Identification of the CRP regulon using in vitro and in vivo transcriptional profiling

Zheng, Dongling; Constantinidou, Chrystala; Hobman, Jonathan; Minchin, Stephen

DOI:
10.1093/nar/gkh908

Document Version
Peer reviewed version

Citation for published version (Harvard):
Zheng, D, Constantinidou, C, Hobman, J & Minchin, S 2004, 'Identification of the CRP regulon using in vitro and in vivo transcriptional profiling', *Nucleic Acids Research*, vol. 32, pp. 5874-5893.
https://doi.org/10.1093/nar/gkh908

Published online November 1, 2004

5874–5893 *Nucleic Acids Research, 2004, Vol. 32, No. 19*
doi:10.1093/nar/gkh908

Identification of the CRP regulon using *in vitro* and *in vivo* transcriptional profiling

Dongling Zheng, Chrystala Constantinidou*, Jon L. Hobman and Stephen D. Minchin

School of Biosciences, University of Birmingham, Birmingham B15 2TT, UK

Received July 2, 2004; Revised September 3, 2004; Accepted October 7, 2004

ABSTRACT

The *Escherichia coli* cyclic AMP receptor protein (CRP) is a global regulator that controls transcription initiation from more than 100 promoters by binding to a specific DNA sequence within cognate promoters. Many genes in the CRP regulon have been predicted simply based on the presence of DNA-binding sites within gene promoters. In this study, we have exploited a newly developed technique, run-off transcription/microarray analysis (ROMA) to define CRP-regulated promoters. Using ROMA, we identified 176 operons that were activated by CRP *in vitro* and 16 operons that were repressed. Using positive control mutants in different regions of CRP, we were able to classify the different promoters into class I or class II/III. A total of 104 operons were predicted to contain Class II CRP-binding sites. Sequence analysis of the operons that were repressed by CRP revealed different mechanisms for CRP inhibition. In contrast, the *in vivo* transcriptional profiles failed to identify most CRP-dependent regulation because of the complexity of the regulatory network. Analysis of these operons supports the hypothesis that CRP is not only a regulator of genes required for catabolism of sugars other than glucose, but also regulates the expression of a large number of other genes in *E.coli*. ROMA has revealed 152 hitherto unknown CRP regulons.

INTRODUCTION

The *Escherichia coli* cyclic AMP receptor protein (CRP) is an important transcription factor that regulates transcription initiation for more than 100 genes mainly involved in catabolism of carbon sources other than glucose (1). *E.coli* preferentially utilizes glucose over other sugars and only catabolizes other sugars when the supply of glucose has become depleted [reviewed in (2)]. The presence of glucose prevents *E.coli* from catabolizing alternative sugars by several mechanisms, one of which is that glucose lowers the level of cAMP, the inducer for CRP.

CRP functions as a dimer in the form of a CRP–cAMP complex, and regulates transcription initiation by binding to a symmetrical DNA sequence (consensus sequence 5′-AAA**TGTGA**TCTAGA**TCACA**TTT-3′), located near or within the promoter regions. At CRP-dependent promoters, CRP activates transcription by making direct protein–protein contacts with RNA polymerase (RNAP). Class I CRP-dependent promoters, e.g. *lac*, contain a single CRP-binding site upstream of the DNA-binding sites for RNAP. At these promoters, CRP activates transcription by interacting with the C-terminal domain of the RNAP α subunit (αCTD) via a surface-exposed patch, known as activating region 1 (AR1) (residues 156–164). Class II CRP-dependent promoters, e.g. *gal*P1 and *melR*, contain a CRP-binding site overlapping the −35 hexamer for RNAP. At these promoters, CRP makes multiple interactions with RNAP, including between AR1 and αCTD, the activating region 2 (AR2) (residues 19, 21 and 101) and the N-terminal domain of α subunit (αNTD) and between AR3 (residues 52–58) and the RNAP σ^{70} subunit region 4. Many promoters contain tandem CRP sites and are known as Class III promoters. CRP activation at these promoters involves a combination of the Class I and Class II mechanisms. Additionally, many CRP-dependent promoters are co-dependent on a second activator (3).

The recent completion of several bacterial genome sequences has facilitated the development of computer-based bioinformatic approaches and microarray techniques to study transcription regulation at a genome-wide level. Bioinformatic analysis has allowed the prediction of the regulon of a particular transcription factor by searching for the consensus sequences, or by using algorithms to search for sequence patterns within the genome. CRP-binding sites present within the *E.coli* genome have been predicted using different computational approaches (4,5). In the latter study, Tan *et al*. have predicted 161 strong (including known sites) and 285 weak candidate CRP-binding sites, using a comparative genomic approach. However, *in silico*-predicted DNA-binding sites may not be occupied *in vivo*, or may only function in the presence of other factors. Conversely, operons with weak, but functional, sites may not have been identified using these approaches. In this study, we have utilized oligonucleotide microarray technology to determine the CRP regulation experimentally, and compared these data with the predicted CRP-binding sites.

The DNA microarray technology allows thousands of genes to be studied simultaneously in a single experiment, and thus provides a powerful tool to investigate gene function at a genomic level. Transcription profiles of *E.coli* in different media (6,7), under various stress conditions (8,9) and between different strains (10,11), have revealed the roles of many regulatory factors. However, the interference of other

*To whom correspondence should be addressed. Tel: +44 121 414 5564; Fax: +44 121 414 5925; Email: C.Constantinidou@bham.ac.uk

regulatory networks makes it difficult to distinguish direct effects on transcription from indirect effects, and therefore it is hard to directly link the results to a specific transcriptional factor. Recently, Helmann and co-workers (12) have developed a novel technique, which combined *in vitro* run-off transcription with macroarray analysis (ROMA), to define the direct effects of σ^W on *Bacillus subtilis* promoters. Run-off transcription reactions in this experimental system used genomic DNA as template. The resulting ^{33}P-labelled RNA products, transcribed by either σ^W holoenzyme or core enzyme, were hybridized to macroarray membranes. Direct visual comparison of the core versus holoenzyme experiments allows the ready identification of genes targeted by σ^W. The result from ROMA for σ^W from *B.subtilis* was consistent with those determined by promoter consensus searching and *in vivo* transcriptional profiling. In this study, the term '*in vitro*' refers to experiments completed with purified DNA and protein components, whereas '*in vivo*' refers to experiments in *E.coli* cells grown in planktonic culture.

In this study, the ROMA procedure was modified so as to exploit oligonucleotide arrays on glass slides and ROMA was then used to identify CRP-regulated operons *in vitro*.

MATERIALS AND METHODS

Strains and growth conditions

E.coli K-12 MG1655 (CGSC 7740) was used in this study (13). A Δ*crp* derivative of MG1655 was constructed using the gene disruption method of Datsenko and Wanner (14). Primers: CRP P1: 5′-GCTCTGGAGAAAGCTTATAACAGAGG ATAACCGCGCGTGTAGGCTGGAGCTGCTTC-3′ and CRP P2:5′-TGGCGCGCTACCAGGTAACGCGCCACTCCGACGGGACATATGAATATCCTCCTTAG-3′ were used to amplify a chloramphenicol resistance gene cassette from plasmid pKD3 (14). The bases underlined in the primer sequences shown above represent the short regions of homology to genomic sequences flanking *crp*, which were used in Red-mediated recombination of the chloramphenicol resistance cassette into the chromosome. Transformants were grown on Luria–Bertani (LB) agar (15) containing 0.2% glucose and 50 μg/ml chloramphenicol. PCR was used to screen chloramphenicol resistant transformants, for replacement of *crp* with the chloramphenicol resistance cassette, using the primers CRPSCREEN 1: 5′-GGATGCTACAGTAATACATTGATG-3′, and CRPSCREEN2: 5′-GACCGAATCGTAATTCGCCAAG-3′. Amplicons generated by the screening PCR were sequenced using a BigDye™ version 3 sequencing kit (Applied Biosystems, Warrington, UK) and analysed on an ABI 3700 sequencer (Applied Biosystems). MG1655 Δ*crp* strains were grown on Maconkey agar containing maltose (1% w/v) to confirm the phenotype.

For in *vivo* microarray experiments, both wild-type and Δ*crp* strains were grown in M9 minimal media (14) containing 0.2% fructose at 37°C to OD_{600} 0.8. Glucose (0.2%) was then added to the cultures, which were grown for a further 15 min, prior to harvesting. The cell samples for RNA preparations were collected before and after the addition of glucose. Two volumes of RNAProtect (Qiagen Ltd, Crawley, UK) were added per volume of bacterial culture, prior to centrifugation.

E.coli oligonucleotide array

E.coli oligonucleotide arrays were produced by the UBEC group (University of Birmingham *E.coli* group). The oligonucleotides were designed and synthesized by Qiagen Operon Ltd as the *E.coli* Array ready oligo set vs 1.0 (Qiagen) and represented 4289 *E.coli* K12 strain (MG1655) open reading frames (ORFs); 1416 ORFs designed to the O157:H7 (EDL933) strain; and 273 ORFs unique to the O157:H7 (Sakai) strain (http://oligos.qiagen.com/arrays/oligosets_ecoli.php). In addition, 110 oligonucleotides representing ORFs from the EDL933 and Sakai plasmids were added to the oligonucleotide array set. Also included within the array set were 12 positive and 12 negative control oligonucleotides, which were printed within each subarray. The average size of each oligonucleotide used in the array was a 70mer, with a $T_m = 75 \pm 5$°C. The position of each oligonucleotide within the ORF was more than 40 bases away from the 3′ end. The oligonucleotides were arrayed on Corning CMT-GAPS II slides in 48 blocks, each containing 324 spots (18 rows by 18 columns) using a MicroGrid II robot (BioRobotics, UK). Each oligonucleotide was printed in duplicate on the array, and Amersham Lucidea™ Universal Scorecard™ (Amersham, UK) controls were printed within each subarray.

Genomic run-off transcription for ROMA

Genomic DNA was isolated from the *E.coli* MG1655 (CGSC 7740) wild-type strain using phenol/chloroform extraction (http://www.research.umbc.edu/~jwolf/m1.htm), digested with EcoRI overnight and purified by phenol/chloroform extraction, followed by isopropanol precipitation. RNAP holoenzyme was purified as described previously (16). Wild-type CRP, AR1-mutated CRP (HL159) and AR2-mutated CRP (KE101) were purified as described previously (17).

For a single ROMA experiment, two run-off transcription reactions were set up in parallel. The control reaction contained 4 μg of EcoRI digested genomic DNA, and 1 mM each of ATP, GTP, CTP and UTP, in transcription buffer (40 mM Tris/acetate, pH 7.9, 10 mM $MgCl_2$, 1 mM DTT, 100 mM KCl, 0.1 mg/ml BSA), which contained RNase Out (Invitrogen, UK) at a concentration of 50 U/reaction. The test reaction contained 2 μl of 5 mM cAMP (final concentration 200 μM), and 4 μl of 10 μM CRP (40 pmol of monomer, at a final concentration of 800 nM) in addition to the components of the control mixture. Both run-off transcription reaction mixtures were incubated for 10 min at 37°C and transcription started by the addition of 20 pmol of RNA polymerase holoenzyme. Run-off transcription reactions were incubated at 37°C for 30 min. After incubation, the reactions were stopped by the addition of 5 μl of 250 mM EDTA and placed on ice.

RNA purification and labelling

Both RNA transcripts from *in vitro* transcription reactions and total RNA from cell cultures were extracted using the Qiagen RNeasy mini kit. On-column DNase I digestion (Qiagen Ltd) was used to remove contaminating genomic DNA from the RNA preparations. The concentrations of RNA or *in vitro* transcription products were determined at OD_{260} and OD_{280}. An indirect labelling method was used to obtain fluorescence-labelled cDNA for any RNA sample. Briefly, all RNA transcripts from each reaction or 10–20 μg of total

RNA were mixed with 6 μg of random hexamers (Amersham Biosciences, Little Chalfont, UK) to a final volume of 18.4 μl, incubated at 70°C for 10 min and snap-cooled in ice. Reverse transcription labelling mixture (11.6 μl) was then added to the RNA template and random hexamers, which contained 0.5 mM dATP, dCTP, dGTP, 0.2 mM dTTP, 0.3 mM aminoallyl-dUTP (aa-dUTP), RNase inhibitor (30 U), 400 U SuperScript II (Invitrogen, Paisley, UK), 10 mM DTT and 1× first strand buffer. The mixture was incubated at 42°C for 3 h or overnight to generate aminoallyl-labelled cDNA. To hydrolyse the RNA template, 10 μl of 0.5 M EDTA and 10 μl of 1 M NaOH were added to the reaction and incubated at 65°C for 15 min. The reaction was neutralized by the addition of 10 μl 1 M HCl. Unincorporated aa-dUTP and free amines were removed by washing the cDNA in a microconcentrator (Microcon YM-30, Millipore) and the sample was then vacuum dried. The aminoallyl-cDNA pellet was resuspended in 4.5 μl of 0.1 M sodium carbonate buffer (pH 9.0) and coupled with Cy3 or Cy5 monoreactive dye (Amersham), prepared in dimethyl sulfoxide for 2 h at room temperature in the dark. For the ROMA experiment, the aa-cDNA from the control reactions was coupled with Cy3 and aa-cDNA from the CRP reactions coupled with Cy5. For *in vivo* transcriptional profiling, the aa-cDNA obtained from either wild-type strain or ΔCRP strains before the addition of glucose was coupled with Cy3 and after the addition of glucose was coupled with Cy5. Uncoupled dyes were removed using a QIAquick PCR purification kit (Qiagen). Lucidea™ Scorecard (Amersham) mRNA spike mixes were added to labelling reactions for both *in vivo* and ROMA experiments.

For the *in vivo* microarray work, the quality and concentration of the RNA prepared was assessed using the Agilent 2100 Bioanalyser. Total RNA (10–20 μg) was labelled using the CyScribe Post-Labelling Kit (Amersham Biosciences) as described by the manufacturer.

Prehybridization and hybridization

Before hybridization, the arrayed slides were prehybridized in a buffer containing 25% formamide, 5× SSC, 0.1% SDS and 10 mg/ml BSA at 42°C for 2 h and washed by dipping twice in distilled water. The slides were then dipped in 95% ethanol for 1 sec and dried in a clean 50 ml centrifuge tube by centrifugation at 1500 *g* for 10 min. The Cy3 and Cy5 labelled cDNAs were mixed, vacuum dried and resuspended in 70 μl hybridization buffer containing 25% formamide, 5× SSC, 2 μl of 50× Denhardt's, 2 μl yeast tRNA (20 μg/ml) and 0.1% SDS. The labelled cDNAs were denatured by heating at 95°C for 5 min, and applied to the prehybridized slide in a CMT-Hybridization chamber (Corning Inc., Corning, NY). A HybriSlip (Sigma) was carefully lowered onto the slide. To maintain humidity inside the chamber, 10 μl of distilled water was added to the two reservoir wells. The chamber was then tightly sealed and incubated at 42°C for 16–20 h in the dark. The slide was then removed from the chamber, washed for 5 min sequentially in 2× SSC/0.1% SDS buffer, 0.1× SSC/0.1% SDS buffer and 0.1× SSC buffer, rinsed in distilled water for 5 s and dried by centrifugation at 2000 r.p.m. for 10 min. The hybridized slides were scanned with a confocal laser scanner (Axon GenePix 4000A) using appropriate gains on the photomultiplier tube to obtain the highest intensity without saturation. Three replicates were completed for each experiment.

