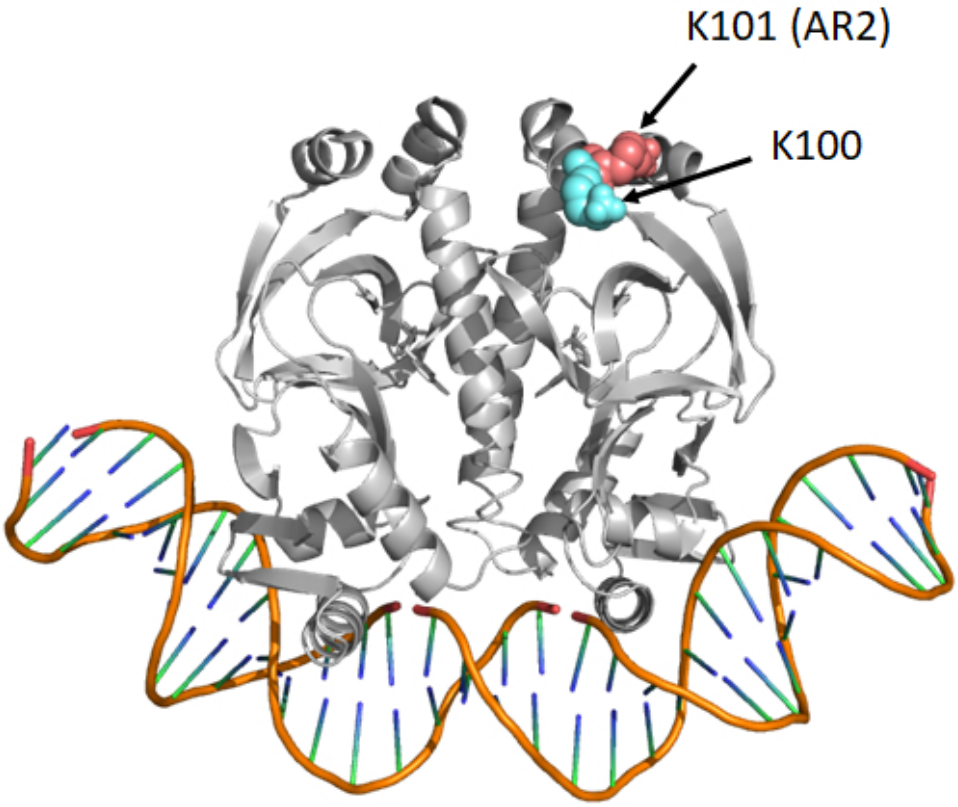
OXFORD UNIVERSITY PRESS | FEMS Microbiology Letters

http://mc.manuscriptcentral.com/fems

An ideal spacing is required for the control of Class II CRP-dependent promoters by the status of CRP K100

Journal:	*FEMS Microbiology Letters*
Manuscript ID	FEMSLE-20-06-0296.R4
Manuscript Type:	Research Letter
Date Submitted by the Author:	25-Sep-2020
Complete List of Authors:	Écija Conesa, Ana ; University of Murcia Faculty of Chemistry, Biochemistry and Molecular Biology and Immunology Gallego-Jara, Julia; University of Murcia Faculty of Chemistry, Biochemistry and Molecular Biology and Immunology Lozano Terol, Gema; University of Murcia Faculty of Chemistry, Biochemistry and Molecular Biology and Immunology Browning, Douglas; University of Birmingham, School of Biosciences Busby, Steve; University of Birmingham, School of Biosciences Wolfe, Alan; Loyola University Chicago, Microbiology and Immunology Canovas, Manuel; Universidad de Murcia, Biochemistry and Molecular Biology and Immunology (B) de Diego Puente, Teresa; University of Murcia Faculty of Chemistry, Biochemistry and Molecular Biology and Immunology
Keywords:	Escherichia coli, Cyclic AMP receptor protein (CRP), Transcription activation, Class II CRP-dependent promoters
Please select the most appropriate subject section for your submission from the drop down list. For full Section descriptions please refer to the Author guidelines at https://academic.oup.com/femsle/pages/instructions_for_authors:	Physiology & Biochemistry (Editor: Jana Jass)



43x44mm (300 x 300 DPI)

RESEARCH LETTER

An ideal spacing is required for the control of Class II CRP-dependent promoters by the status of CRP K100

Ana Écija-Conesa[1], Julia Gallego-Jara[1], Gema Lozano Terol[1], Douglas F. Browning[2], Steve J. W. Busby[2], Alan J. Wolfe[3], Manuel Cánovas Díaz[1] and Teresa de Diego Puente[1*].

[1]Department of Biochemistry and Molecular Biology and Immunology (B), Faculty of Chemistry, University of Murcia, Campus of Espinardo, Regional Campus of International Excellence ''*Campus Mare Nostrum''*, P.O. Box 4021, Murcia E-30100, Spain.

[2]School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK.

[3]Department of Microbiology and Immunology, Stritch School of Medicine, Health Sciences Division, Loyola. University Chicago, Maywood, IL 60153, USA.

*Corresponding author:

Teresa de Diego Puente

Department of Biochemistry and Molecular Biology and Immunology (B), Faculty of Chemistry, University of Murcia, Campus of Espinardo, Regional Campus of International Excellence ''Campus Mare Nostrum'', P.O. Box 4021, Murcia E-30100, Spain.

Tel.: +34 868887395; Fax: +34 968364148

E-mail: tdp@um.es

Keywords: *Escherichia coli*, cyclic AMP receptor protein (CRP), transcription activation, class II CRP-dependent promoters.

Abstract

Transcription activation by the *Escherichia coli* CRP at Class II promoters is dependent on direct interactions between RNA polymerase and CRP, therefore the spatial proximity between both proteins plays a significant role in the ability of CRP to activate transcription. Using both *in vivo* and *in vitro* techniques, here we demonstrate that the CRP K100 positive charge, adjacent to AR2, is required for full promoter activity when CRP is optimally positioned. Accordingly, K100 mediated activation is very position-dependent and our data confirm that the largest impact of the K100 status on transcription activation occurs when the spacing between the CRP binding site and the A2 of the -10 element is 22 bp. From the results of this study and the progress in the understanding about open complex DNA scrunching, we propose that CRP-dependent promoters should now be numbered by the distance from the centre of the DNA site for CRP and the most highly conserved base at position 2 of the -10 hexamer in bacterial promoters.

