAAGATAAGATACTTGGTTATTTTTGTGCATAAGCTCACATAAAAATAATATTAATTATGTTTTCTTAGCGATAG

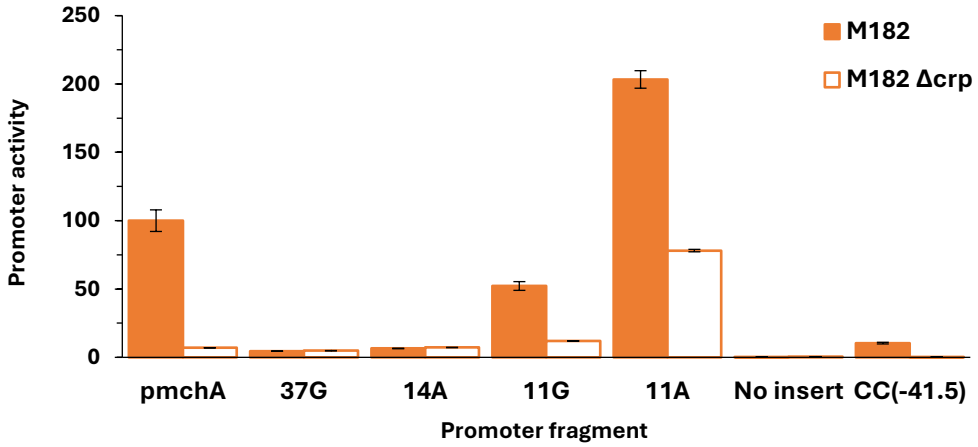
DNA site for CRP -10 element

10 40

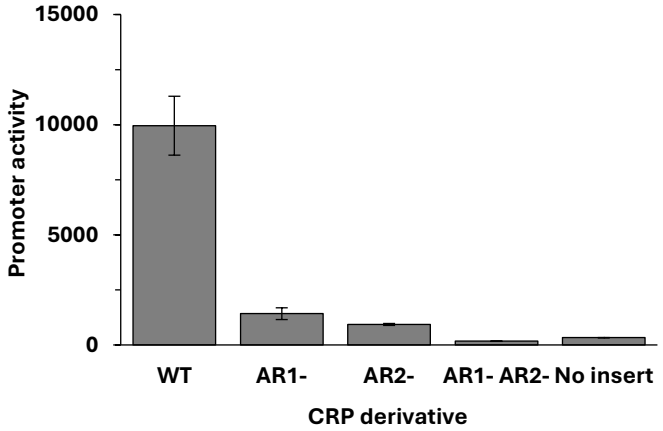
AAAAATAAACAGGTAAGCAGAAGACGGACTTTTTTA**ATG**CGAAAACGTA**AAGCTT**

***mchA* start codon** ***Hind*III**

(b)



(c)



(d)