Image extraction and data normalization

Scanned images for Cy3 and Cy5 were then overlaid with GenePix Pro 3.0 software. Only data generated from spots representing *E.coli* MG1655 genes were analysed in our studies. Spots with background-subtracted intensity lower than 100 in both Cy3 and Cy5 channels were filtered out. GeneSpring software (SiliconGenetics) was used for global normalization and density-dependent normalization (Lowess) to correct artefacts dependent on density or caused by different dye incorporation rates for the two dyes. The duplicate spots for each gene on a single slide were taken as two individual spots. Three independent experiments were performed for each comparison and, therefore, six replicate data sets were obtained for each gene. The normalized data from GeneSpring software were exported to a Microsoft Excel spreadsheet.

Data reproducibility

To assess data reproducibility, the spot-to-spot variation was calculated as described by Loos *et al*. (18) and correlation between data sets analysed by the Pearson correlation coefficient (*r*). The spot-to-spot variation was represented by the percent error calculated by dividing the SD by the average ratio for each gene ($\% = \sigma/\mu$). Average variation within a slide was calculated as the average of the gene spot-to-spot variations on that slide (two spots), and average variation across slides was calculated as the average of gene spot-to-spot variations based on the six spots of the data set. The latter variation included spot, slide and cDNA variability. The Pearson correlation coefficient (*r)* was calculated for Cy3 data sets from any two slides and for Cy5/Cy3 from any two replicate slides.

Identification of differentially transcribed genes

Differentially transcribed genes were selected using an outlier iteration method (18,19). The data for each gene were averaged and the geometric mean and SD were calculated for the entire population. Any gene with a log-ratio more than three SDs away from the mean was considered an outlier. Outliers were then removed from the population and retained within the differentially expressed subset, and the mean and SDs were recalculated for the rest of the data. The step was repeated until few or no outliers were detected. The 99% predictive interval (PI) was set for the final cut-off ratio to define the remaining differentially transcribed genes within the now symmetrical and well-defined distribution.

The *in vitro* transcription profiles of CRP derivatives were compared with wild-type CRP, and significant changes were identified by outlier iteration. The average of the log-ratios for each gene from the CRP derivative experiments was normalized to the wild-type CRP experiment. The geometric mean and SD were calculated for the entire population. The outliers were selected as stated above and the 95% PI was set as a final cut-off. The genes with a log-ratio change between two experiments falling in the outlier group or beyond the PI were regarded as differentially regulated by CRP mutants and wild-type CRP.

Sequence processing

The differentially transcribed genes were grouped into operons using the RegulonDB database at http://www.cifn.unam.mx/Computational_Genomics/regulondb. The current version of the RegulonDB database (RegulonDB 4.0) identifies 87 CRP-regulated operons that have been verified experimentally. In this study, 24 of these operons were shown in the ROMA experiments to be regulated by CRP and, therefore, were further analysed. CRP-binding site predictions were based on the previous study (5), where a Positional Weight Matrix was generated from alignment of known CRP-binding sites and used to determine the strength of a site. In this study, a strong site was defined as a sequence with a score more than 10 (consistent with the prediction by Tan *et al.*) and a weak site as a sequence with a score between 5 and 10. Sequences with a score below 5 were rejected.

RESULTS

Combination of run-off transcription and microarray analysis (ROMA) on glass slides

The original ROMA procedure used macroarray membranes to analyse radioactively labelled RNA transcripts (12). In our study, the procedure was modified by using a post-transcriptional fluorescence-labelling procedure to enable hybridization to a sense-oligonucleotide array. To gauge data reproducibility, data from each of three replicate slides were subjected to statistical analysis. Gene spot-to-spot variation was 5.0% (range, 4.1–6.3%) between replicate spots within a given slide, and 13.9% (range, 12.6–15.4%) among six spots across three slides. The latter number includes both the hybridization variation and biological variation. Comparison of the duplicated probe sets within a single hybridization showed good reproducibility of the fluorescence signals with a Pearson correlation coefficient (r) of >0.98. This result illustrates uniform hybridization of the microarray. The reproducibility of signal between replicate reactions was lower, with a Pearson correlation coefficient (r) of >0.80. This variability was considered to be mainly due to variations from independent reactions. In addition, the detail of signal intensities showed that ~90% of all genes gave a detectable signal.

CRP regulatory profiles in the ROMA system

Comparison of RNA transcripts from reactions containing wild-type CRP to control reactions containing only RNAP revealed many differentially transcribed genes (Figure 1a). Data for the majority of genes fall within the threshold values, with a ratio between +2 and −2, and therefore indicates that

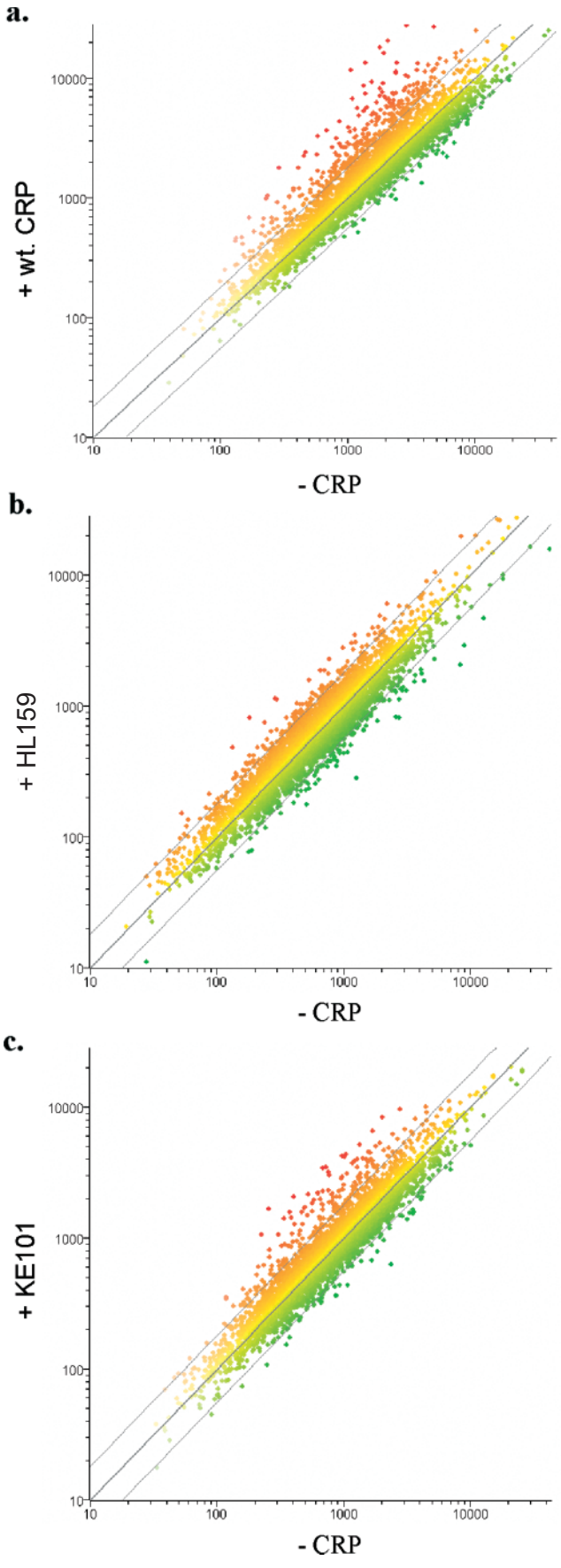

Figure 1. Logarithmic-scale scatter plots of spot intensities. The data were subjected to global and density-dependent normalization, and plotted on GeneSpring 4.2.1. Each data point corresponds to the average of two spots representing an individual MG1655 gene. The ratios were colour coded, with red for ratio above one and green for ratio below one. The middle line passes genes with no change at two conditions (ratio = 1) and the other two lines demarcate threshold values for genes with significant increase or decrease in transcription in response to CRP deletion (ratio = ±2). (**a**) Transcriptional profiles from reactions with wild-type CRP (wt. CRP) versus without CRP (−CRP). (**b**) Transcriptional profiles from reactions with AR1-mutated CRP (HL159) versus without CRP (−CRP). (**c**) Transcriptional profiles from reactions with AR2-mutated CRP (KE101) versus without CRP (−CRP).

these genes are not significantly regulated by CRP. As expected, transcripts of many genes were increased by the addition of CRP and transcripts of very few genes were reduced, which is consistent with the view that CRP activates most of the genes that it regulates. A total of 280 genes, present in 188 different operons, had significantly higher transcriptional levels and 20 genes, in 16 operons, had reduced transcriptional levels in response to CRP. Taken together, these data show that ~7% of the genes in the *E.coli* genome were transcribed differentially upon the addition of CRP in the ROMA experiment. The highest activation, 13-fold induction, was observed at the *sdaC* gene encoding a protein that is a putative serine transporter. Of the 176 operons identified by ROMA in this study, 24 are known members of the CRP regulon (Table 2), i.e. the promoter regions and CRP sites have been mapped experimentally. For genes present within the same polycistronic operon, the signal for the promoter-proximal gene should be the same or higher than the signal for promoter distal genes. Of the 188 operons activated by CRP, the promoter-proximal gene was activated in the vast majority (176 operons). For 8 of these 176 operons, the promoter-proximal gene was subjected to only marginal activation (above 95% PI), while the promoter distal gene(s) were activated at a significant level. In only 12 operons was a promoter-distal gene up-regulated and not the promoter-proximal gene (Tables 1 and 2). These 12 promoters were not included in the CRP-regulated set. Up-regulation of promoter-distal genes in the absence of regulation of the promoter-proximal genes may represent false positives in the assay; alternatively it may be due to CRP-dependent promoters within operons that are active *in vitro*.

Determination of CRP-binding sites

We scanned the upstream regions of the 176 up-regulated operons for putative CRP-binding sites to determine whether the activation observed is due to a site that matches the CRP-binding site consensus. A previous study using a comparative genomic approach (5) identified 447 operons on the *E.coli* genome that contain putative CRP sites upstream from the translation start site. Using the same approach as Tan *et al*., 55 of the 176 (31%) operons identified in our study were found to contain a CRP site. Tan *et al*. predicted CRP-binding sites with a good match to the consensus (10 as a cut-off score). However, some known CRP-dependent promoters contain weak CRP sites that are not identified when using a cut-off of 10 (see Table 2), e.g. the *melR* CRP site has a score of 5.5 and *rpoH* as two sites with scores of 2.2 and 7.4. In addition, they also identified many false positives.

We therefore scanned the regulatory regions of the remaining operons for sequences with a lower match to the consensus (5 as a cut-off score) and found an additional 70 operons that contain CRP sites with a score between 5 and 10 (Tables 1 and 2). Thus, a total of 125 operons identified in the ROMA assay contain CRP sites identified by either Tan *et al*. (5) or this study. The remaining 51 operons (29%) identified by ROMA do not contain a recognizable CRP site (score <5).

Effect of CRP AR1 and AR2 on activation

Previous studies have shown that CRP activation is dependent on the interactions of its activating regions with subunits of RNAP [reviewed in (3)]. In this study, the effects of two CRP derivatives containing a mutation at either AR1 or AR2 were analysed. The HL159-CRP carries a His-to-Leu mutation at position 159 within AR1 and has been shown to be defective at both Class I and Class II promoters. Whereas KE101-CRP, which contains a Lys-to-Glu mutation at position 101 within AR2, fails to activate transcription at Class II promoters, but still functions at Class I promoters (20).

The transcriptional level of each gene in the presence of either HL159-CRP or KE101-CRP was compared with those from the control reaction. As expected, the results showed that HL159-CRP failed to activate transcription of 86% (151/176) of the CRP-dependent genes identified in the initial ROMA experiment (Table 1, Figure 2 and Figure 1b). Therefore only 14% (25/176) of operons were activated by HL159-CRP to the same level as by wild-type CRP. The AR2 mutant KE101-CRP was defective for activation at 59% (104/176) of CRP-dependent operons (Table 1, Figure 2 and Figure 1c). The AR2 mutant, therefore, still activated transcription for many genes although the level of activation was reduced. We predict that the promoter regions of the operons whose transcription is not up-regulated by the AR2 mutant contain a CRP-binding site overlapping the −35 hexamer (Class II binding site).

To validate the ROMA data, we considered the 24 up-regulated operons whose promoter regions and CRP sites have been mapped experimentally. Of these 24 operons, 7 contain a single Class I CRP-dependent promoter (possess a CRP site far upstream of the −35 hexamer), 9 contain a single Class II promoter (possess a CRP site overlapping the −35 hexamer) and 8 contain a Class III promoter with tandem CRP-binding sites (Table 2). All Class III promoters contain a Class II binding site and a Class I binding site. In the ROMA assay, HL159-CRP failed to activate all promoters, except the two Class III promoters, *deoCABD* and *rpoH*. Both the *deoCABD* and the *rpoH* operons have several promoter regions that are under differential regulation by CRP (21,22) and therefore the effect of the AR1 mutation may be compromised by transcription from different promoters. In contrast to HL159-CRP, KE101-CRP could still activate class I promoters, but failed to activate all class II promoters and half of the class III promoters. It is therefore clear from the experiments with activating region mutants that the CRP regulation observed in ROMA is similar to that observed *in vivo*. In addition, it allows us to predict that 59% of the CRP-regulated operons identified by ROMA contain a class II CRP site and are therefore class II or class III promoters.