Introduction

CRP (cAMP receptor protein or catabolite activator protein, CAP) is a global regulator in *Escherichia coli* (*E. coli*) that controls the expression of a large number of genes in response to changes in cAMP levels, which in turn are modulated by glucose. Different regions of CRP interact with RNA polymerase (RNAP) depending on the specific class of CRP-dependent promoter . Activating region 1 (AR1) of CRP, composed of amino acid residues 156 to 164, interacts with specific surface determinant, 287 of the C-terminal of the alpha subunit (alpha-CTD) of the RNAP in both Class I and Class II promoters. The 265 determinant of alpha -CTD interacts with A-T-rich sequences in DNA, particularly with the UP element, in CRP-dependent promoters. The interaction at Class II CRP-dependent promoters also involves interaction between a second CRP region called AR2 (H19, H21, E96, and K101) and the N-terminal domain (alpha -NTD) of the RNAP alpha-subunit {Formatting Citation}. A third surface patch, known as AR3 of CRP (amino acid residues 52 to 55 and 58), has been shown to be important in CRP-dependent Class II promoter activity (Busby and Ebright 1999; Harman 2001). Recently, we demonstrated that the positive charge of lysine 100 (K100) promotes CRP activity at Class II promoters (Davis *et al.* 2018). Thus, CRP K100 lost of possitive charge by acetylation acts downregulating Class II promoter activation. In that work, we proposed that CRP K100 acetylation required for full promoter activity could be a fine-tuning of gene expression dynamics in stress response.

Furthermore, it is well known that for Class II promoters, the DNA target site for CRP is located upstream of the DNA site for RNAP, centred at position -41.5 or at various distances upstream from the transcription start point (from -39.5 to -46.5 bp), being the DNA target site for CRP overlapping with the target site for RNAP. In consequence, there is considerable variation in the location of binding sites of CRP at Class II promoters. We hypothesized that the effect of CRP K100 will depend on the length of the spacer relative to the transcription start site (TSS) and therefore on the proximity between CRP and the RNAP alpha-NTD. In this work, using semi-synthetic promoters containing different lengths of the spacer relative to the TSS, as well as natural Class II promoters in the presence of wild type (WT) or K100 mutants CRP, we evidence that the K100 impact on CRP activation is very position-dependent.

Materials and methods

Bacterial strains, plasmids and promoter constructs

The bacterial strains, plasmids, promoter fragments and the primers used in this study are listed in Table 1. Standard methods were employed for the isolation and manipulation of DNA fragments. Different derivatives of pDCRP encoding mutant *crp* genes, were made by site directed mutagenesis and donated by A. J Wolfe. The well-characterized CRP-dependent semi-synthetic CC(-41.5) promoter (Gaston et al. 1990) was manipulated to construct a set of spacing mutants that have varying

lengths between the CRP binding site and TSS: CC (-39.5), CC (-40.5), CC (-41.5) and CC (-42.5) by inserting one or more cytosine residues between the -21 and -22 positions. All constructed promoters include the synthetic consensus, 5'-TGTGATGTACATCACA-3' where CRP strongly binds (Morita *et al.* 1988; Ebright, Ebright and Gunasekera 1989). For promoter activity assays, CC(-41.5), derivatives and naturals promoters fragments were cloned into the lacZ expression vector, pRW50. To generate templates for *in vitro* transcription assays and, as a source of DNA fragments for gel retardation assays, EcoRI-HindIII promoter fragments were cloned into pSR as described previously (Browning *et al.* 2004).

Protein preparation

Purified RNA polymerase was purchased from New England Biolabs. The purification of WT and K100 mutants of CRP was carried out by affinity chromatography on cAMP-agarose as described previously (Ghosaini, Brown and Sturtevant 1988).

β-Galactosidase assays

To assay the activity of semi-synthetic promoter fragments containing spacing mutants and natural promoters (listed in Table 1), they were cloned into the low-copy-number lac expression vector, pRW50, placing the lac genes. The resulting plasmids were transformed into *E. coli* M182 Δ*crp* cells carrying pDCRP derivatives encoding wild type or K100 mutants CRP and β-galactosidase levels were determined as described previously (Beatty *et al.* 2003) using the Miller protocol (1972) (Miller 1972). In all cases, the doubly transformed cells were grown aerobically in Luria–Bertani (LB) broth at 37 ºC to measure β-galactosidase. Control experiments indicated that semi-synthetics promoters expression was completely dependent on CRP (14.15±0.88 Miller units). The empty pDU9 vector serves as a negative control. Each value is the average of three independent experiments.

Electrophoretic mobility shift assays (EMSAs)

Electromobility shift assays were performed with each of the EcoRI–HindIII fragments, end-labelled with [γ32P]-ATP, as described in (Lloyd, Busby and Savery 1998) with varying amounts of purified WT CRP or mutants derivatives in the absence or presence of RNA polymerase (New England Biolabs). EMSA reactions (10 µl) contained 0.5 nM of [^{32}P]-labeled fragments, 20mM Hepes (pH 8), 5 mM $MgCl_2$, 50 mM Potassium L-glutamate, 1 mM DTT, 25 µg/mL sonicated herring-sperm DNA, 0.5 mg/mL BSA, 200 nM AMPc and 5% glycerol. Mixtures of protein and DNA were incubated for 20 min at 37 ºC. The gel running buffer was 0.25 x Tris/borate/EDTA and 200 µM AMPc containing 2% glycerol, and samples were run on 6% acrylamide gels containing 7.5% glycerol and 0.25 x Tris/borate/EDTA using the Protean II system (Bio-Rad). DNA was visualized using dried gels to expose a phosphor storage screen and the data were analysed by densitometry. EMSA gel quantitative

analysis was performed using Quantity One software v4.4 (Bio-Rad)to measure the different intensity of each shifted band. Relative intensity quantification for the ternary complex band was calculated based on the sum of the intensities of the two bands in each vertical lines (binary complex + ternary complex) expressed in terms of percentages where the band intensity for binary complex in lane without RNAP corresponds at 100% for each assay. For each lane, the program subtracts the background that is automatically estimated using the average intensities**.**

.

***In vitro* transcription assays**

All semi-synthetic promoter fragments and *gatY* promoter were cloned into pSR plasmid, placing each promoter upstream of the λ oop transcriptional terminator. *In vitro* transcription assays were initiated by incubating each pSR derivative with purified WT or mutants CRP and RNAP. In these experiments, the 105-nucleotide RNA I transcript, encoded by pSR, served as an internal transcriptional control.