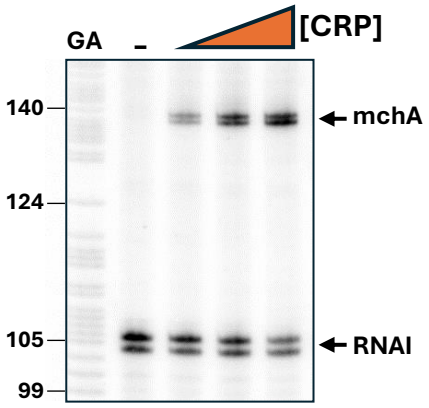
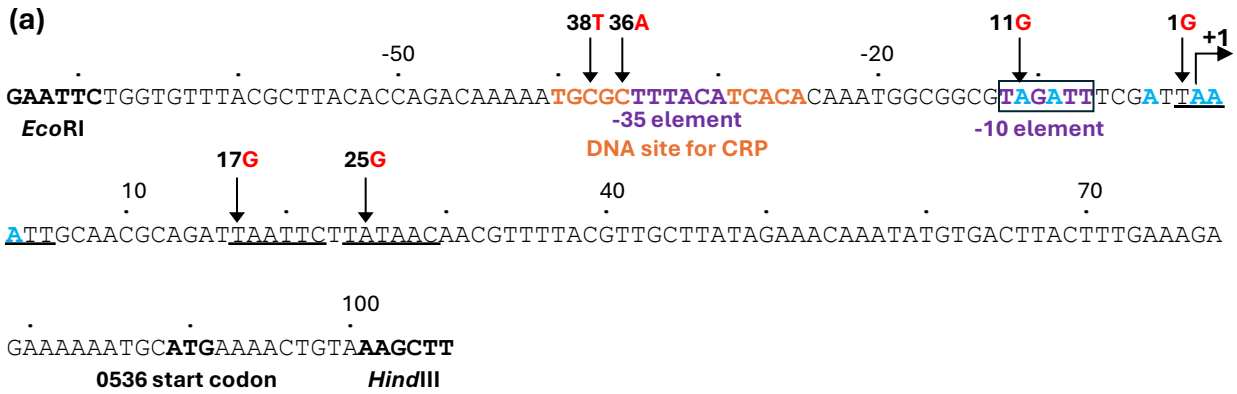
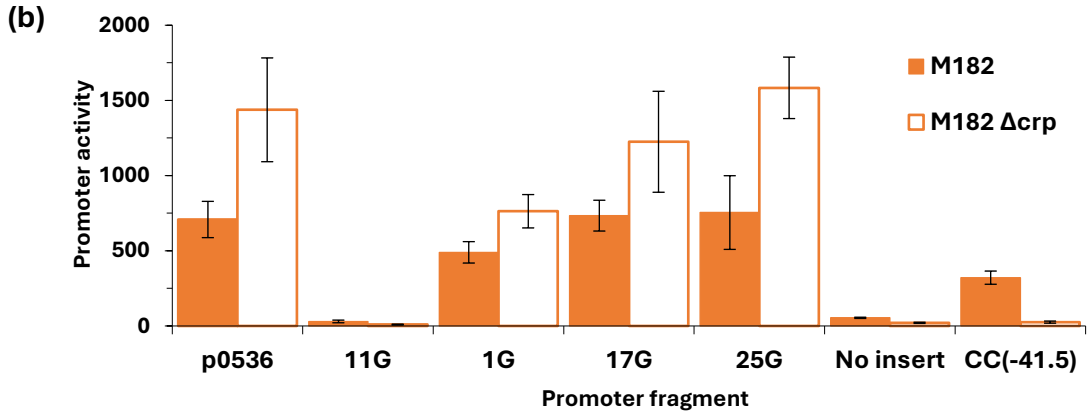
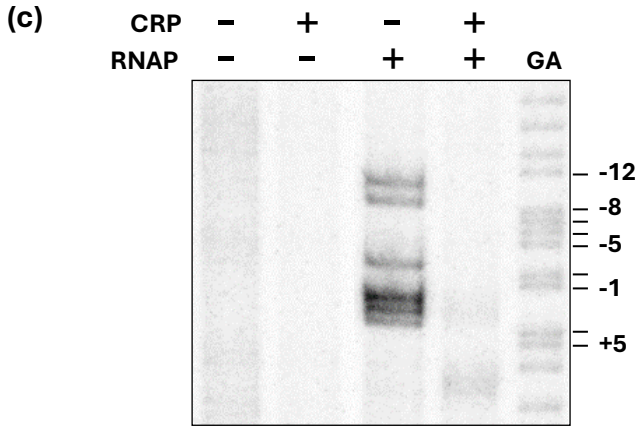
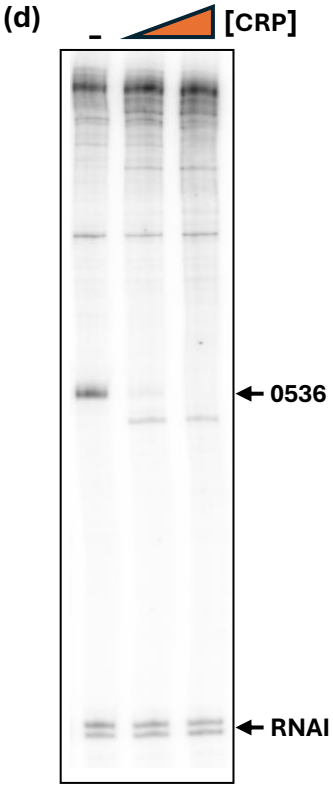