CRP-repressed operons in the ROMA system

Although CRP is predominantly an activator, we found 16 operons where the promoter-proximal gene had reduced levels of transcription in response to CRP in the ROMA experiments (Table 3). The strongest repression, 3-fold reduction, occurred at the *yjcB* gene. In contrast to activation, the repression levels of most operons were not significantly affected by either HL159-CRP or KE101-CRP, which indicates that most CRP repression is not dependent on either the AR1 or the AR2 determinant. However, repression at four operons was slightly affected by the AR1 mutation, suggesting some role for AR1 at these operons. We have identified possible

Table 1. Operons activated by wild-type CRP in ROMA experiment

Operon name[a]	ROMA WT[b]	ROMA HL[c]	HL/wt[d]	ROMA KE[e]	KE/wt[f]	CRP-binding site Position[g]	CRP-binding site Site sequence	Score[h]	Gene product	Gene function
acrE	1.86	1.35	↓	**1.89**	—	−289	AATCGTTAAATAAATAATATAT	7.2	Transmembrane protein affects septum formation	Membrane; inner membrane
adhE	2.49	1.13	↓	1.02	↓	−241	AAATTTGATTTGGATCACGTAA	18.72	Coa-linked acetaldehyde dehydrogenase and iron-dependent alcohol dehydrogenase	Enzyme; energy metabolism, carbon: Fermentation
arcA	1.84	1.11	↓	0.99	↓	−295	GAACTTGATATATGTCAACGAA	6.78	Negative response regulator of genes in aerobic pathways	Regulator; global regulatory functions
						−358	TAATGTGACGAAAGCTAGCATT	5.12		
argR	1.86	1.58	—	**2.05**	—	−33	TAATGTTGTATCAACCACCATA	6.42	Repressor of arg regulon	Regulator; amino acid biosynthesis: Arginine
						−193	CAATTTGATAACAATTAATTTA	5.65		
b0846	1.98	0.91	↓	**2.42**	—	−335	CATCGTGACCAGGATGACATTA	8.37	Putative DEOR-type transcriptional regulator	Putative regulator; not classified
						−40	TTGTGTACATTTGTACACAATT	7.94		
b1329	2.3	1.17	↓	**2**	—	−101	TATATTGATACTTAAAACATTT	6.94	Putative transport periplasmic protein	Putative transport; not classified
						−38	TTCTTTAAGGCGAAACAAATAA	5.8		
b1431	1.94	**1.72**	—	**1.59**	—	−216	GGATGTGAAATTAATCACAGTA	14.71	Orf, hypothetical protein	Orf; unknown
b1513_ydeYZ_b1516_yneB	4.56	1.2	↓	**1.56**	↓	−183	ATCTGTGATGGCAACCACAGTT	13.9	Putative ATP-binding component of a transport system	Putative transport; not classified
b2085_b2084_b2083	2.23	**1.9**	—	1.44	↓	−29	TATTATAATCCTATTCAATTAT	7.03	Orf, hypothetical protein	Orf; unknown
b2448_b2449	2.52	1.17	↓	1.04	↓	−156	TTTTGTGATGCAGATCGCTTTT	16.65	Orf, hypothetical protein	Orf; unknown
						−4	AACTGTGAAGGAGGACGCCATG	5.4		
b2450	2.35	1.16	↓	1.44	↓	−69	TTTTGTGATCTGCGTCAATATT	16	Orf, hypothetical protein	Orf; unknown
						−229	CAATGAAACTGAGTTCAAACTT	9.14		
b2462_b2461_b2459_eutI_cchA	2.03	1.07	↓	1.18	↓	−31	CTTAGTGATCTACCTCACCTTT	13.48	Orf, hypothetical protein	Orf; unknown
b2463	2.72	0.93	↓	0.99	↓	−257	ATGAGTGCGTTAATTCACACTT	13.2	Putative multimodular enzyme	Putative enzyme; not classified
b2611_ypjE_yfjD	1.7	1.02	↓	1.25	↓				Orf, hypothetical protein	Orf; unknown
b2710_ygbD	3.31	**2.02**	↓	**1.62**	↓	−90	TAGAGTAAAAACAATCAGATAA	7.05	Putative flavodoxin	Putative enzyme; not classified
b2736_b2737	2.13	1.51	—	1.4	↓	−90	TTATGTGAATCAGATCACCATA	18.82	Putative dehydrogenase	Putative enzyme; not classified
						−51	TTATATTCACAATATCAAACAA	5.14		
b2740	2.26	1.58	—	1.26	↓				Putative transport protein	Putative transport; not classified
b2876_b2875	3.24	1.03	↓	1.41	↓	−277	TTTTGATACTTGTATCAAGAAT	7.57	Orf, hypothetical protein	Orf; unknown
						−366	GATATTGCTCTGGTTCGCATAA	6.7		
b2997_hybABCDEF	2.47	**1.96**	—	**1.84**	—	−114	TATCATGATATCGATAACCATA	7.2	Putative hydrogenase subunit	Putative enzyme; not classified
						−53	TTTTGCGATTGAGGCAAAATAT	6.4		
b3001	2.62	1.42	↓	1.42	↓	−379	ATACGTGATGTACTCAGCAACA	5.15	Putative reductase	Putative enzyme; not classified
b3836_b3837_b3838_yigU_yigW_1	1.98	0.88	↓	1.2	↓				Orf, hypothetical protein	Orf; unknown
cchB_eutEJG	1.65	0.88	↓	1.03	↓				Detox protein	Phenotype; not classified
celABCDF_ydjC	2.05	0.92	↓	**1.63**	—	−201	AAATGTGAAGAGGGTCATAACC	11.91	PEP-dependent phosphotransferase enzyme IV for cellobiose	Enzyme; transport of small molecules: carbohydrates, organic acids, alcohols
csgDEFG	1.78	1.17	↓	1.12	↓	−201	TTTAGTTACATGTTTAACACTT	8.81	Putative 2-component transcriptional regulator for second curli operon	Putative regulator; not classified
						−35	AGGTGTGCGATCAATAAAAAAA	5.86		
						−329	TTATTAGTAAGTTATCACCATT	5.25		
cyoABCDE	2.13	1.23	↓	**2.46**	—	−245	ATAATTGTTTTATTTCACATTG	9.95	Cytochrome *o* ubiquinol oxidase subunit II	Enzyme; energy metabolism, carbon: aerobic respiration

Table 1. Continued

Operon name[a]	ROMA WT[b]	HL[c]	HL/wt[d]	KE[e]	KE/wt[f]	CRP-binding site Position[g]	Site sequence	Score[h]	Gene product	Gene function
						−91	GAATTTAATCATGTTTACAGTA	7.43		
dapB	3.28	1.18	↓	**2.08**	↓				Dihydrodipicolinate reductase	Enzyme; amino acid biosynthesis: Lysine
ecpD_htrE	1.9	1.51	—	1.43	—				Probable pilin chaperone similar to papd	Putative factor; surface structures
exuT	2.79	0.87	↓	1.33	↓	−152	ATTTGTGATGGCTCTCACCTTT	16.3	Transport of hexuronates	Transport; transport of small molecules: carbohydrates, organic acids, alcohols
						−274	TTTCGTGAGTTAGATCAATAAA	11.9		
fadBA	3.41	1.04	↓	**3.89**	—				3-Hydroxyacyl-coa dehydrogenase; 3-hydroxybutyryl-coa epimerase	Enzyme; degradation of small molecules: Fatty acids
fadD	1.91	1.15	↓	**1.79**	—	−131	AATAGTGACGCGCTTCGCAACC	8.72	Acyl-coa synthetase, long-chain-fatty-acid–coa ligase	Enzyme; degradation of small molecules: Fatty acids
fucAO	3.95	1.26	↓	**1.55**	↓	−361	TTATGTGACTACCATCACTTTA	16.9	L-fuculose-1-phosphate aldolase	Enzyme; degradation of small molecules: carbon compounds
						−144	TTAGTTGAACCAGGTCACAAAA	15.59		
						−54	TAGTGTGAAAGGAACAACATTA	12.46		
fucPI	10.79	**2.17**	↓	**4.03**	↓	−207	TAAAGTGATGGTAGTCACATAA	16.9	Fucose permease	Transport; transport of small molecules: carbohydrates, organic acids, alcohols
						−167	AAGTGTGACCGCCGTCATATTA	16.54		
fumA	4.27	1.52	↓	**2.1**	↓	−11	GTGAGAGAACAATGTCAAACAA	7.18	Fumarase A = fumarate hydratase Class I	Enzyme; energy metabolism, carbon: TCA cycle
gatYZA	2.38	1.42	↓	0.96	↓	−75	TTTTGTGATCGTTATCTCGATA	13.62	Tagatose-bisphosphate aldolase 1	Enzyme; degradation of small molecules: carbon compounds
gef	2.31	1.26	↓	**2.5**	—				Gef protein interferes with membrane function when in excess	Membrane; cell killing
glcC	2.33	**2.38**	—	**1.81**	—	−95	ATGTTAAATTGATGTAACATAA	6.93	Transcriptional activator for glc operon	Regulator; degradation of small molecules: carbon compounds
gntP	3.32	1.16	↓	0.97	↓	−90	GGATGTGACATTCATCGCAACA	9.87	Gluconate transport system permease 3	Transport; transport of small molecules: carbohydrates, organic acids, alcohols
						−69	AATGGTTGACCAATTTACATAA	6.74		
grpE	3.38	1.25	↓	**1.77**	↓	−293	GAGCGTGCCAGTTTTCACATTC	7.54	Phage lambda replication; heat shock protein	IS, phage, Tn; Phage-related functions and prophages
hrsA_ybgG	2.07	1.67	—	1.26	↓	−21	TCAAGTGAAATTGATCACATAA	11.07	Protein modification enzyme, induction of ompc	Enzyme; proteins: translation and modification
						−321	TTCTTTGACGTAAGTCCCGCTG	5.31		
hydHG	2.68	1.32	↓	1.25	↓	−255	TTTCGTGTTCCGTTTCATGGTT	6.48	Sensor kinase for hydg, hydrogenase 3 activity	Enzyme; energy metabolism, carbon: fermentation
inaA	4.93	**1.73**	↓	**2.19**	↓				Ph-inducible protein involved in stress response	Phenotype; adaptations, atypical conditions
kbl_tdh	1.66	1.02	↓	**1.73**	—				Glycine acetyltransferase	Enzyme; central intermediary metabolism: pool, multipurpose conversions

Table 1. Continued

Operon name[a]	ROMA WT[b]	HL[c]	HL/wt[d]	KE[e]	KE/wt[f]	CRP-binding site Position[g]	Site sequence	Score[h]	Gene product	Gene function
kdgT	4.13	1.27	↓	1.44	↓	−171	TTTTGTGATCAATTTCAAAATA	17.05	2-Keto-3-deoxy-D-gluconate transport system	Transport; transport of small molecules: carbohydrates, organic acids, alcohols
kdgT						−111	TGATGTGGTTTTGATCACTTTT	13.87		
leuLABCD	1.92	1.16	↓	1.45	—				Leu operon leader peptide	Leader; amino acid biosynthesis: Leucine
marRAB	1.93	0.9	↓	**2.69**	—	−147	AAACGTGGCATCGGTCAATTCA	7.43	Repressor of mar operon	Regulator; drug/analog sensitivity
						−61	TAATGTGAAAAGTACCAGCGAT	6.74		
mgtA	2.02	1.02	↓	**1.9**	—				$Mg2^+$ transport atpase, P-type 1	Enzyme; transport of small molecules: cations
mhpR	1.82	1.33	↓	1.23	↓				Transcriptional regulator for mhp operon	Regulator; not classified
modABC	1.59	0.88	↓	**1.99**	—	−100	TTTTCTTATCTACCTCACAAAG	9.06	Molybdate transport permease protein	Transport; transport of small molecules: anions
nadA_pnuC	2.1	1.14	↓	**2.28**	—				Quinolinate synthetase, A protein	Enzyme; biosynthesis of cofactors, carriers: pyridine nucleotide
ndk	2.24	0.82	↓	**2.3**	—	−173	GACAGTGAAATTTGTCATGCAA	5.25	Nucleoside diphosphate kinase	Enzyme; purine ribonucleotide biosynthesis
nrdAB_yfaE	2.29	0.8	↓	1.4	↓	−257	AACAGTTATTTTTAACAAATTT	6.49	Ribonucleoside diphosphate reductase 1, alpha subunit, B1	Enzyme; 2′-deoxyribonucleotide metabolism
pflA	2.03	1.66	—	1.05	↓				Pyruvate formate lyase activating enzyme 1	Enzyme; energy metabolism, carbon: anaerobic respiration
ribB	2.18	1.12	↓	**1.5**	—				3,4 Dihydroxy-2-butanone-4-phosphate synthase	Enzyme; biosynthesis of cofactors, carriers: Riboflavin
rpiR_yjcXWVUTS	2.69	**2.13**	—	1.36	↓	−285	ATTTGTGATGTTAATGAATTAA	6.38	Transcriptional repressor of rpib expression	Regulator; central intermediary metabolism: non-oxidative branch, pentose pathway
						−47	TTTTGTGAAGTCGCCAGCATCT	5.93		
						−103	AAATTTTAAGCCACTCGCCATT	5.47		
sbmC	2.3	1.33	↓	**2.02**	—	−93	GAGTGCGAGTCTGCTCGCATAA	8.91	Sbmc protein	Orf; unknown function
sdaC_sdaB_exo	13.43	**2.01**	↓	**3.89**	↓	−179	ATTTGAGATCAAGATCACTGAT	14.46	Probable serine transporter	Putative transport; transport of small molecules: Amino acids, amines
seqA_pgm	2.46	0.89	↓	1.4	↓				Negative modulator of initiation of replication	Regulator; DNA-replication, repair, restriction/modification
sodA	1.98	1.23	↓	**1.51**	—	−161	GTGGGTGATTTGCTTCACATCT	13.61	Superoxide dismutase, manganese	Enzyme; detoxification
speAB	2.62	1.15	↓	**2.15**	—	−45	AGTCGTTAACTGTTTTACACTT	8.81	Biosynthetic arginine decarboxylase	Enzyme; central intermediary metabolism: polyamine biosynthesis
speF_potE	1.99	0.87	↓	0.78	↓	−11	AGAGATGAAAAATGTCAAAATT	7.64	Ornithine decarboxylase isozyme, inducible	Enzyme; central intermediary metabolism: polyamine biosynthesis
srlAEBD_gutM_srlR_gutQ	3.07	**1.97**	↓	**1.62**	↓	−93	TTTTGCGATCAAAATAACACTT	13.12	PTS system, glucitol/sorbitol-specific IIC component	Transport; transport of small molecules: carbohydrates, organic acids, alcohols