In vitro transcription reactions (20 µl) were performed in a buffer containing 100 mM KCl, 40 mM Tris-acetate (pH 7.9), 10 mM MgCl2, 1 mM dithiothreitol, 100 µg/mL of bovine serum albumin, 200 nM cAMP, 500 µM ATP, 500 µM CTP, 500 µM GTP, 5 µM UTP, and 5 µCi of [α-^{32}P]UTP (Perkin Elmer). pSR plasmids containing promoter fragments DNA tested was added at a final concentration of 10 nM. Reactions were initiated by the addition of RNA polymerase (New England Biolabs) to a final concentration 50 nM and incubated at 37 °C for 20 min. Samples were analysed by denaturing gel electrophoresis and quantified with Quantity One software v4.4 (Bio-Rad). The transcription level was expressed as the ratio of intensities bands corresponding to the transcript promoter and RNA I transcript control

Insert Table 1

Results and discussion

First, to confirm if side chain of lys-100 adjacent to CRP AR2, is critical on CRP activation, alanine substitution of K100 was performed. Then, to evaluate the impact of the K100 positive charge, substitution of lys-100 with glutamine and arginine was also carried out. The arginine residue would mimic the electrostatic charge of a lysine residue (K100R) and the glutamine residue would mimic the loss of positive charge (K100Q). In order to examine the effect of the spacer length on CRP activation we generated a set of artificial promoters with various spacing lengths: CC (-39.5), CC (-40.5), CC (-41.5) and CC (-42.5). Therefore, the lengths of the spacer between the CRP site and the adenine residue (A) at position 2 (A2) of the -10 element of RNAP binding site were changed from 20 to 23 bp (see Supplementary Figure S1). β-Galactosidase assays were used to analyze the expression of semi-synthetic promoters-lacZ fusions.

The results illustrated in Fig. 1A show that the expression profiles of semi-synthetic promoters moving the CRP site were similar between the K100 mutants and WT CRP with the highest activity at the 41.5 length. However, these results display that when K100A or K100Q CRP were positioned at -41.5, resulted in about 2-fold decrease in promoter activity. In contrast, the loss of the positive charge in 100 residue had little effect on promoter activity moving the CRP site to -42.5, -40.5 and -39.5 positions. Furthermore, the decrease in the activity of K100Q CRP compared to K100R CRP at CC (-41.5) seems to be the same as the defect with K100A CRP compared to WT CRP. These results suggested a preferred position of WT CRP for class II promoters according to the following criterion: -41.5, -42.5, -40.5 and -39.5 and further that the effect of the K100 positive charge of CRP on promoter activity was also position-dependent.

In order to investigate the effect of the substitutions of K100 on the ternary complex formation, EMSA assays were carry out employing purified WT or K100 mutants of CRP and ^{32}P end-labelled semi-synthetic promoter fragments and RNAP (see Supplementary Fig. S2). In all cases, when increasing concentrations of RNAP, a supershifted complex was observed for all fragments except for the CC (-39.5) promoter. EMSA gel quantitative analysis results are shown in supplementary Figure S2 B. When the WT or K100R CRP were incubated with the CC (-41.5) promoter fragments and 150 nM RNAP, a maximum in the formation of supershifted complex was reached (see Supplementary Fig.S2B2). However, when K100Q mutant was employed, an 85% of the ternary complex was formed and the single shifted band was still detected (see Supplementary Fig. S2B3). In the same way, the K100Q mutant was also defective on the ternary complex formation using 75 nM RNAP. In contrast, K100 substitution played little or no role in the formation of the ternary complex when both mutants were positioned at - 42.5. Our results confirm that, when CRP was suboptimally positioned, it was compromised in its ability to activate transcription, in accordance with previous results from Rossiter et al 2015 (Rossiter *et al.* 2015), that showed that when CRP is bound at position -40.5, there was an inefficient recruitment of the RNAP. In the same way, CRP-dependent promoter activity was strongly affected by a 1 bp insertion between the CRP binding site and the -10 region of the promoter, in accordance with previous results from Gaston et al1990 (Gaston *et al.* 1990). Nevertheless, apart from this, our results suggested that a positive charge at position 100 would help in the ternary complex formation when CRP was positioned at -41.5.

To analyse the contribution of K100 positive charge on transcription activation, in vitro transcription assays were carried out. The results illustrated in supplementary Figure S3 confirm that when WT or K100 mutants of CRP were positioned at -39.5 and at -40.5, transcripts were not observed. In contrast, the results shown in Fig. 1B display that when CRP is positioned at -41.5, a K100-dependent transcript was observed. The increasing WT or K100R CRP concentrations led to an increase in transcription level, reaching a maximum transcription level at 150 nM, while the K100Q substitution always reduced expression. However, when CRP was positioned at -42.5 (Fig. 1C), a lower level of transcription was reached with either WT or K100 mutants CRP, resulting in a similar transcription

profile between them. These results corroborate those shown in Fig 1 indicating that both the positive charge of K100 and the CRP position at -41.5 were required for full CRP activation.

Insert Figure 1

Taken together, the results demonstrate the full CRP activation and the positive charge of Lys-100 impact occurs with an "ideal spacing" between the CRP site and RNAP α-NTD, equivalent to 22 bp between the DNA site for CRP and the A2 base of the -10 element. The CRP-binding site for most natural Class II CRP dependent promoters is centred on -41.5 (34% of a total of 85 promoters, although for almost 50% of these there is not any experimental evidence). However not all of them have the same number of base pairs between the DNA site for CRP and the A2 base of the -10 element. Therefore, twenty fivesigma70 Class II CRP-dependent promoters were selected from Regulon DB (Boyle *et al.* 2012) to evaluate their sequence and architecture. Validated CRP binding sequence by strong evidence (binding of purified protein and site mutation experiments) was the selection criteria for these promoters (Salgado *et al.* 2013). The architecture of these promoters is shown in Supplementary Table S1. Fig. 2 shows Class II CRP-dependent promoters distribution according to the distance between the CRP binding site and the TSS (Fig. 2A) or even considering the distance between the A2 of the -10 element and the CRP site as the ruler of spacing (Fig. 2B). Surprisingly, most of the promoters are grouped in the "ideal spacing".