Figure 5


(a)
38T 36A
11G
1G
+1
-50
-20
GAATTCTGGTGTTTACGCTTACACCAGACAAAAATGCGCTTTACATCACACAAATGGCGGCGTAGATTTCGATTAA
EcoRI
-35 element
DNA site for CRP
-10 element
17G
25G
10
40
70
ATTGCAACGCAGATTAATTCTTATAACAACGTTTTACGTTGCTTATAGAAACAAATATGTGACTTACTTTGAAAGA
100
GAAAAAATGCATGAAAACTGTAAAGCTT
0536 start codon
HindIII


(b)
Promoter activity
2000
1500
1000
500
0
M182
M182 Δcrp
p0536
11G
1G
17G
25G
No insert
CC(-41.5)
Promoter fragment


(c)
CRP
RNAP
GA
-12
-8
-5
-1
+5


(d)
[CRP]
0536
RNAI

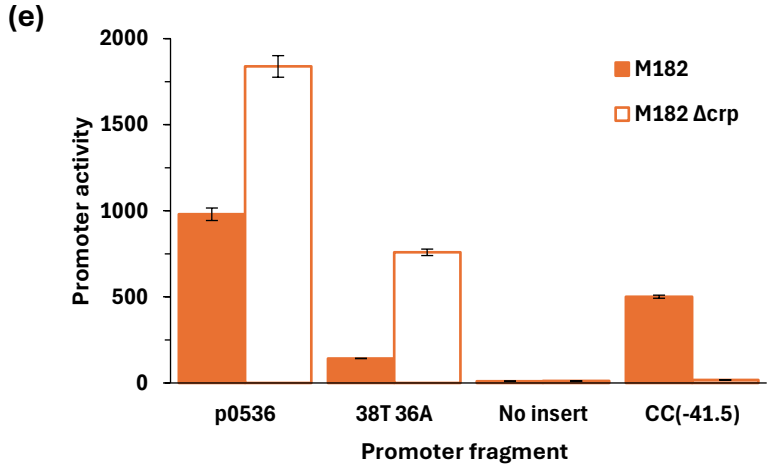
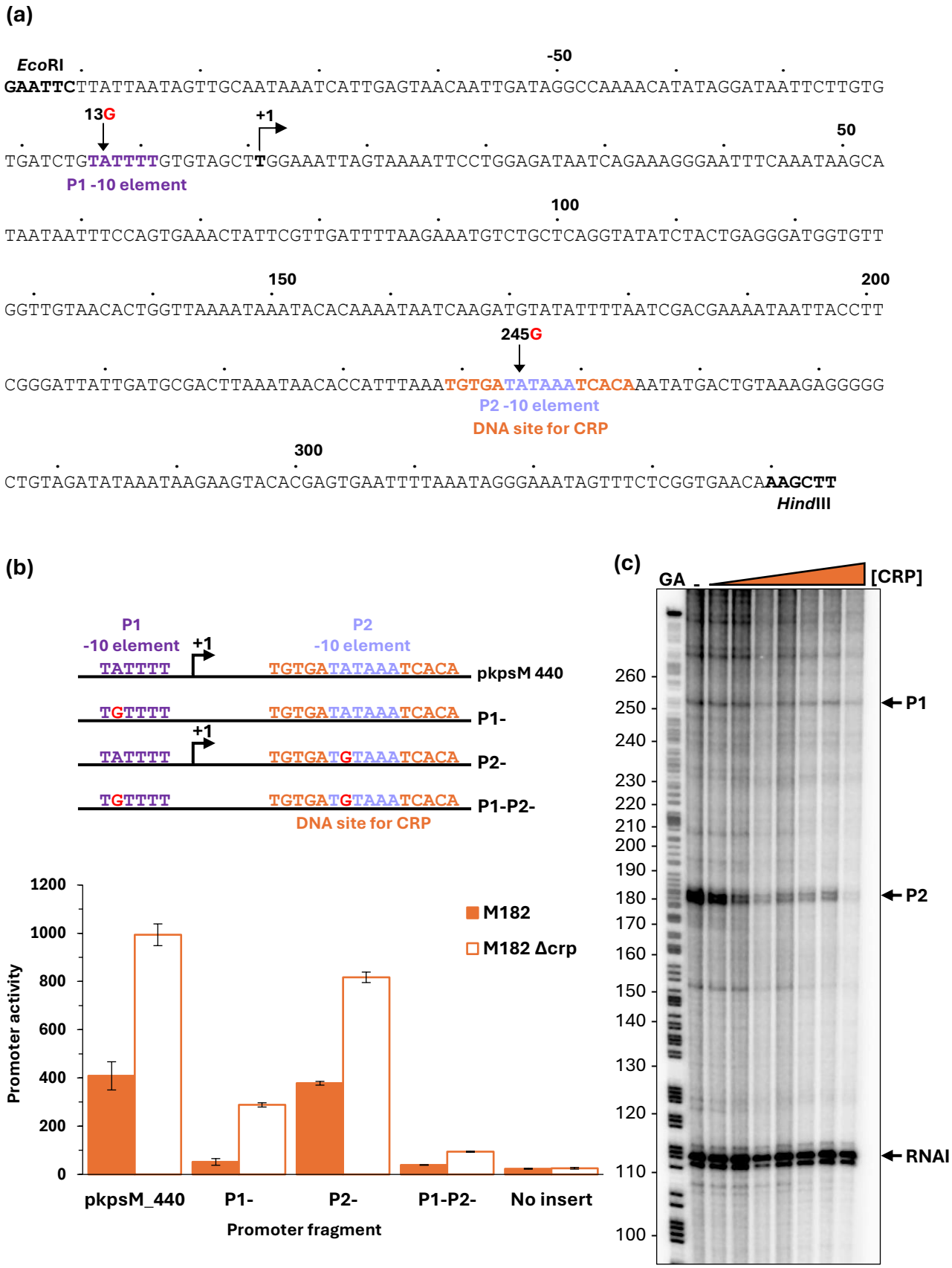