Table 1. Continued

Operon name[a]	ROMA WT[b]	HL[c]	HL/wt[d]	KE[e]	KE/wt[f]	CRP-binding site Position[g]	Site sequence	Score[h]	Gene product	Gene function
						−394	CATTGTGCTGCGGCTCAAGCAG	5.63		
thrS_infC_rpmI_rplT	1.85	1.3	↓	1.43	—	−322	TGCTGTAAATCCGCTCGCAGTA	7.27	Threonine trna synthetase	Enzyme; aminoacyl tRNA synthetases, tRNA modification
						−230	AATTGCGAATCGAATCAATGTG	6.87		
tpiA	2.06	1.11	↓	1.08	↓				Triosephosphate isomerase	Enzyme; energy metabolism, carbon: glycolysis
ubiB	1.83	1.03	↓	1.49	—				NADPH:flavin oxidoreductase	Enzyme; energy metabolism, carbon: electron transport
ucpA	1.94	1.11	↓	1.22	↓	−291	AACGGTGATTGCCATTACTTCT	7.62	Putative oxidoreductase	Putative enzyme; not classified
						−27	ATTAGTGAGCTGATCCGCAGCA	5.76		
uidR	2.01	1.04	↓	**1.83**	—				Repressor for uid operon	Regulator; degradation of small molecules: carbon compounds
uspA	1.99	1.42	↓	1.01	↓				Universal stress protein; broad regulatory function?	Putative regulator; adaptations, atypical conditions
uxaB	2.46	1.3	↓	1.47	↓	−145	AACCATGATCCGCGCCACACTT	10.86	Altronate oxidoreductase	Enzyme; degradation of small molecules: carbon compounds
wzzB	2.09	1.69	—	**3.14**	—	−73	TTTTGTTACTTACACAACAATT	9.43	Regulator of length of O-antigen component of lipopolysaccharide chains	Regulator; outer membrane constituents
yaaF	4.37	1.29	↓	**2.45**	↓	−83	ATTCGTGAAGTCGATTAAGTCA	7.31	Orf, hypothetical protein	Orf; unknown
yabQ	1.94	1.25	↓	1.07	↓	−279	TATCGAGTTAAGTGTCACTTTT	7.69	Orf, hypothetical protein	Orf; unknown
						−149	GATTGTTACCTTGGTAAAAAAA	6.8		
yacL	1.96	1.59	—	1.1	↓				Orf, hypothetical protein	Orf; unknown
yadN	1.83	1.39	↓	**1.53**	—	−342	TTCTGAAAATCATATCTCATCT	8.46	Putative fimbrial-like protein	Putative structure; not classified
						−198	AATTTTGAATTAAGTAAATTTA	8.44		
yaeH	2.61	1.19	↓	1.28	↓				Putative structural protein	Putative structure; not classified
yahM	2.2	1.66	—	**1.99**	—	−68	AAATATGCTGTAAGGCTCATAT	8.82	Orf, hypothetical protein	Orf; unknown
						−260	TGTTGTGATCGGCGACACTTCG	6.52		
yahO	1.86	1.52	—	**1.94**	—	−127	GGGTGCGAGAGAGATCACAAAG	8.62	Orf, hypothetical protein	Orf; unknown
						−188	CATAGTGATTTCATCCATAAAT	5.66		
yaiB_phoA	1.83	1.6	—	**1.54**	—	−323	AGATGTGCGCAAGATCACAAAA	15.37	Orf, hypothetical protein	Orf; unknown
						−41	TAATGTTAACTTCTCCATACTT	8.63		
						−341	GAAAGTGCGTTATCTCAAAGAT	7.24		
yajD	2.31	1.2	↓	1.23	↓				Orf, hypothetical protein	Orf; unknown
ybbB	1.86	**1.75**	—	1.46	—	−144	TATTGTGACCTTGTTTACCCAG	10.73	Putative capsule anchoring protein	Phenotype; Surface polysaccharides and antigens
ybeL	2.3	1.16	↓	1.01	↓				Putative alpha helical protein	Phenotype; not classified
ybfP	1.81	1.16	↓	1.01	↓				Putative pectinase	Putative enzyme; not classified
ybhD	1.92	1.22	↓	**1.96**	—				Putative transcriptional regulator LYSR-type	Putative regulator; not classified
ybhJ	1.99	1.34	↓	**2**	—				Putative enzyme	Putative enzyme; not classified
ybhK	2.39	1.32	↓	1.43	↓				Putative structural protein	Putative structure; not classified
ybiH_b0795_ybhFSR	1.98	1.28	↓	**1.64**	—	−111	TTTTGTGACGCAGCGCATAAAT	12.53	Putative transcriptional regulator	Putative regulator; not classified
ybiP	2.42	0.99	↓	**3.09**	—	−190	ATTTGCAATCCCCTTCGCAAAA	8.89	Putative enzyme	Putative enzyme; not classified
ybiP						−325	TTGTGCGATATGGTCAACAACA	6.83		Putative enzyme; not classified
						−218	TGGTCTGAAAAACCCCACTTTT	5.11		
ycdZ	4.78	1.49	↓	**2.08**	↓	−55	AAGTGTGATCTACGTCACTCAT	19.82	Orf, hypothetical protein	Orf; unknown
						−4	AAATGTGTGCTCGATCTCATTC	13.87		

Table 1. Continued

Operon name[a]	ROMA WT[b]	HL[c]	HL/wt[d]	KE[e]	KE/wt[f]	CRP-binding site Position[g]	Site sequence	Score[h]	Gene product	Gene function
ychH	5.63	**4.5**	—	**2.1**	↓	−105	AATTGTGATCACGCCCGCACAT	12.05	Orf, hypothetical protein	Orf; unknown
ydcH	1.8	1.36	↓	1.38	—	−153	AAACACGATCCCGCTCGCATTT	8.83	Orf, hypothetical protein	Orf; unknown
ydeD	1.91	0.96	↓	**1.96**	—				Orf, hypothetical protein	Orf; unknown
ydeF	2.2	0.87	↓	**3.36**	—				Putative transport protein	Putative transport; not classified
ydeI	2.04	1.16	↓	1.32	↓	−173	ATGTTTAAGACTTTTCAATCTT	5.79		Orf; unknown
						−218	CAATATAATAATTTTCAAACAT	5.57		
yeaA_b1777	1.86	0.98	↓	**2.08**	—	−137	AAATGTGATTTTCATCACGATT	19.64	Orf, hypothetical protein	Orf; unknown
						−34	TAATGTGAGCACGATTAAAGTG	10.37		Orf; unknown
yedEF	2.65	0.94	↓	**1.73**	↓	−109	AAATTTGATCTAACTATTATTT	9.78	Orf, hypothetical protein	Putative transport; not classified
						−204	AATTGCGAGTTTTTTCATCATT	7.64		
yedK	3.37	1.04	↓	**2.06**	↓				Orf, hypothetical protein	Orf; unknown
yedL	2.28	1.41	↓	**1.61**	—	−64	CTATTAGACAAAGATTTCATTA	6.74	Orf, hypothetical protein	Orf; unknown
yedM	2.37	1.03	↓	1.22	↓	−74	ATTTTTGATAGTGGTAATTTTA	5.6	Orf, hypothetical protein	Orf; unknown
yedN_b1933	2.41	1.29	↓	1.47	↓	−27	AAATGTGTTGAACTCCACAATA	13.14	Orf, hypothetical protein	Orf; unknown
						−271	TTAAGTTATTGCACTCAGAATA	6.04		
yeeA	3.14	1.14	↓	**2.14**	—	−34	ATGAGCGCTTTTAATCTCATTA	7.01	Orf, hypothetical protein	Orf; unknown
						−273	TTATTTGAACAATGGCGCGGAA	5.32		
yeeX	2.46	1.26	↓	**1.82**	—				Putative alpha helix protein	Phenotype; not classified
yejK	1.87	1.3	↓	1.03	↓	−192	ATTTGTGGCATAAATCGAAATC	8.6	Orf, hypothetical protein	Orf; unknown
						−112	ATGCGTTAAATCATTCTCGTTA	6.88		
yfaD	3.29	1.61	↓	**2.03**	↓	−367	AAAAGAGAACTACCCCACCTCA	6.94	Orf, hypothetical protein	Orf; unknown
yfaH	4.63	1.2	↓	**1.96**	↓				Orf, hypothetical protein	Orf; unknown
yfaO	2.35	1.46	↓	**1.6**	—	−80	AAAAGAGATTACTGTCACTTTC	8.84	Orf, hypothetical protein	Orf; unknown
yfeA	2.18	0.86	↓	1.11	↓				Orf, hypothetical protein	Orf; unknown
yfeCD	3.63	**4.08**	—	**2.54**	—	−84	AGTAGTTATTCATGTCACGGTT	8.3	Orf, hypothetical protein	Orf; unknown
yfiD	1.83	1.01	↓	0.99	↓	−125	TTTATTGATTTAAATCAAAGAT	11.26	Putative formate acetyltransferase	Putative enzyme; energy metabolism, carbon: anaerobic respiration
ygbLM	2.36	**1.91**	—	1.36	↓	−4	TATCATGAGCGATTTCGCAAAA	8.37	Orf, hypothetical protein	Putative enzyme; not classified
ygbLM						−285	AACAGCAATACGGTGCACAAAA	6.03		
ygcB	1.99	1.29	↓	1.13	↓	−70	ATTTATGAGCAGCATCGAAAAA	8.23	Orf, hypothetical protein	Orf; unknown
yggG	1.85	1.07	↓	**1.97**	—				Orf, hypothetical protein	Orf; unknown
yghA	2.96	1.47	↓	1.4	↓				Putative oxidoreductase	Putative enzyme; not classified
ygjU	2.49	0.86	↓	**2.53**	—	−152	AAATATGACCTCTCTTTAAAAT	7.48	Putative transport protein	Putative transport; not classified
ygjV	2.1	1.06	↓	**2.28**	—				Orf, hypothetical protein	Orf; unknown
yhaB_yhaC	1.89	1.39	↓	**1.66**	—	−203	TTTTGTGTTGAGGATCACAAAA	16.06	Orf, hypothetical protein	Orf; unknown
yhcB	1.86	1.18	↓	1.05	↓				Orf, hypothetical protein	Orf; unknown
yhcM	1.94	1.18	↓	1.15	↓				Orf, hypothetical protein	Orf; unknown
yhdG_fis	1.79	1.08	↓	**2.74**	—	−297	AAGTGCGAGCAAGCTCACAAAA	15.81	Putative dehydrogenase	Putative enzyme; not classified
yhdG_fis						−212	AATTGAGAACTTACTCAAATTT	14.56		
yhdJ	2.11	1.31	↓	**2.34**	—	−25	CTTTTTTATAGGTGTCACAAAG	6.42	Putative methyltransferase	Putative enzyme; not classified
yhdU	1.88	1	↓	**1.89**	—	−113	GAAGTTGACCCCGATCTCATTA	11.45	Orf, hypothetical protein	Orf; unknown
yhfMNO	2.18	1.24	↓	1.22	↓	−75	TTTTGTGATCTTCCTCCACATT	7.75	Putative amino acid/amine transport protein	Putative transport; not classified
						−40	TGTTGTTATATTCATCATGCAT	6.95		
yhfPQR	1.73	1.21	↓	0.97	↓	−296	AAAAGCGGGCGGCATCAAACAA	7.02	Orf, hypothetical protein	Orf; unknown
yhfZY	3.37	1.04	↓	**3.08**	—	−314	AAATGTGAAGTGCCTCGCCGTT	13.39	Orf, hypothetical protein	Orf; unknown
yhiP	3.29	**3.03**	—	1.41	↓	−180	AACATTTAACACCATCATATTT	5.04	Putative transport protein	Putative transport; not classified
yhjJ	2.22	0.97	↓	**1.54**	—				Orf, hypothetical protein	Orf; Unknown
yi81_1	2	1.28	↓	1.4	↓	−76	TAAAGTGGCGGGGATCACTCCC	6.61	IS186 hypothetical protein	IS, phage, Tn; transposon-related functions

Table 1. Continued

Operon name[a]	ROMA WT[b]	HL[c]	HL/wt[d]	KE[e]	KE/wt[f]	CRP-binding site Position[g]	Site sequence	Score[h]	Gene product	Gene function
yi81_2	1.88	0.82	↓	1.27	↓	−185	AACGGTGCTGGGATTTACGCTT	5.95	IS186 hypothetical protein	IS, phage, Tn; transposon-related functions
yi82_1	2.01	1.09	↓	1.34	↓				IS186 and IS421 hypothetical protein	IS, phage, Tn; transposon-related functions
yidE	2.95	0.84	↓	1.03	↓	−58	ATGTGCGCTATAAGGCAAATCT	5.7	Putative transport protein	Putative transport; not classified
yidP	2.09	1.33	↓	1.26	↓	−185	ATTCGTGATCGCTTTCATGCTT	10.02	Putative transcriptional regulator	Putative regulator; not classified
yidW_b3694_dgoKA	7.82	1.52	↓	**2.75**	↓	−154	TTTTGTGATCTAAATTGTAGTA	10.24	Regulator protein for dgo operon	Putative regulator; not classified
						−96	AAACGCGAACGTCATCACGCTG	8.31		
yiePO	1.85	0.81	↓	**1.52**	—	−213	TTTTGCAACCGTAATCACACTT	12.09	Orf, hypothetical protein	Orf; unknown
						−243	TAATGTGCCATAAAACAAGCAA	8.42		
yigC	2.51	1.15	↓	**1.89**	—	−370	AGTCGTGGTATGAATCACTTCT	8.59	Putative oxidoreductase	Putative enzyme; not classified
yihVWX_rbn_yihZ_ yiiD	2.17	**1.87**	—	1.03	↓	−70	AAAATTGACAGCCGTCACTTTT	11.71	Putative kinase	Putative enzyme; not classified
yiiM	2.09	1.42	↓	1.38	↓				Orf, hypothetical protein	Orf; unknown
yjaE	1.82	1.4	↓	**2.1**	—				Putative transcriptional regulator	Putative regulator; not classified
yjeM	2.35	0.93	↓	1.31	↓	−75	ATATTTGATATCATCCAGGTAT	6.84	Putative transport	Putative transport; not classified
yjeNO	1.87	0.93	↓	1.24	↓	−217	ATTTGCGAACGTCTTCACCATC	8.45	Orf, hypothetical protein	Orf; unknown
						−283	GTTTGTGATTTTCAAAACGCAT	6.6		
yjfG	2.64	1.46	↓	1.45	↓	−34	AAATGAGCGGCAGATTAAAAAA	9.06	Putative ligase	Putative enzyme; not classified
yjhBC	1.92	1.19	↓	1.26	↓	−109	TAGAGTGAAATATGTTAAGAAG	5.44	Putative transport protein	Putative transport; not classified
yjhQP	3.42	1.05	↓	**4.23**	—				Orf, hypothetical protein	Orf; unknown
yjiMLKJ	3.04	1.37	↓	1.38	↓	−70	AACCGCGAGAGAGATCAAATAA	10.71	Orf, hypothetical protein	Orf; unknown
yjjM	2.81	1.27	↓	1.35	↓	−173	TTTTGTGAAAACACACGCATAA	8.94	Orf, hypothetical protein	Orf; unknown
yjjY	1.9	0.84	↓	1.08	↓				Orf, hypothetical protein	Orf; unknown
ynfM	2.12	1.32	↓	1.22	↓	−94	TTATTTGAGATTATTAATATAT	8.26	Putative transport protein	Putative transport; not classified
yqeA	1.93	0.95	↓	0.97	↓				Putative kinase	Putative enzyme; not classified
yqgA	2.37	0.81	↓	**1.6**	—				Putative transport protein	Putative transport; not classified
yqgB	2.97	1.25	↓	**3.49**	—				Orf, hypothetical protein	Orf; unknown
yqhA	2.35	1.28	↓	1.44	↓				Orf, hypothetical protein	Orf; unknown
yqiEB_icc_yqiA_parE	1.89	1.24	↓	1.29	↓	−51	AACGTCGCTGAAATTCACATTT	6.03	Orf, hypothetical protein	Orf; unknown
ytfQ	3.38	1.21	↓	**2.03**	↓	−98	AAGTGTGATGTAACGCAATCTG	8.65	Putative LACI-type transcriptional regulator	Putative regulator; not classified
ytfRST_yjfF	2.1	0.96	↓	**1.58**	—				Putative ATP-binding component of a transport system	Putative transport; not classified
zipA	2.16	1.69	—	1.12	↓	−250	AGATGTGAATGATGAAACCATA	6.28	Cell division protein involved in ftsz ring	Membrane; cell division
						−313	TAGTGTAGAGCAGAAAACAAAA	5.88		