Insert Figure 2

To assess the effect of lysine 100 of CRP on Class II natural promoters, we generated eight natural promoter:lac fusions. The natural promoters were examined based on the distance between the centre of the DNA binding site of CRP and the TSS: -40.5 for *focA* and *gatY*, -41.5 for *malX* and *mglB*, -42.5 for *fepA* and -43.5 for *dadA* and *nupG* and also considering the distance between the CRP site and the A2 of the -10 element: 22 bp for the *focA, gatY, malX, mglB* and *dadA* promoters and 23 bp for the *fepA* and *nupG* promoters. The *fucP* promoter was also analysed but no clear dependence on CRP activation was detected when comparing promoter activity with and without WT CRP (18.12 ± 0.58 and 18.67 ± 0.49, respectively).

Insert Figure 3

The results depicted in Fig. 3 confirm that alanine substitution of K100 significantly reduced expression from the *focA, gatY, malX, mglB* and *dadA* promoters but not from the *fepA* and *nupG* promoters. These results suggested that the K100 influence was dependent on the bp spacing between the CRP binding site and the A2 of the -10 element.

To evaluate K100 positive charge on natural Class II promoters activation, β-galactosidase activity assays were also performed using K100R or K100Q mutant CRP. However, differences in expression between all mutant CRP proteins compared to the wild type protein reached statistical significance only in case *gatY* promoter in wthe K100Q substitution reduced the expression of this promoter

Insert Figure 4.

To deepen the analysis of the impact of bp spacing on transcription activation, -40.5 *gatY* promoter was selected. Moreover, *gatY* promoter has the same spacing between the DNA site for CRP and the A2 base of the-10 element as CC (-41.5) promoter with 22 bp, although it is conventionally classified as -40.5 promoter, considering the distance to TSS (Fig. 4A). Then, to evaluate the RNAP-CRP-*gatY* ternary complex formation we also used EMSA assays employing purified wild-type and K100 mutants CRP, ^{32}P end-labelled *gatY* promoter fragment and RNAP (Fig. 4B and C). When increasing the concentrations of RNAP a supershifted complex was observed. However, less ternary complex was detected using the K100Q mutant compared with WT and K100R mutant CRP. These results also corroborate that the positive K100 charge favoured the ternary complex formation with *gatY* natural promoter. Additionally, differences in the transcription level shown in Fig. 4D and E demonstrate that the status of the K100 charge affected to *gatY* promoter activity. However, Glutamine substitution of K100 resulted in a decrease in the *gatY* promoter activity, whereas the positive charge of K100 increased transcriptional activation. Besides, WT and K100R mutant CRP showed the same level of transcription. These results also confirm that "ideal" CRP binding site for *gatY* promoter was the same as that for the -41.5 semi-synthetic promoter.

By convention, the CC(-n) numbering system classification of CRP-dependent promoters is based on the distance between CRP binding site and TSS. However, recent findings show that TSS can vary due to pre-initiation scrunching of the transcription 'bubble' that is driven by the core recognition element (CRE), and also, in a few cases, by initiator NTP levels (Vvedenskaya *et al.* 2015; Fishburn, Galburt and Hahn 2016; Winkelman and Gourse 2017; Yu *et al.* 2017). The data in Figure 2 emphasise the importance of the positioning of the DNA site for CRP with respect to the -10 hexamer element, and our experiments underscore that this is the crucial consideration. Thus CRP-dependent promoters should now be numbered by the distance between the DNA site for CRP and the most highly conserved base at position 2 of the -10 hexamer, because the principal role of CRP is to hold RNAP in place so that formation of the transcription 'bubble', that initiates at the -10 hexamer element, can occur (Harley and Reynols 1987; Helmann and DeHaseth 1999; Chen *et al.* 2020).. Considering the reversible modification of K100 by acetylation in response to dynamic metabolic state of the cells, the result may help understand why 34% of natural Class II CRP dependent promoters are centred on -41.5

Insert Figure 5

Furthermore, CRP might activate transcription in two ways: interacting directly with RNAP and acting upon DNA to change its structure. The electrostatic interaction between both proteins decrease linearly in strength with the increasing distance, destabilizing the transition state and impairing the optimal transcription activation (Niu *et al.* 1996). When CRP binds, the DNA is bent 90°, which likely significantly affects protein-protein and protein-DNA interactions in promoter regions (Lee and Borukhov 2016). Fig. 5 shows the crystal structure of CRP. In this structure, K101 (AR2) point down the DNA, so even if the site rotates around the DNA (1 bp corresponds to 34° rotation angle), it points

down the helix to interact with the alpha NTD. Likewise, K100 points off to the side and it is likely that moving up and down the helix by 1 bp, leads to have a greater effect on its ability to interact with the alpha NTD. Additionally, in a previous work we show that K101, the closest AR2 neighbour to K100, was not required to promote Class II activity by K100 (Davis et al., 2018). This would explain the results reported in this work, showing why K100 mediated activation was very position-dependent. This proposition is also correlated this open complex formation model in which A2 is a first melted nucleotide interacting with the aromatic pocket of sigma-factor.

This work unveils an additional transcription regulation by the positive charge of CRP K100 whose impact depends on the optimal position of CRP. The K100 charge status, reversibly modified by acetylation, would serve as a fine-tuning mechanism in response to the metabolic state of the cells. This work contributes to increase our knowledge on the complex machinery within transcription regulatory motifs in a model organism as *E. coli*.

Funding

This work was supported by the Ministry of Science, Innovation and Universities, the State Research Agency and the European Regional Development Fund (RTI2018-094393-B-C21-MCIU/AEI/FEDER, UE), and the Seneca Foundation (20786/PI/18).

Acknowledgements

We thank Prof. J.C. García-Borrón for his generous gift of specific material for in vitro transcription experiments.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Beatty CM, Browning DF, Busby SJW *et al.* Cyclic AMP Receptor Protein-Dependent Activation of the Escherichia coli acsP2 Promoter by a Synergistic Class III Mechanism. *J Bacteriol* 2003;**185**:5148–57.

Boyle AP, Hong EL, Hariharan M *et al.* Annotation of functional variation in personal genomes using RegulomeDB. Genome Research 2012, 22(9):1790-1797. PMID: 22955989. *Genome Res* 2012;**22**:1790–7.

Browning DF, Beatty CM, Sanstad EA *et al.* Modulation of CRP-dependent transcription at the Escherichia coli acsP2 promoter by nucleoprotein complexes: Anti-activation by the nucleoid proteins FIS and IHF. *Mol Microbiol* 2004;**51**:241–54.