(e)
Promoter activity
2000
1500
1000
500
0
M182
M182 Δcrp
p0536
38T 36A
No insert
CC(-41.5)
Promoter fragment

Figure 6


(a)
EcoRI
-50
GAATTCTTATTAATAGTTGCAATAAATCATTGAGTAACAATTGATAGGCCAAAACATATAGGATAATTCTTGTG
13G
+1
50
TGATCTGTATTTTGTGTAGCTTGGAAATTAGTAAAATTCCTGGAGATAATCAGAAAGGGAATTTCAAATAAGCA
P1 -10 element
100
TAATAATTTCCAGTGAAACTATTCGTTGATTTTAAGAAATGTCTGCTCAGGTATATCTACTGAGGGATGGTGTT
150
200
GGTTGTAACACTGGTTAAAATAAATACACAAAATAATCAAGATGTATATTTTAATCGACGAAAATAATTACCTT
245G
CGGGATTATTGATGCGACTTAAATAACACCATTTAAATGTGATATAAATCACAAATATGACTGTAAAGAGGGGG
P2 -10 element
DNA site for CRP
300
CTGTAGATATAAATAAGAAGTACACGAGTGAATTTTAAATAGGGAAATAGTTTCTCGGTGAACAAAGCTT
HindIII
(b)
P1
-10 element
+1
P2
-10 element
TATTTT
TGTGATATAAATCACA
pkpsM 440
TGTTTT
TGTGATATAAATCACA
P1-
+1
TATTTT
TGTGATGTAAATCACA
P2-
TGTTTT
TGTGATGTAAATCACA
P1-P2-
DNA site for CRP
1200
1000
800
600
400
200
0
Promoter activity
M182
M182 Δcrp
pkpsM_440
P1-
P2-
P1-P2-
No insert
Promoter fragment
(c)
GA
[CRP]
260
250
240
230
220
210
200
190
180
170
160
150
140
130
120
110
100
P1
P2
RNAI