[a]Gene or genes of an operon that are activated by wild-type CRP in ROMA experiment.
[b]Average ratios of RNA transcripts in wild-type CRP reaction versus control reaction.
[c]Average ratios of RNA transcripts in HL159-CRP reaction versus control reaction. Operons that are still significantly activated by HL159-CRP (outliers) are indicated in boldface.
[d]Significance of ratio change between HL159-CRP and wild-type CRP experiments. Operons that are not activated by HL159-CRP or regulated in a reduced level (outliers) are indicated by '↓', and operons that are activated by HL159-CRP at the same level as by wild-type CRP indicated by '—'.
[e]Average ratios of RNA transcripts in KE101-CRP reaction versus control reaction. Operons that are still significantly activated by KE101-CRP (outliers) are indicated in boldface.
[f]Significance of ratio change between KE101-CRP and wild-type CRP experiments. Operons that are not activated by KE101-CRP or regulated in a reduced level (outliers) are indicated by '↓', and operons that are activated by KE101-CRP at the same level as by wild-type CRP indicated by '—'.
[g]The position of 5′ end of a CRP site relative to corresponding translation start site.
[h]A site score was calculated from Positional Weight Matrices generated by Tan *et al.* (5) to indicate the conservation of a site sequence. The higher the score the better match to the consensus sequence.

Table 2. Operons which have been experimentally verified as being regulated by CRP

Operon name	ROMA					CRP sites			Other regulators[b]	Gene function	Reference
	WT	HL	HL/wt	KE	KE/wt	Position[a]	Sites	Score			
bglG	1.96	0.85	↓	1.73	—	−61.5	AACTGCGAGCATGGTCATATTT	11.2	Fis	Positive regulation of bgl operon	(49)
dctA	2.44	1.09	↓	1.85	—	−81.5	TTGTGCGAGCCAGCTCAAACTT	14.1	DcuR, ArcA	Uptake of C4-dicarboxylic acids	(50)
rbsDACBKR	3.80	0.89	↓	4.10	—	−61.5	CGTTTCGAGGTTGATCACATTT	9.4	RbsR	D-ribose high-affinity transport system	(51)
rhaSR	1.46	0.86	↓	1.66	—	−113.5	TGATGTGATGCTCACCGCATTT	12.4	RhaR	Positive regulator for rhabad operon	(52)
sdhCDAB	2.71	1.11	↓	2.38	—	−83.5	TATCGTGACCTGGATCACTGTT	16.4	ArcA, FNR	Succinate dehydrogenase, cytochrome b556	(53)
tnaLAB	1.98	0.95	↓	2.08	—	−61.5	GATTGTGATTCGATTCACATTT	19.6		Tryptophanase leader peptide	(54)
treBC	3.09	1.35	↓	4.73	—	−60.5	AATTGTGATCTTCGCTGCGTTT	8.8		PTS system enzyme II, trehalose specific	(55)
focA_pflB	3.90	1.59	↓	1.39	↓	−41.5	AGATATGATCTATATCAATTTC	10.1	FNR	Probable formate transporter	(43)
galS	2.49	1.22	↓	1.09	↓	−41.5	TGCTGTGACTCGATTCACGAAG	10.4	GalR, GalS	Mgl repressor, galactose operon inducer	(56)
glpTQ	4.51	0.89	↓	2.21	↓	−41.5	ATGTGTGCGGCAATTCACATTT	17.2	GlpR	Sn-glycerol-3-phosphate permease	(57)
malXY	6.24	1.71	↓	2.94	↓	−49.5	TTATGTGACAGATAAAACGTTT	11.2		PTS system, maltose and glucose-specific II ABC	(58)
melR	2.29	1.20	↓	1.17	↓	−41.5	AACCGTGCTCCCACTCGCAGTC	5.5	MelR	Regulator of melibiose operon	(59)
mglBAC	6.78	1.49	↓	1.78	↓	−41.5	ATCTGTGAGTGATTTCACAGTA	16.6		Galactose-binding transport protein	(56)
ptsG	2.11	1.14	↓	1.13	↓	−40.5	AAACGTGATAGCCGTCAAACAA	14.3	Mlc	PTS system, glucose-specific IIBC component	(35)
ptsHI_crr	2.95	1.01	↓	1.56	↓	−42.5	TTTTATGATTTGGTTCAATTCT	9.5		PTS system protein hpr	(60)
yhfA	1.88	0.97	↓	1.10	↓	−44.5	TAATGTGACGTCCTTTGCATAC	11.7		Orf, hypothetical protein	(61)
aspA	6.29	2.50	↓	2.65	↓	−90.5	AGCGGTGATCTATTTCACAAAT	15.5	FNR	Aspartate ammonia-lyase (aspartase)	(62)
						−40.5	TAAAGTGATCCAGATTACGGTA	14.2			
deoCABD	3.22	2.37	—	2.37	—	−94.5	TTATTTGAACCAGATCGCATTA	17.1	CytR, DeoR	2-Deoxyribose-5-phosphate aldolase	(21)
						−41.5	AATTGTGATGTGTATCGAAGTG	11.9			
glpACB	4.38	1.72	↓	3.31	—	−90.5	AAATGTGAATTGCCGCACACAT	17.2	FNR, ArcA, FlhD, GlpR	Sn-glycerol-3-phosphate dehydrogenase	(57)
						−40.5	AATGACGCATGAAATCACGTTT	3.3			
manXYZ	4.26	2.59	↓	2.17	↓	−92.5	GAATGTGACAAGGATATTTTAC	2.4	Mlc, NagC	PTS enzyme IIAB, mannase-specific	(63,36)
						−40.5	ATTACGGATCTTCATCACATAA	7.6			
nupC	3.25	1.17	↓	1.39	↓	−89.5	AAATGTATGACAGATCACTATT	12.6	CytR	Permease of transport system for 3 nucleosides	(64)
						−40.5	TAGTGTGTGTCAGATCTCGTTT	12.6			
nupG	4.45	1.01	↓	3.25	—	−92.5	AAATGTTATCCACATCACAATT	20.4	CytR, DeoR	Transport of nucleosides, permease protein	(65)
						−42.5	TTATTTGCCACAGGTAACAAAA	10.6			
rpoH	2.18	1.75	—	1.15	↓	−93.5	ATTTCATCTCTATGTCACATTT	7.4	CytR	Sigma(32) factor	(22)
						−41.5	ACTTGTGGATAAAATCACGGTC	2.2			
tsx	3.96	1.19	↓	2.40	↓	−40.5	AACTGTGAAACGAAACATATTT	12.9	CytR	Receptor of phage T6 and colicin K	(66)
						−74.5	AAACGTGAACGCAATCGATTAC	5.8			

[a]The centre of an experimentally verified CRP-binding site relative to corresponding transcription start site.
[b]Factors that regulate transcription in addition to CRP.

CRP-binding sites within 11 of the 16 promoter regions, 8 operons contain a strong CRP-binding site (>10) and 3 operons contain a weak site (5–10). The *yjcB* and *ycfR* operons contain a CRP site almost identical to the consensus.

Of the 16 repressed operons, promoter regions of two, *nirB* and *pncB*, have been determined experimentally (23,24). Sequence analysis indicates that their CRP-binding sites overlap the binding sites for RNAP (Figure 3) and it is therefore probable that CRP represses transcription at the *nirB* and *pncB* promoters by either blocking the interaction between RNAP and the promoter DNA, or blocking transcription from an alternative upstream promoter. The position of putative CRP-binding sites relative to the RNAP-binding site was determined for the other repressed operons (Figure 3). The position of the CRP site is variable, but at most promoters the CRP site overlaps the region between the −35 and −10 hexamers. Therefore, repression at these operons might also involve a simple blocking mechanism. However, at the *metK*, *ybiS* and *yjcB* predicted promoter regions, the CRP sites are located 3 bp upstream of the −35 hexamer. The *yjcB* promoter contains a very strong CRP-binding site (score = 22). The promoter

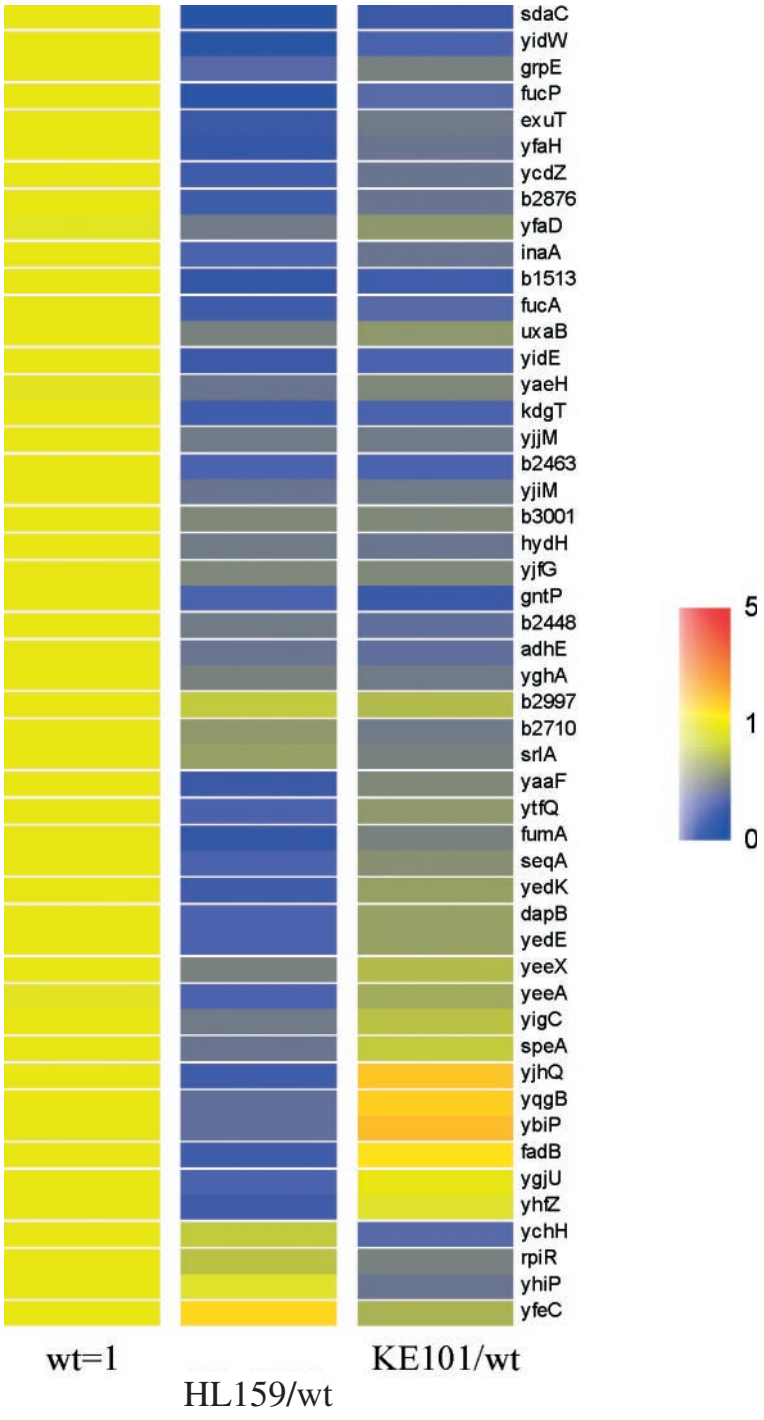

Figure 2. Regulatory map of 50 highest CRP-regulated operons. The figure indicates the activation by HL159-CRP (middle) and KE101-CRP (right) relative to wild-type CRP (=1), generated by GeneSpring clustering in distance order, so that those with similar regulation pattern were grouped together. The relative activation changes are colour coded: yellow indicates that a gene is regulated at the same level as wild-type CRP, blue indicates that a gene is regulated at a much lower level than wild-type CRP and red indicates that a gene is regulated at a higher level.

architecture might be such that the CRP-binding site still stops productive binding of RNAP.