Busby S, Ebright RH. Transcription activation by catabolite activator protein (CAP). *J Mol Biol* 1999;**293**:199–213.

Chen J, Chiu C, Gopalkrishnan S *et al.* Stepwise Promoter Melting by Bacterial RNA Polymerase. *Mol Cell* 2020;**78**:275-288.e6.

Davis R, Écija-Conesa A, Gallego-Jara J *et al.* An acetylatable lysine controls CRP function in E. coli. *Mol Microbiol* 2018;**107**:116–31.

Ebright RH, Ebright YW, Gunasekera A. Consensus DNA site for the Escherichia coli catabolite gene activator protein (CAP): CAP exhibits a 450-fold higher affinity for the consensus DNA site than for the E.coli lac DNA site. *Nucleic Acids Res* 1989;**17**:10295–305.

Fishburn J, Galburt E, Hahn S. Transcription start site scanning and the requirement for ATP during transcription initiation by RNA polymerase II. *J Biol Chem* 2016;**291**:13040–7.

Gaston K, Bell A, Kolb A *et al.* Stringent spacing requirements for transcription activation by CRP. *Cell* 1990;**62**:733–43.

Ghosaini LR, Brown AM, Sturtevant JM. Scanning Calorimetric Study of the Thermal Unfolding of Catabolite Activator Protein from Escherichia coli in the Absence and Presence of Cyclic Mononucleotides. *Biochemistry* 1988;**27**:5257–61.

Harley C, Reynols RP. Analysis of E. coli promoter sequences. *Nucleic Acids Res* 1987;**15**:465–75.

Harman JG. Allosteric regulation of the cAMP receptor protein. *Biochim Biophys Acta - Protein Struct Mol Enzymol* 2001;**1547**:1–17.

Helmann JD, DeHaseth PL. Protein-nucleic acid interactions during open complex formation investigated by systematic alteration of the protein and DNA binding partners. *Biochemistry* 1999;**38**:5959–67.

Lee J, Borukhov S. Bacterial RNA polymerase-DNA interaction-The driving force of gene expression and the target for drug action. *Front Mol Biosci* 2016;**3**, DOI: 10.3389/fmolb.2016.00073.

Lloyd GS, Busby SJW, Savery NJ. Spacing requirements for interactions between the C-terminal domain of the α subunit of Escherichia coli RNA polymerase and the cAMP receptor protein. *Biochem J* 1998;**330**:413–20.

Miller J. *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory, NY, 1972.

Morita T, Shigesadal K, Kimizuka F *et al.* Regulatory effect of a synthetic CRP recognition sequence placed downstream of a promoter. *Nucleic Acids Res* 1988;**16**:7269–85.

Niu W, Kim Y, Tau G *et al.* Transcription activation at class II CAP-dependent promoters: Two interactions between CAP and RNA polymerase. *Cell* 1996;**87**:1123–34.

Rossiter AE, Godfrey RE, Connolly JA *et al.* Expression of different bacterial cytotoxins is controlled by two

global transcription factors, CRP and Fis, that co-operate in a shared-recruitment mechanism. *Biochem J* 2015;**466**:323–35.

Salgado H, Peralta-Gil M, Gama-Castro S *et al.* RegulonDB v8.0: Omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acids Res* 2013;**41**:203–13.

Vvedenskaya IO, Zhang Y, Goldman SR *et al.* Massively Systematic Transcript End Readout, "MASTER": Transcription Start Site Selection, Transcriptional Slippage, and Transcript Yields. *Mol Cell* 2015;**60**:953–65.

Winkelman JT, Gourse RL. Open complex DNA scrunching: A key to transcription start site selection and promoter escape. 2017;**39**:1–12.

Yu L, Winkelman JT, Pukhrambam C *et al.* The mechanism of variability in transcription start site selection. *Elife* 2017;**6**:1–22.

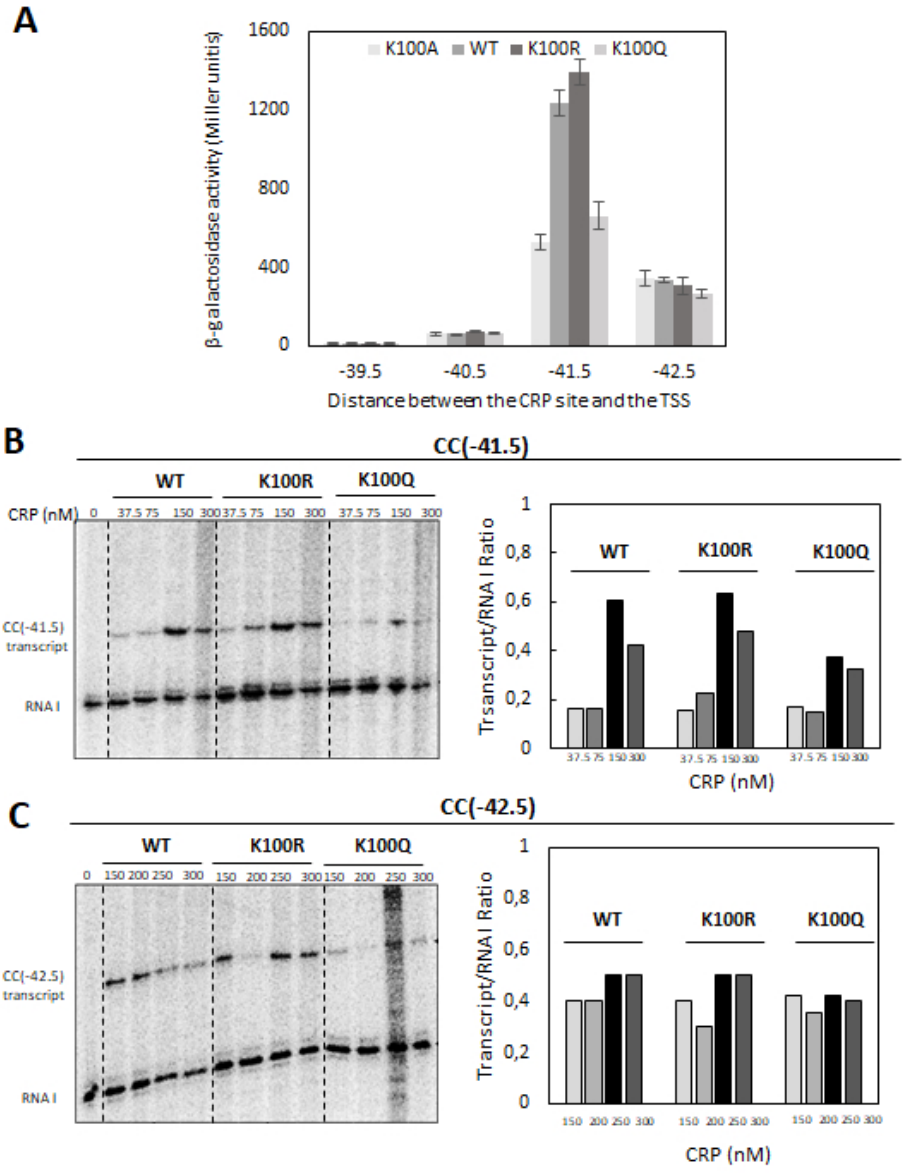


Fig. 1. A. β-galactosidase activities in M182 Δcrp cells containing pRW50 derivatives encoding different semi-synthetic promoter:lacZ fusions and WT, K100A, K100R or K100Q CRP. Data represent the means and standard deviations from three separate experiments. **B** and **C**. Left panel: Autoradiography of gels used for in vitro transcription: pSR plasmids containing semi-synthetic promoter fragments CC(-41.5) and CC(-42.5) were incubated with WT and K100 mutants CRP in the presence of 50 nM RNAP. RNA I transcript, encoded by the pSR vector, was used as an internal control. Right panel: Quantitative analysis for in vitro transcription assays expressed as the ratio of intensities bands corresponding to the transcript promoter and RNA I transcript control using promoter fragments CC(-41.5) and CC(-42.5). Each value derives from the analysis of the images shown, but three independent experiments were carried out with standard deviation of inter-assay less than 10%.

45x59mm (300 x 300 DPI)

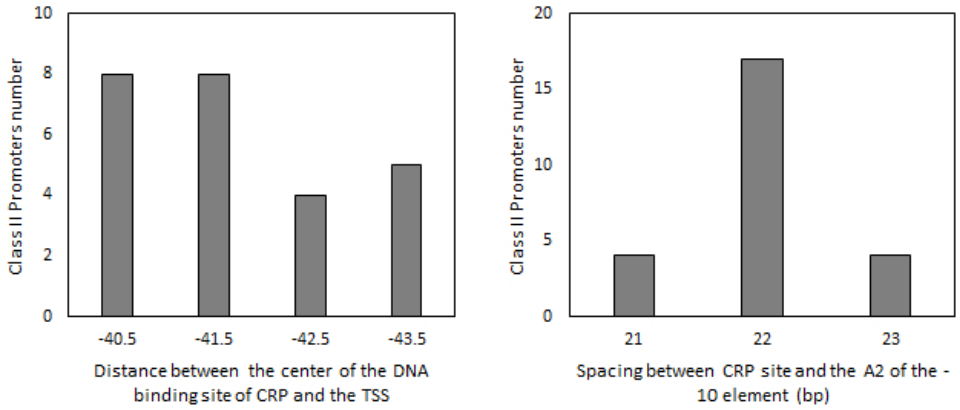


Fig. 2. Class II CRP-dependent promoters distribution according the CC(-n) numbering system classification (A) or bp spacing (B).

55x24mm (300 x 300 DPI)

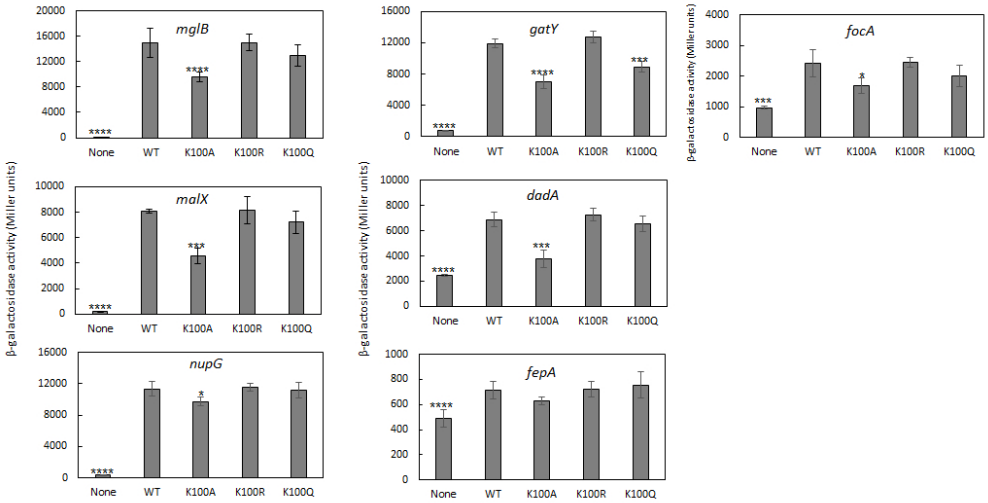


Fig. 3. Expression of promoter: lac fusions in cells expressing wild-type or mutant CRP. The bar charts illustrate measured β-galactosidase activities in M182 Δcrp cells containing pRW50 derivatives encoding different promoter: lacZ fusions, as shown in each panel, and pDCRP encoding wild-type or mutant CRP proteins as indicated. For each promoter, the none control corresponds to cells carrying an empty pDU9 vector. Statistical analysis by one-way ANOVA was carried out in order to evaluate differential significance between the WT and K100 mutants CRP (p-value < 0.0001 (****), < 0.001 (***), < 0.05(*).