Identification of CRP regulon by *in vivo* transcriptional profiling

An *in vivo* microarray experiment was also designed to identify CRP-regulated genes. Both wild-type and Δ*crp* strains were grown in minimal media containing fructose and then pulsed with glucose, and the transcriptome before the addition of glucose for both strains was compared with that after glucose addition. In the presence of glucose, the cAMP level is expected to be decreased, and hence the cAMP CRP level, thus the CRP-dependent regulation is repressed. However, glucose also affects gene expression through

Table 3. Operons repressed by wild-type CRP in ROMA experiment

Operon name	ROMA					CRP-binding site			Product
	wt	HL	HL/wt	KE	KE/wt	Position[a]	CRP sites	Score	
yjcB	0.34	0.36	—	0.40	—	−80	AATTGTGATATAGTTCACAAAA	22.26	Orf, hypothetical protein
ycfR	0.38	0.48	—	0.46	—	−87	GTATGTGATCCAGATCACATCT	20.52	Orf, hypothetical protein
						−136	AAATTTAAAGATTTTTAAATTA	6.68	
metK	0.40	0.54	—	0.52	—	−200	GAATGAGACACGATTCAAAAAA	12	Methionine adenosyltransferase 1
nirB_nirD	0.42	0.54	—	0.50	—	−76	GAATTTGATTTACATCAATAAG	8.5	Nitrite reductase (NAD(P)H) subunit
						−331	CTTTGTGATGTGCTTCCTGTTA	6.13	
ybiS	0.43	0.56	—	0.48	—	−116	AAATGTGATTTCGTACACATCT	16.59	Orf, hypothetical protein
						−201	AGATATGACAAACCGCGCATTA	6.75	
ykfC	0.45	0.44	—	0.46	—				Orf, hypothetical protein
yhfC	0.46	0.48	—	0.64	—	−141	TTCCGTGATCAAAATCACCTCT	12.33	Putative transport
						−88	AACATTTAAACAGATCACAAAA	10.22	
rbfA_truB	0.49	0.72	—	0.66	—				Ribosome-binding factor A
panF_prmA	0.52	0.81	—	0.64	—				Sodium/pantothenate symporter
amiC	0.53	0.89	↑	0.71	—				*N*-acetyl-muramyl-L-alanine amidase
ydfK	0.54	0.94	↑	0.56	—	−195	AATTGTCAACTATATCATATAT	10.99	Orf, hypothetical protein
yeeF	0.54	0.80	—	0.65	—	−252	TATTCTGACAAGCCTCTCATTC	8.98	Putative amino acid/amine transport protein
						−29	TTACGCGACGGTTATCACCGTA	8.17	
ygfJ	0.54	0.92	↑	0.75	—				Orf, hypothetical protein
ynaE	0.57	0.78	—	0.57	—	−195	AATTGTCAACTATATCATATAT	10.99	Orf, hypothetical protein
pncB	0.58	0.63	—	0.64	—	−106	TGGTGTGATCGGGGTTCAATAA	7.19	Nicotinate phosphoribosyltransferase
						−276	TGTTGAGTCATAAATAACCTTT	5.41	
apbA_yojL	0.58	0.96	↑	0.61	—	−151	ATTTTTGATGCGAAGCATAATA	10.45	Involved in biotin biosynthesis

[a]The position of 5′ end of a CRP site relative to corresponding translation start site.

other mechanisms, thus glucose effects will be observed in the Δ*crp* strain. The CRP-regulated genes in this study were defined as genes regulated by glucose in the wild-type strain, but not regulated in the Δ*crp* strain. In total, we identified only 17 operons repressed by glucose (CRP-activated operons) and six glucose-activated operons (CRP-repressed operons), listed in Table 4. In the 17 CRP-activated operons, 9 operons are known members of the CRP regulon and 7 operons were activated by CRP in the ROMA experiment. Our *in vivo* experiments failed to identify most CRP-regulated genes.

DISCUSSION

The combination of run-off transcription and microarray analysis (ROMA) has shown advantages in defining σ^W-dependent promoters in *B.subtilis* (12). In this study, we have established ROMA exploiting microarray glass slides and further demonstrated its application to investigate a specific transcription factor. The microarray glass slide has several advantages over a macroarray membrane. First, the slide array allows more genes to be studied in a single experiment (up to 50 000 genes) compared with a few thousand genes that can be printed on a macroarray membrane. Second, RNA from two different sources of interest can be dual fluorescence labelled and simultaneously hybridized on a single slide, whereas nylon arrays are generally probed in serial or parallel hybridization reactions to allow comparison. Third, the utilization of fluorescence instead of radioactive material is safer, easier to handle and produces higher resolution and lower background. Therefore, the establishment of ROMA on glass slides provides a convenient system to obtain high quality information. Our results have shown the high sensitivity and reproducibility of the system. However, our procedure does have some disadvantages compared with the original macroarray procedure. First, several purification steps are required which might result in the loss of short transcripts. Second, a large amount of template DNA and protein is required. Third, any trace of template DNA remaining after RNA purification will result in high background hybridization that might mask the actual induction level.

Using this newly established ROMA system, we have identified 152 novel CRP-dependent operons and 24 operons corresponding to previously known CRP regulons. Regulation at these operons was further verified by using two CRP mutants. Our experiments indicate that ROMA is a very powerful technique to identify direct regulation by a transcription initiation factor. The ROMA system was, however, unable to identify some known CRP-regulated operons. There are several reasons that could account for these false negatives. First, some operons require other factors for CRP to activate transcription, e.g. AraC at the *araBAD* promoter (25), which are missing in the ROMA system. Second, transcripts that can initiate from different promoters upstream from the same operon cannot be resolved. So genes where CRP alters the transcription start point will not be identified. For example, the *gal*P1 promoter is a well-known Class II CRP-dependent promoter. However, in the absence of CRP, transcription of the *gal* operon can initiate from the upstream *gal*P2 promoter (26), which will mask the weak transcription from *gal*P1. Third, certain promoter structures result in a situation where CRP can activate transcription by binding to an upstream site but can also repress transcription by binding to a downstream site, e.g. the *crp* promoter (27). In addition, some operons may not be transcribed efficiently due to DNA relaxation, or in the experimental system described here, the existence of an EcoRI restriction site within the promoter region or coding region. Finally, because of the necessity for data manipulation to identify only the most robust changes in transcription, promoters that are only weakly activated by CRP *in vitro* will not be detected. The CRP

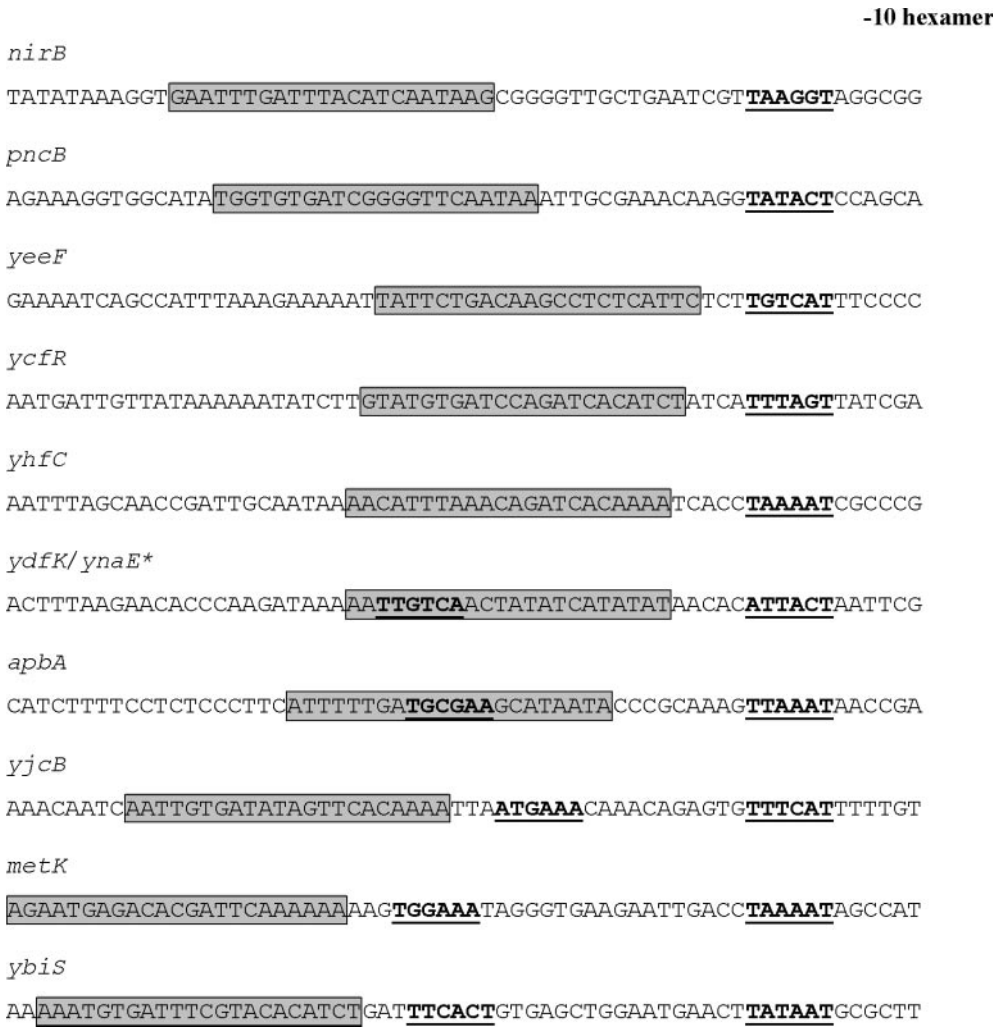

Figure 3. Non-template strand sequences of promoters repressed by CRP. Except *nirB* and *pncB*, promoter regions are predicted by searching the −10 and −35 sequences upstream of 10 down-regulated genes with a CRP site. The predicted −10 and −35 hexamers are underlined and in boldface with the −10 regions aligned together. The potential CRP-binding sites are shown as shaded boxes. *ydfK and ynaE are both associated with different prophage (Quin and Rac, respectively) but have essentially identical sequences.

regulon may therefore contain more operons than were identified in this study.

In the σ^W ROMA experiment, 50% of the genes identified using ROMA were also identified using consensus sequence searching and *in vivo* transcriptional profiling (12). However, in the case of CRP, these approaches have serious limitations. Comparison of the ability of each method to identify known members of the CRP regulon demonstrates that sequence searching generates the highest number of hits, and *in vivo* transcriptional profiling generates the fewest (Figure 4). A combination of the three methods can identify more than 80% of known operons, but failed to identify 14 operons known to be regulated by CRP. Although the sequence prediction generates the highest number of hits, it is probable that this includes many false positives and is therefore not necessarily the best approach. In addition, many weak sites are difficult to identify by sequence searching. Lowering the cut-off score would facilitate the identification of known sites but would inevitably lead to more false positives. Besides DNA sequence, CRP activation is dependent on the location of the binding site. Previous studies have indicated that at a Class I promoter, the CRP-binding site is normally positioned at −61.5, −71.5, −81.5 or −91.5, and no activation occurs when CRP sites are located further than −113.5, which might be beyond the reach of αCTD [reviewed in(28)]. Therefore, many *in silico* predicted sites might be null sites that are not functional during transcription. For promoters containing tandem CRP sites, the space between the two sites is an important determinant for synergistic activation (29–32). Therefore, although some operons contain several predicted binding sites, we hypothesized that at least some sites are not functional due to improper spacing. This may also explain why some operons with strong predicted binding sites were not regulated by CRP in our experiments. It is also probable that sequence searching generates many false negatives due to the ability of CRP to bind variable sequences. For example, previous work has identified the CRP sites within the *melR, rpoH* and *rbs* promoters that play important roles in regulation of these promoters (22). These sites have scores below 10 and were, therefore, not identified in the sequence search. It

Table 4. CRP-regulated genes[a] identified by *in vivo* transcription profiling

Operons	Genes	Glucose effect[b]		CRP-binding site			Product
		Wt strain	crp-strain	Position	Sequence	Score	
Glucose repression (CRP-activated operons)							
acs	acs	0.542	1.131	−100	TTGCGTGATCTGTCGCCCAAAT	8.47	Acetyl-CoA synthetase
aldA[c]	aldA	0.524	1.057	−112	TTTTATGAAGCCCTTCACAGAA	12.79	Aldehyde dehydrogenase, NAD-linked
fruBKA	fruB	0.218	0.515	−178	AATTGTGCAGCACATCAAACTT	15.13	PTS system, fructose-specific IIA/fpr component
	fruK	0.410	0.755				Fructose-1-phosphate kinase
gatYZABCD_gatR_2[c]	gatY	0.177	1.410	−75	TTTTGTGATCGTTATCTCGATA	13.62	Tagatose-bisphosphate aldolase 1
	gatZ	0.122	0.841	−30	TATTTTGAAATCGAAAACAAAC	6.62	Putative tagatose 6-phosphate kinase 1
	gatA	0.122	0.921				Galactitol-specific enzyme IIA of phosphotransferase system
	gatC	0.265	1.240				PTS system galactitol-specific enzyme IIC
	gatD	0.368	0.589				Galactitol-1-phosphate dehydrogenase
gcvP	gcvP	0.492	0.810				Glycine decarboxylase
glpFK[c]	glpF	0.137	1.096	−142	TTTTATGACGAGGCACACACAT	10.45	Facilitated diffusion of glycerol
	glpK	0.144	0.828				Glycerol kinase
glpTQ[c]	glpQ	0.154	0.785	−129	ATGTGTGCGGCAATTCACATTT	17.23	Glycerophosphodiester phosphodiesterase, periplasmic
	glpT	0.509	1.390				sn-Glycerol-3-phosphate permease
mglBAC[c]	mglB	0.519	0.958	−270	ATCTGTGAGTGATTTCACAGTA	16.64	Galactose-binding transport protein; receptor for galactose taxis
	mglA	0.400	0.934				ATP-binding component of methyl-galactoside transport and galactose taxis
	mglC	0.549	0.967				Methyl-galactoside transport and galactose taxis
nmpC_trs5_2	nmpC	0.490	0.823	−69	AATAGAGATCTACTTCACAAAT	16.15	Outer membrane porin protein
ompF	ompF	0.518	2.915	−242	AAATATGACGGTGTTCACAAAG	12.53	Outer membrane protein 1a (Ia;b;F)
rbsDACBK	rbsD	0.426	0.973	−66	CGTTTCGAGGTTGATCACATTT	9.35	D-ribose high-affinity transport system
	rbsB	0.497	1.488				D-ribose periplasmic binding protein
ribB	ribB	0.505	1.119				3,4 Dihydroxy-2-butanone-4-phosphate synthase
tnaLAB[c]	tnaL	0.437	0.802	−94	GATTGTGATTCGATTCACATTT	19.57	Tryptophanase leader peptide
ybeK	ybeK	0.504	1.910	−97	AATTGCGCGCCATCTCACGCTT	10.64	Putative tRNA synthetase
yihQPO_yshA	yihQ	0.310	1.081	−76	TTATGAGAATCATTTTACATAA	14.51	Putative glycosidase
sdhCDAB_b0725_sucABCD[c]	sdhD	0.605	1.018	−313	TATCGTGACCTGGATCACTGTT	16.39	Succinate dehydrogenase, hydrophobic subunit
	sdhA	0.509	1.250				Succinate dehydrogenase, flavoprotein subunit
	sucC	0.494	1.101				Succinyl-CoA synthetase, beta subunit
	sucD	0.381	1.067				Succinyl-CoA synthetase, alpha subunit
Glucose induction (CRP-repressed operons)							
gcd	gcd	2.014	0.864	−80	AATTGTGATGACGATCACACAT	20.43	Glucose dehydrogenase
proP	proP	2.413	1.111	−227	ATGTGTGAAGTTGATCACAAAT	20.42	Proline permease II
ptsG[c]	ptsG	3.632	1.628	−154	AAACGTGATAGCCGTCAAACAA	14.25	PTS system, glucose-specific IIBC component
soda	sodA	2.012	1.174	−161	GTGGGTGATTTGCTTCACATCT	13.61	Superoxide dismutase, manganese
trpLEDCBA	trpL	2.200	1.087				Trp operon leader peptide
	trpE	2.806	1.061				Anthranilate synthase component I
	trpD	2.152	0.972				Anthranilate synthase component II, glutamine amidotransferase and phosphoribosylanthranilate transferase
	trpC	2.096	1.202				*N*-(5-phosphoribosyl)anthranilate isomerase and indole-3-glycerolphosphate synthetase
	trpA	2.144	0.982				Tryptophan synthase, alpha protein
yagU	yagU	1.872	1.022				Orf, hypothetical protein

[a]The CRP-regulated genes in this study were defined as genes regulated by glucose (i.e. the ratio of gene expression plus glucose divided by the gene expression minus glucose) in the wild-type strain, but not regulated in the Δ*crp* strain.
[b]The ratio of gene expression plus glucose divided by the gene expression minus glucose.
[c]Genes that were also identified as CRP-regulated by Gosset *et al*. (40).

is therefore possible that several promoters identified by ROMA that do not contain a putative CRP-binding site, e.g. *inaA* and *yfaH*, may still be regulated by CRP.