87x46mm (300 x 300 DPI)

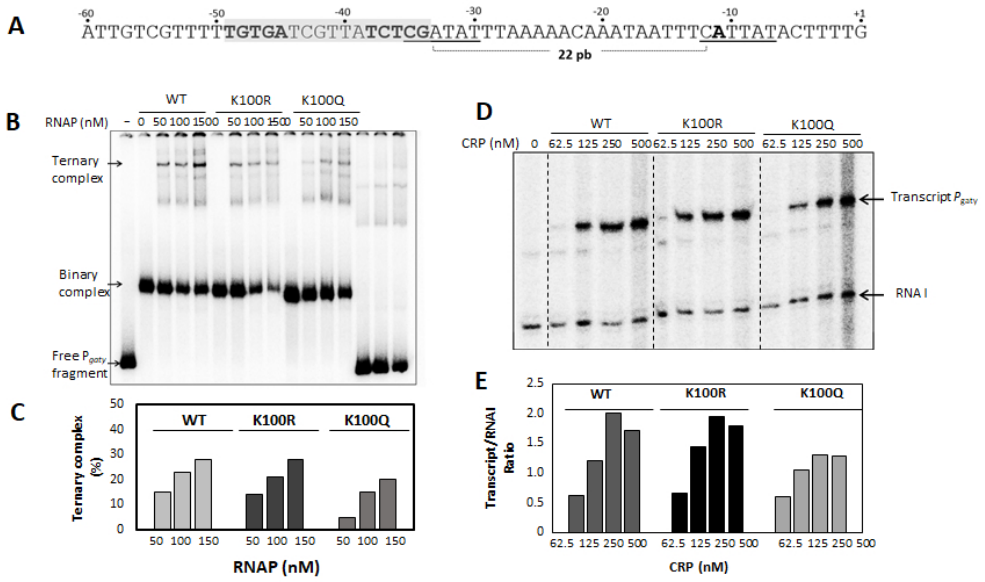


Fig 4. A. gatY promoter sequence. The locations of the CRP binding sites are shaded in grey. The bp spacing between the A2 base of the -10 element (in bold) and the CRP site is indicated (22 bp). The - 35 and -10 sequences are underlined and the TSS is indicated as +1. **B**. Ternary complex formation by electrophoretic mobility shift assays: end-labelled gatY fragments were incubated with increasing concentrations of RNAP in the presence of 100 nM purified WT CRP and K100R and K100Q mutants (lanes 14-16 show the RNAP controls that confirm the non-interaction between RNAP and gatY fragment in the absence of CRP). **C**. EMSA gel quantitative analysis for RNAP concentration dependence of ternary complex formation for WT CRP (■) and K100R (■) and K100Q (■) mutants. Each value derives from the analysis of the gel image shown but three independent experiments were carried out with standard deviation of inter-assay less than 10%. **D**. Autoradiography of a gel used for in vitro transcription. pSR plasmids containing gatY promoter fragments were incubated with purified WT or mutants CRP in the presence of 50 nM RNAP. RNA I transcript, encoded by the pSR vector, was used as an internal control. **E**. Quantitative analysis for in vitro transcription assays expressed as the ratio of intensities bands corresponding to the transcript promoter and RNA I transcript control. Each value derives from the analysis of the image shown but three independent experiments were carried out with standard deviation of inter-assay less than 10%.

72x43mm (300 x 300 DPI)

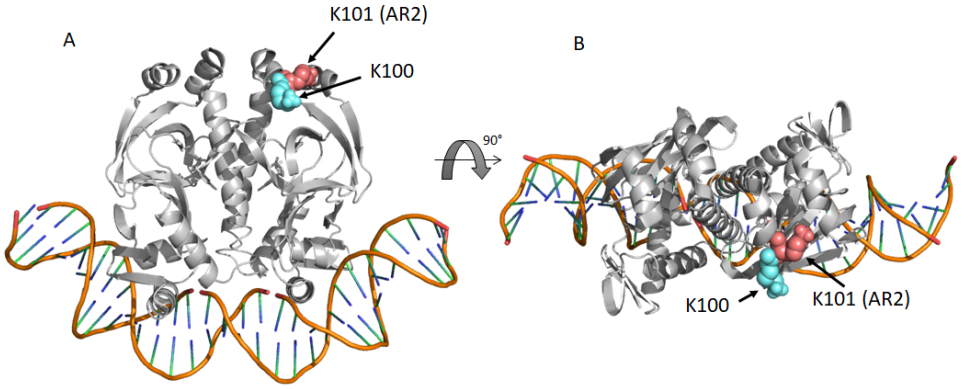


Fig. 5. The crystal structure of CRP (pdb code 1CGP). Front image (A) and 90° rotated image (B).

92x44mm (300 x 300 DPI)

Figure legends.

Fig. 1. **A**. β-galactosidase activities in M182 Δ*crp* cells containing pRW50 derivatives encoding different semi-synthetic promoter:lacZ fusions and WT, K100A, K100R or K100Q CRP. Data represent the means and standard deviations from three separate experiments. **B** and **C. Left panel:** Autoradiography of gels used for *in vitro* transcription: pSR plasmids containing semi-synthetic promoter fragments CC(-41.5) and CC(-42.5) were incubated with WT and K100 mutants CRP in the presence of 50 nM RNAP. RNA I transcript, encoded by the pSR vector, was used as an internal control. **Right panel**: Quantitative analysis for *in vitro* transcription assays expressed as the ratio of intensities bands corresponding to the transcript promoter and RNA I transcript control using promoter fragments CC(-41.5) and CC(-42.5). Each value derives from the analysis of the images shown, but three independent experiments were carried out with standard deviation of inter-assay less than 10%.

Fig. 2. Class II CRP-dependent promoters distribution according the CC(-n) numbering system classification (A) or bp spacing (B).

Fig. 3. Expression of promoter: lac fusions in cells expressing wild-type or mutant CRP. The bar charts illustrate measured β-galactosidase activities in M182 Δ*crp* cells containing pRW50 derivatives encoding different promoter: lacZ fusions, as shown in each panel, and pDCRP encoding wild-type or mutant CRP proteins as indicated. For each promoter, the none control corresponds to cells carrying an empty pDU9 vector. Statistical analysis by one-way ANOVA was carried out in order to evaluate differential significance between the WT and K100 mutants CRP (p-value < 0.0001 (****), < 0.001 (***), < 0.05(*).

Fig 4. A. *gatY* promoter sequence. The locations of the CRP binding sites are shaded in grey. The bp spacing between the A2 base of the -10 element (in bold) and the CRP site is indicated (22 bp). The - 35 and -10 sequences are underlined and the TSS is indicated as +1. **B**. Ternary complex formation by electrophoretic mobility shift assays: end-labelled *gatY* fragments were incubated with increasing concentrations of RNAP in the presence of 100 nM purified WT CRP and K100R and K100Q mutants (lanes 14-16 show

the RNAP controls that confirm the non-interaction between RNAP and *gatY* fragment in the absence of CRP). **C.** EMSA gel quantitative analysis for RNAP concentration dependence of ternary complex formation for WT CRP (■) and K100R (■) and K100Q (■) mutants. Each value derives from the analysis of the gel image shown but three independent experiments were carried out with standard deviation of inter-assay less than 10%. **D**. Autoradiography of a gel used for *in vitro* transcription. pSR plasmids containing *gatY* promoter fragments were incubated with purified WT or mutants CRP in the presence of 50 nM RNAP. RNA I transcript, encoded by the pSR vector, was used as an internal control. **E**. Quantitative analysis for in vitro transcription assays expressed as the ratio of intensities bands corresponding to the transcript promoter and RNA I transcript control. Each value derives from the analysis of the image shown but three independent experiments were carried out with standard deviation of inter-assay less than 10%.