Comparing activation levels of operons with strong CRP sites to those with weak sites indicated that there is no significant relationship between the degree of activation and the 'quality' of CRP-binding sites. For example, the *focA-pflB* operon which is activated 3.9-fold contains a relatively weak CRP site (score = 10.1), whereas the *tnaL* operon which is activated 2-fold contains a very strong CRP-binding site (score = 19.6) (Table 2). Therefore, conservation of CRP-binding sequence is not sufficient to predict CRP activation levels.

In our study, the *in vivo* transcriptional profiles failed to identify many CRP-regulated genes. The main reason for this is the complexity of the CRP regulon. CRP activation at some promoters, such as *melAB* and *araBAD*, is dependent on the presence of an additional regulator (MelR and AraC, respectively) (33,25) that is only induced by a specific substrate. It is impossible to include all inducers in the growth media and, moreover, the presence of extra inducers may trigger expression of other genes that are independent on CRP, further

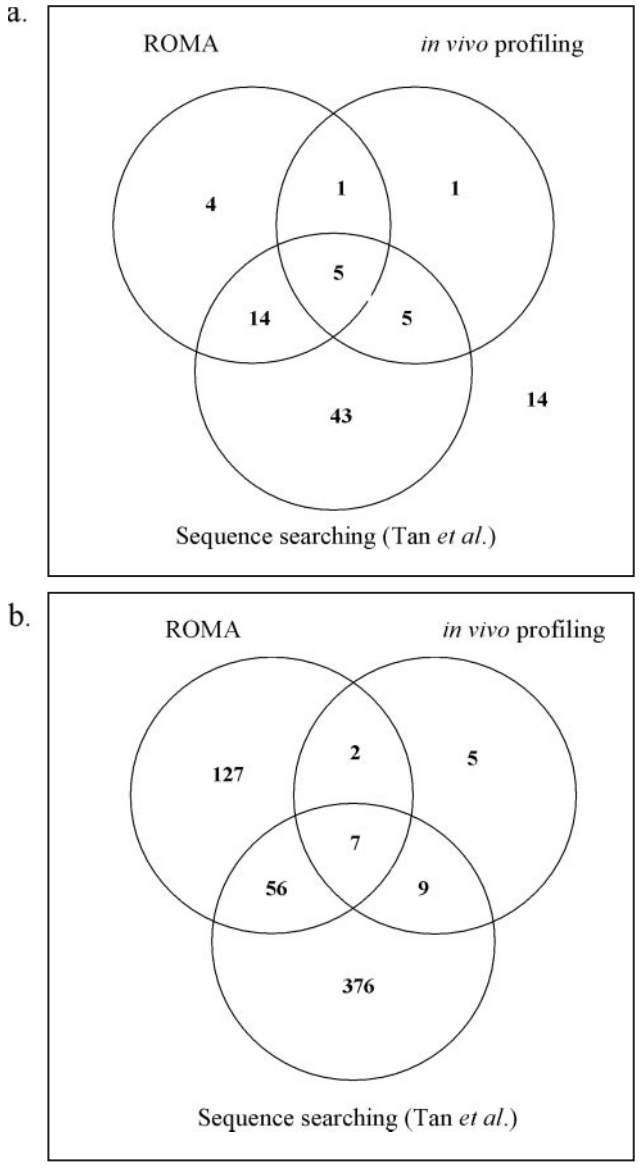

Figure 4. Venn diagram of the CRP regulon as identified by three genomic approaches: ROMA, *in vivo* transcriptional profiling, sequence search (5). The result for (**a**) 87 known (experimentally verified) CRP regulon collected in RegulonDB database and (**b**) all operons are shown. The number of operons identified by each approach is presented in a coloured circle. The numbers covered by more than one circles are operons identified by two or three methods.

complicating the interpretation of the data. In addition, most CRP-dependent promoters are subjected to repression by other factors, which will mask the CRP activation *in vivo*. For example, the *galE* promoter is repressed by GalR, the *lacZ* promoter repressed by LacI and many CRP-activated promoters are repressed by CytR (Table 2). Therefore, the variability of CRP regulation at different operons makes it difficult to design an *in vivo* microarray experiment to distinguish direct effects from indirect effects. ROMA is better suited for identifying the regulon of a factor that is part of a complex regulatory network. In the six glucose-induced operons, two operons, *gcd* and *proP*, are known CRP-repressed operons, which contain a strong CRP-binding site between the −10 and −35 regions, thus blocking RNAP binding. As observed in previous studies (34,35), the *ptsG* gene encoding the major glucose transporter was highly induced by glucose in the wild-type stain. Previous studies have indicated that the *ptsG* promoter, which contains a CRP activation site (see Table 2) (36), is both activated by CRP and strongly repressed by the pleiotropic transcriptional repressor Mlc. The glucose induction of *ptsG* is through a complicate Mlc-dependent mechanism that involves several layers of regulation (34–39). The case of *ptsG* demonstrates that an *in vivo* indirect effect can mask the real regulation by CRP, sometimes leading to the opposite conclusion. The regulation observed in the *trp* operon might also have resulted from an indirect effect, as there is no evidence from previous studies to indicate that CRP regulates this operon.

Further evidence of the limitations of *in vivo* transcription profiling of a complex regulon has come from a recent study of CRP using an Affymetrix *E.coli* array (40). Gosset *et al* completed a similar *in vivo* transcriptomic study of CRP and identified 39 operons under CRP-dependent glucose repression and 19 operons under CRP-dependent glucose activation. Among these operons, only seven glucose-repressed operons and one glucose-activated operon were identified in our *in vivo* transcriptional profiling (Table 4). However, direct comparison of the two data sets is difficult because of differences in the strain and the growth conditions used. Gosset *et al* in their study used the *E.coli* strain BW25113 and an isogenic *crp* mutant derivative and grew their cells in LB media with or without glucose. The difference in composition between the LB media used in the Gosset study and minimal media used in this study could contribute to a different profile of induction of some genes. For example, the presence of fructose in our growth conditions meant that the *fru* operon was identified as CRP regulated in our study, but not regulated in the Gosset study due to the lack of fructose in LB media. Gosset *et al*. (40) used long-term growth in glucose plus media rather than a glucose shock which was used in this study. Long-term growth in glucose should result in higher levels of induction but can also lead to a larger number of indirect effects, such as regulation of ribosomal protein-encoding genes, RNA-encoding genes, stress-related genes and temperature shock genes, which were significantly affected in their study. One-third of operons subject to CRP-dependent glucose repression (13/39) in Gosset *et al*. were activated by CRP in ROMA experiment. This figure is similar to the data obtained by *in vivo* transcription profiling in this study (7/17). This indicates that ROMA has limitations to detect a certain set of operons that may be subjected to complex regulation or regulated indirectly *in vivo*. However, as seen in this study, Gosset *et al*. failed to identify many CRP-regulated genes, such as *rpoH*, *melR* and *rbs*, so neither study produced a definitive list of members of the CRP regulon. This further indicates that the study of complex regulons, such as CRP, *in vivo* has many limitations. However, combining several approaches, such as *in vivo* profiling, sequence analysis and ROMA increases the likelihood of obtaining a more accurate definition of a complex regulon.

The mechanism of CRP activation, in particular the role of different activating regions, has been studied using several well-characterized promoters. At Class I promoters, CRP binds to a DNA sequence upstream of the RNAP-binding site and makes direct protein–protein contact to αCTD via

AR1 of the downstream subunit of the CRP dimer, and this interaction recruits αCTD to its DNA target immediately downstream of the CRP-binding site. At Class II promoters, CRP binds to a site overlapping the −35 hexamer and makes several contacts with RNAP: AR1 of the upstream subunit of the CRP dimer binds αCTD, AR2 of the downstream subunit of the CRP dimer binds αNTD and AR3 of the downstream subunit binds region 4 of σ^{70}. The αCTD binds to its target upstream of the CRP site. At Class III promoters that contain tandem sites, CRP activation involves both Class I and Class II mechanisms [reviewed in (3)]. Nearly 90% of operons identified by ROMA showed dependence on the AR1 determinant of CRP for full activation, while AR2 is required for activation of 59% (104) of operons. We predicted that these 104 AR2-dependent operons (including 14 known operons) contain a Class II binding site in their promoter regions.

It is not surprising that transcription of some Class III operons was still activated by KE101-CRP though they all contain a Class II binding site. It has been previously shown that CRP activation at this kind of promoter involves a Class I mechanism (upstream site) and a Class II mechanism (downstream site) (30). Inactivation of AR2 only affects the proximal Class II activation, while the distal Class I site could still drive efficient activation. In some cases, where CRP activation is mainly dependent on the proximal site, the AR2 determinant will be crucial.

Interestingly, a few known FNR-regulated operons, *adhE*, *sdhC*, *sodA*, *cyoA* and *focA*, were also activated by CRP in the ROMA experiments. FNR is a CRP homologue and regulates genes during anaerobic growth. The consensus sequence for FNR is similar to that for CRP consensus, and previous studies have shown that FNR activates transcription initiation in a similar way to CRP and that both proteins can bind to the DNA site for the other protein (41,42). Therefore, we propose that some CRP-regulated operons identified in this study are actually regulated by FNR *in vivo* and contribute to anaerobic growth.

At some promoters, transcription activation by CRP can be repressed by a transcriptional regulator protein, CytR. The CytR-regulated promoters usually contain two DNA sites for CRP, centred at positions −41.5 and −93.5 with respect to the transcription start point, and a DNA site for CytR located between the two CRP-binding sites (43). Cytidine is an allosteric inducer of CytR. In the presence of cytidine, CytR is inactive, and thus CRP can activate transcription in a Class III dependent manner. In the absence of cytidine, CytR is active and binds between the two CRP dimers, repressing CRP-dependent transcription [reviewed in (42)]. CytR binds to two 5′-TTGCAA-3′ motifs between two CRP-binding sites separated by 10–30 bp (44). The *deoC*, *nupG* and *tsx* operons that are under CytR regulation were activated by CRP in our study. Many operons identified by ROMA possess tandem CRP-binding sites separated by 30–60 bp (Table 1). Sequence analysis of the promoter sequence of these operons may identify more CytR-regulated promoters.

In addition to activation, our data also shows that CRP can serve as a repressor at some promoters. There are several different mechanisms by which repressors can inhibit transcription initiation. The simplest mechanism is by blocking the interaction between RNAP and a promoter. This can occur if a binding site for a repressor protein is located overlapping the binding site for RNAP at a promoter, e.g. over the transcription start site or the −10 hexamer. The blocking mechanism is common for most repressors, such as LacI at the *lacUV5* promoter and IclR at the *iclR* promoter (45–47). CRP repression at *nirB*, *pncB*, *yeeF*, *ycfR*, *yhfC*, *ydfK*, *ynaE* and *apbA* might also involve this mechanism since at these promoter regions the CRP-binding site overlaps the sites for RNAP. Repressors also can inhibit transcription initiation by direct contact with RNAP, a process more commonly associated with activators. The P4 protein of the *B.subtilis* bacteriophage Φ29 represses the transcription of early promoter *A2C* by inhibiting promoter clearance (48). The binding of P4 overstabilizes the open complex formed by RNAP and DNA, thus impeding the following promoter escape step. The repression at *metK*, *ybiS* and *yjcB* might involve this mechanism since they all contain a strong CRP-binding site immediately upstream of a recognizable −35 hexamer. Interestingly, the CRP-binding site and the −35 region at three promoters are separated by 3 bp, which suggests that the structure of protein–protein interaction might play an important role.

CRP has been identified as a regulator of genes required for catabolism of sugars other than glucose and a large number of other genes [reviewed in (1)]. The CRP regulon defined in this study includes several operons that are involved in carbohydrate transport and metabolism, such as *fucPI* for fucose, *kdgT* for gluconate and *fucAO* for fuculose. CRP also regulates transcription of genes required for energy production, amino acid metabolism, nucleotide metabolism and ion transport systems. In addition, CRP can regulate transcription of other transcription factors, such as MelR, RpoH, BlgG, Fis and PdhR, which could further regulate transcription of genes involved in the above processes. We predict that there are several hundred genes that are directly or indirectly subjected to regulation by CRP. In our list, the functions of more than one-third of the gene products are unknown. Identification of function of the gene product of these genes will help to define the biological function of CRP in *E.coli*.

ACKNOWLEDGEMENTS

We thank Nigel J. Savery and Georgina S. Lloyd for the CRP protein used in this study. This work was supported by Biotechnology and Biological Sciences Research Council Grants EGA16107 and JIF 13209, and by a Darwin Trust of Edinburgh studentship (to D.Z.). The authors thank Steve J.W. Busby for his interest in this work and stimulating discussions.

REFERENCES

1. Kolb,A., Busby,S., Buc,H., Garges,S. and Adhya,S. (1993) Transcriptional regulation by cAMP and its receptor protein. *Annu. Rev. Biochem.*, **62**, 749–795.
2. Postma,P. (1986) Catabolite repression and related phenomena. *Symp. Soc. Gen. Microbiol.*, **39**, 21–49.
3. Busby,S. and Ebright,R. (1999) Transcription activation by catabolite activator protein (CAP). *J. Mol. Biol.*, **293**, 199–213.
4. Robison,K., McGuire,A.M. and Church,G.M. (1998) A comprehensive library of DNA-binding site matrices for 55 proteins applied to the complete *Escherichia coli* K-12 genome. *J. Mol. Biol.*, **284**, 241–254.
5. Tan,K., Moreno-Hagelsieb,G., Collado-Vides,J. and Stormo,G.D. (2001) A comparative genomics approach to prediction of new members of regulons. *Genome Res.*, **11**, 566–584.

6. Tao,H., Bausch,C., Richmond,C., Blattner,F.R. and Conway,T. (1999) Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media. *J. Bacteriol.*, **181**, 6425–6440.
7. Oh,M.K., Rohlin,L., Kao,K.C. and Liao,J.C. (2002) Global expression profiling of acetate-grown *Escherichia coli*. *J. Biol. Chem.*, **277**, 13175–13183.
8. Zheng,M., Wang,X., Templeton,L.J., Smulski,D.R., LaRossa,R.A. and Storz,G. (2001) DNA microarray-mediated transcriptional profiling of the *Escherichia coli* response to hydrogen peroxide. *J. Bacteriol.*, **183**, 4562–4570.
9. Courcelle,J., Khodursky,A., Peter,B., Brown,P.O. and Hanawalt,P.C. (2001) Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient *Escherichia coli*. *Genetics*, **158**, 41–64.
10. Arfin,S., Long,A., Ito,E., Riehle,M., Paegle,E. and Hatfield,G. (2000) Global gene expression profiling in *Escherichia coli* K12. The effects of integration host factor. *J. Biol. Chem.*, **275**, 29672–29684.
11. Hommais,F., Krin,E., Laurent-Winter,C., Soutourina,O., Malpertuy,A., Le Caer,J.P., Danchin,A. and Bertin,P. (2001) Large-scale monitoring of pleiotropic regulation of gene expression by the prokaryotic nucleoid-associated protein, H-NS. *Mol. Microbiol.*, **40**, 20–36.
12. Cao,M., Kobel,P.A., Morshedi,M.M., Wu,M.F., Paddon,C. and Helmann,J.D. (2002) Defining the *Bacillus subtilis* σ^W regulon: a comparative analysis of promoter consensus search, run-off transcription/ macroarray analysis (ROMA), and transcriptional profiling approaches. *J. Mol. Biol.*, **316**, 443–457.
13. Blattner,F.R., Plunkett,G.,III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F., Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J., Mau,B. and Shao,Y. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science*, **277**, 1453–1474.
14. Datsenko,K.A. and Wanner,B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA*, **97**, 6640–6645.
15. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
16. Tang,H., Severinov,K., Goldfarb,A. and Ebright,R.H. (1995) Rapid RNA polymerase genetics: one-day, no-column preparation of reconstituted recombinant *Escherichia coli* RNA polymerase. *Proc. Natl Acad. Sci. USA*, **92**, 4902–4906.
17. Rhodius,V.A., West,D.M., Webster,C.L., Busby,S.J. and Savery,N.J. (1997) Transcription activation at class II CRP-dependent promoters: the role of different activating regions. *Nucleic Acids Res.*, **25**, 326–332.
18. Loos,A., Glanemann,C., Willis,L.B., O'Brien,X.M., Lessard,P.A., Gerstmeir,R., Guillouet,S. and Sinskey,A.J. (2001) Development and validation of *Corynebacterium* DNA microarrays. *Appl. Environ. Microbiol.*, **67**, 2310–2318.
19. Britton,R.A., Eichenberger,P., Gonzalez-Pastor,J.E., Fawcett,P., Monson,R., Losick,R. and Grossman,A.D. (2002) Genome-wide analysis of the stationary-phase sigma factor (sigma-H) regulon of *Bacillus subtilis*. *J. Bacteriol.*, **184**, 4881–4890.
20. Rhodius,V. and Busby,S. (2000) Transcription activation by the *Escherichia coli* cyclic AMP receptor protein: determinants within activating region 3. *J. Mol. Biol.*, **299**, 295–310.
21. Valentin-Hansen,P., Holst,B., Josephsen,J., Hammer,K. and Albrechtsen,B. (1989) CRP/cAMP- and CytR-regulated promoters in *Escherichia coli* K12: the *cdd* promoter. *Mol. Microbiol.*, **3**, 1385–1390.
22. Kallipolitis,B.H. and Valentin-Hansen,P. (1998) Transcription of rpoH, encoding the *Escherichia coli* heat-shock regulator σ^{32}, is negatively controlled by the cAMP-CRP/CytR nucleoprotein complex. *Mol. Microbiol.*, **29**, 1091–1099.
23. Jayaraman,P.S., Peakman,T.C., Busby,S.J., Quincey,R.V. and Cole,J.A. (1987) Location and sequence of the promoter of the gene for the NADH-dependent nitrite reductase of *Escherichia coli* and its regulation by oxygen, the Fnr protein and nitrite. *J. Mol. Biol.*, **196**, 781–788.
24. Wubbolts,M.G., Terpstra,P., van Beilen,J.B., Kingma,J., Meesters,H.A. and Witholt,B. (1990) Variation of cofactor levels in *Escherichia coli*. Sequence analysis and expression of the pncB gene encoding nicotinic acid phosphoribosyltransferase. *J. Biol. Chem.*, **265**, 17665–17672.
25. Lobell,R. and Schleif,R. (1991) AraC-DNA looping: orientation of distance-dependent loop breaking by the cyclic AMP receptor protein. *J. Mol. Biol.*, **218**, 45–54.
26. Busby,S., Truelle,N., Spassky,A., Dreyfus,M. and Buc,H. (1984) The selection and characterization of two novel mutation in overlapping promoters of the *Escherichia coli* galactose operon. *Gene*, **28**, 201–209.
27. Ishizuka,H., Hanamura,A., Inada,T. and Aiba,H. (1994) Mechanism of the down-regulation of cAMP receptor protein by glucose in *Escherichia coli*: role of autoregulation of the *crp* gene. *EMBO J.*, **13**, 3077–3082.
28. Ebright,R. (1993) Transcription activation at Class I Cap-dependent promoters. *Mol. Microbiol.*, **8**, 797–802.
29. Ushida,C. and Aiba,H. (1990) Helical phase dependent action of CRP: effect of the distance between the CRP site and the -35 region on promoter activity. *Nucleic Acids Res.*, **18**, 6325–6330.
30. Belyaeva,T., Rhodius,V., Webster,C. and Busby,S. (1998) Transcription activation at promoters carrying tandem DNA sites for the *Escherichia coli* cyclic AMP receptor protein: organisation of the RNA polymerase α subunits. *J. Mol. Biol.*, **277**, 789–804.
31. Law,E., Savery,N. and Busby,S. (1999) Interactions between the *Escherichia coli* cAMP receptor protein and the C-terminal domain of the α subunit of RNA polymerase at class I promoters. *Biochem. J.*, **337**, 415–423.
32. Tebbutt,J., Rhodius,V.A., Webster,C.L. and Busby,S.J. (2002) Architectural requirements for optimal activation by tandem CRP molecules at a class I CRP-dependent promoter. *FEMS Microbiol. Lett.*, **210**, 55–60.
33. Belyaeva,T., Wade,J., Webster,C., Howard,V., Thomas,M., Hyde,E. and Busby,S. (2000) Transcription activation at the *Escherichia coli melAB* promoter: the role of MelR and the cyclic AMP receptor protein. *Mol. Microbiol.*, **36**, 211–222.
34. Kimata,K., Inada,T., Tagami,H. and Aiba,H. (1998) A global repressor (Mlc) is involved in glucose induction of the *ptsG* gene encoding major glucose transporter in *Escherichia coli*. *Mol. Microbiol.*, **29**, 1509–1519.
35. Plumbridge,J. (1998b) Expression of *ptsG*, the gene for the major glucose PTS transporter in *Escherichia coli*, is repressed by Mlc and induced by growth on glucose. *Mol. Microbiol.*, **29**, 1053–1063.
36. Plumbridge,J. (1998a) Control of the expression of the *manXYZ* operon in *Escherichia coli*: Mlc is a negative regulator of the mannose PTS. *Mol. Microbiol.*, **27**, 369–380.
37. Decker,K., Plumbridge,J. and Boos,W. (1998) Negative transcriptional regulation of a positive regulator: the expression of *malT*, encoding the transcriptional activator of the maltose regulon of *Escherichia coli*, is negatively controlled by Mlc. *Mol. Microbiol.*, **27**, 381–390.
38. Lee,S.J., Boos,W., Bouche,J.P. and Plumbridge,J. (2000) Signal transduction between a membrane-bound transporter, PtsG, and a soluble transcription factor, Mlc, of *Escherichia coli*. *EMBO J.*, **19**, 5353–5361.
39. Shin,D., Cho,N., Heu,S. and Ryu,S. (2003) Selective regulation of ptsG expression by Fis. Formation of either activating or repressing nucleoprotein complex in response to glucose. *J. Biol. Chem.*, **278**, 14776–14781.
40. Gosset,G., Zhang,Z., Nayyar,S., Cuevas,W.A. and Saier,M.H. (2004) Transcriptome analysis of crp-dependent catabolite control of gene expression in *Escherichia coli*. *J. Bacteriol.*, **186**, 3516–3524.
41. Li,B., Wing,H., Lee,D., Wu,H.C. and Busby,S. (1998) Transcription activation by *Escherichia coli* FNR protein: similarities to, and differences from, the CRP paradigm. *Nucleic Acids Res.*, **26**, 2075–2081.
42. Sawers,G., Kaiser,M., Sirko,A. and Freundlich,M. (1997) Transcriptional activation by FNR and CRP: reciprocity of binding-site recognition. *Mol. Microbiol.*, **23**, 835–845.
43. Valentin-Hansen,P., Søgaard-Anderson,L. and Petersen,H. (1996) A flexible partnership: the CytR ant-activator and the cAMP-CRP activator protein, comrades in transcriptional control. *Mol. Microbiol.*, **20**, 461–466.
44. Perini,L.T., Doherty,E.A., Werner,E. and Senear,D.F. (1996) Multiple specific CytR binding sites at the *Escherichia coli deo*P2 promoter mediate both cooperative and competitive interactions between CytR and cAMP receptor protein. *J. Biol. Chem.*, **271**, 33242–33255.
45. Yamamoto,K. and Ishihama,A. (2003) Two different modes of transcription repression of the *Escherichia coli* acetate operon by IclR. *Mol. Microbiol.*, **47**, 183–194.
46. Müller-Hill,B. (1998) Some repressors of bacterial transcription. *Curr. Opin. Microbiol.*, **1**, 145–151.
47. Roy,S., Garges,S. and Adhya,S. (1998) Activation and repression of transcription by different contact: two sides of a coin. *J. Biol. Chem.*, **273**, 14059–14062.

48. Monsalve,M., Calles,B., Mencia,M., Salas,M. and Rojo,F. (1997) Transcription activation or repression by phage Ψ29 protein p4 depends on the strength of the RNA polymerase-promoter interactions. *Mol. Cell*, **1**, 99–107.
49. Gulati,A. and Mahadevan,S. (2000) Mechanism of catabolite repression in the *bgl* operon of *Escherichia coli*: involvement of the anti-terminator BglG, CRP-cAMP and EIIAGlc in mediating glucose effect downstream of transcription initiation. *Genes Cells*, **5**, 239–250.
50. Davies,S.J., Golby,P., Omrani,D., Broad,S.A., Harrington,V.L., Guest,J.R., Kelly,D.J. and Andrews,S.C. (1999) Inactivation and regulation of the aerobic C(4)-dicarboxylate transport (dctA) gene of *Escherichia coli*. *J. Bacteriol.*, **181**, 5624–5635.
51. Bell,A.W., Buckel,S.D., Groarke,J.M., Hope,J.N., Kingsley,D.H. and Hermodson,M.A. (1986) The nucleotide sequences of the *rbsD*, *rbsA*, and *rbsC* genes of *Escherichia coli* K12. *J. Biol. Chem.*, **261**, 7652–7658.
52. Holcroft,C.C. and Egan,S.M. (2000) Interdependence of activation at *rhaSR* by cyclic AMP receptor protein, the RNA polymerase α subunit C-terminal domain, and rhaR. *J. Bacteriol.*, **182**, 6774–6782.
53. Cunningham,L. and Guest,J.R. (1998) Transcription and transcript processing in the *sdhCDAB-sucABCD* operon of *Escherichia coli*. *Microbiology*, **144**, 2113–2123.
54. Deeley,M.C. and Yanofsky,C. (1982) Transcription initiation at the tryptophanase promoter of *Escherichia coli* K-12. *J. Bacteriol.*, **151**, 942–951.
55. Klein,W., Horlacher,R. and Boos,W. (1995) Molecular analysis of *treB* encoding the *Escherichia coli* enzyme II specific for trehalose. *J. Bacteriol.*, **177**, 4043–4052.
56. Weickert,M.J. and Adhya,S. (1993) Control of transcription of gal repressor and isorepressor genes in *Escherichia coli*. *J. Bacteriol.*, **175**, 251–258.
57. Larson,T.J., Cantwell,J.S. and van Loo-Bhattacharya,A.T. (1992) Interaction at a distance between multiple operators controls the adjacent, divergently transcribed *glpTQ-glpACB* operons of *Escherichia coli* K-12. *J. Biol. Chem.*, **267**, 6114–6121.
58. Otsuka,J., Watanabe,H. and Mori,K.T. (1996) Evolution of transcriptional regulation system through promiscuous coupling of regulatory proteins with operons; suggestion from protein sequence similarities in *Escherichia coli*. *J. Theor. Biol.*, **178**, 183–204.
59. Webster,C., Gaston,K. and Busby,S. (1988) Transcription from the *Escherichia coli melR* promoter is dependent on the cyclic AMP receptor protein. *Gene*, **68**, 297–305.
60. Ryu,S., Ramseier,T.M., Michotey,V., Saier,M.H.,Jr and Garges,S. (1995) Effect of the FruR regulator on transcription of the *pts* operon in *Escherichia coli*. *J. Biol. Chem.*, **270**, 2489–2496.
61. Hanamura,A. and Aiba,H. (1991) Molecular mechanism of negative autoregulation of *Escherichia coli crp* gene. *Nucleic Acids Res.*, **19**, 4413–4419.
62. Golby,P., Kelly,D.J., Guest,J.R. and Andrews,S.C. (1998) Transcriptional regulation and organization of the *dcuA* and *dcuB* genes, encoding homologous anaerobic C4-dicarboxylate transporters in *Escherichia coli*. *J. Bacteriol.*, **180**, 6586–6596.
63. Plumbridge,J. and Kolb,A. (1991) CAP and Nag repressor binding to the regulatory regions of the *nagE-B* and *manX* genes of *Escherichia coli*. *J. Mol. Biol.*, **217**, 661–679.
64. Craig,J.E., Zhang,Y. and Gallagher,M.P. (1994) Cloning of the *nupC* gene of *Escherichia coli* encoding a nucleoside transport system, and identification of an adjacent insertion element, IS 186. *Mol. Microbiol.*, **11**, 1159–1168.
65. Munch-Petersen,A. and Jensen,N. (1990) Analysis of the regulatory region of the *Escherichia coli nupG* gene, encoding a nucleoside-transport protein. *Eur. J. Biochem.*, **190**, 547–551.
66. Gerlach,P., Søgaard-Andersen,L., Pedersen,H., Martinussen,J., Valentin-Hansen,P. and Bremer,E. (1991) The cyclic AMP (cAMP)-cAMP receptor protein complex functions both as an activator and as a corepressor at the *tsx*-p2 promoter of *Escherichia coli* K-12. *J. Bacteriol.*, **173**, 5419–5430.