Fig. 5. The crystal structure of CRP (pdb code 1CGP). Front image (A) and 90° rotated image (B).

Table 1. Strains, plasmids, promoter fragments and primers.

Strain, promoter, plasmid, or primer	Relevant characteristic	Source
Bacterial strains		
E. coli M182 Δ*crp*	Δ*crp* derivative of M182	(Busby, Kotlarz and Buc 1983)
Plasmids		
pDCRP	pBR322 derivative encoding wild type *crp*	(West *et al.* 1993)
pDU9	Derivative of pDCRP with *crp* gene deleted	(Bell *et al.* 1990)
pDCRPK100R	pBR322 derivative encoding *crp* gene carrying the KR 100 substitution	Donated by A.J. Wolfe
pDCRPK100Q	pBR322 derivative encoding *crp* gene CRP carrying the KQ 100 substitution	Donated by A. J. Wolfe
pDCRPK100A	pBR322 derivative encoding *crp* gene CRP carrying the KA 100 substitution	Donated by A. J. Wolfe
pRW50	Broad host range low copy lac expression vector for cloning EcoRI–HindIII promoter fragments	(Lodge *et al.* 1992)
pRW50/CC(-39.5)	pRW50, CC(-39.5) semi-synthetic promoter	Donated by S. Busby
pRW50/CC(-40.5)	pRW50, CC(-40.5) semi-synthetic promoter	Donated by S. Busby
pRW50/CC(-41.5)	pRW50, CC(-41.5) semi-synthetic promoter	(West *et al.* 1993)
pRW50/CC(-42.5)	pRW50, CC(-42.5) semi-synthetic promoter	Donated by S. Busby
pRW50/*gatY*	pRW50, *gatY* promoter	(Hollands, Busby and Lloyd 2007)
pRW50/focA	pRW50, *focA* promoter	This study
pRW50/malX	pRW50, *malX* promoter	(Hollands, Busby and Lloyd 2007)
pRW50/mglB	pRW50, *mglB* promoter	(Hollands, Busby and Lloyd 2007)
pRW50/fepA	pRW50, *fepA* promoter	This study
pRW50/dadA	pRW50, *dadA* promoter	This study
pRW50/nupG	pRW50, *nupG* promoter	This study
pRW50/fucP	pRW50, *fucP* promoter	This study
pSR	pBR322 derivative, vector for in vitro transcription	(Kolb *et al.* 1995)
pSR/*gatY*	pSR, *gatY* promoter	This study
pSR/CC(-39.5)	pSR, CC(-39.5) semi-synthetic promoter	This study
pSR/CC(-40.5)	pSR, CC(-40.5)semi-synthetic promoter	This study
pSR/CC(-41.5)	pSR, CC(-41.5) semi-synthetic promoter	(Savery *et al.* 1998)
pSR/CC(-42.5)	pSR, CC(-42.5) semi-synthetic promoter	This study
Promoter fragments		
CC(-41.5)	*E. coli melR* promoter derivative with consensus DNA site for CRP centred at -41.5	(Gaston *et al.* 1990)
CC(-40.5)	*CC(-41.5)* with 1 bp deleted between the DNA site for CRP and -10 element	(West *et al.* 1993)
CC(-39.5)	*CC(-41.5)* with 2 bp deleted between the DNA site for CRP and -10 element	(West *et al.* 1993)
CC(-42.5)	*CC(-41.5)* with 1 bp inserted between the DNA site for CRP and -10 element	(West *et al.* 1993)

gatY	EcoRI–HindIII fragment carrying *gatY* promoter	(Hollands, Busby and Lloyd 2007)
fepA	EcoRI–HindIII fragment carrying *fepA* promoter	This study
malX	EcoRI–HindIII fragment carrying *malX* promoter	(Hollands, Busby and Lloyd 2007)
nupG	EcoRI–HindIII fragment carrying *nupG* promoter	This study
focA	EcoRI–HindIII fragment carrying *focA* promoter	This study
mglB	EcoRI–HindIII fragment carrying *mglB* promoter	(Hollands, Busby and Lloyd 2007)
fucP	EcoRI–HindIII fragment carrying *fucP* promoter	This study
dadA	EcoRI–HindIII fragment carrying *dadA* promoter	This study
Primers		
Seq. pSR (F)	CAAAATGCCGCAAAAAAGGGAA	This study
pRW50-EcoRI (F)	AAGTTTCCAACGCGAGCCATT	This study
pRW50-HindIII (R)	CAGGTCGTTGAACTGAGCCT	This study

F, refers to the forward primer and R to the reverse primer.

Supplementary material

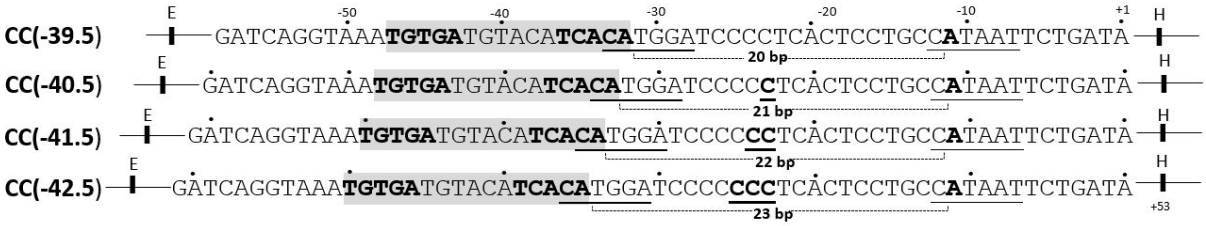


Fig. S1. Semi-synthetic promoter fragments containing different lengths of the spacer relative to the TSS designated +1. The bp spacing between the A2 of the -10 element (in bold) and the CRP site is indicated (20, 21, 22 and 23 bp). The locations of the CRP binding sites are shaded in grey. The inserted cytosine residues and the - 35 and -10 sequences are underlined. The relevant restriction sites are also shown, EcoRI, E and HindIII, H.

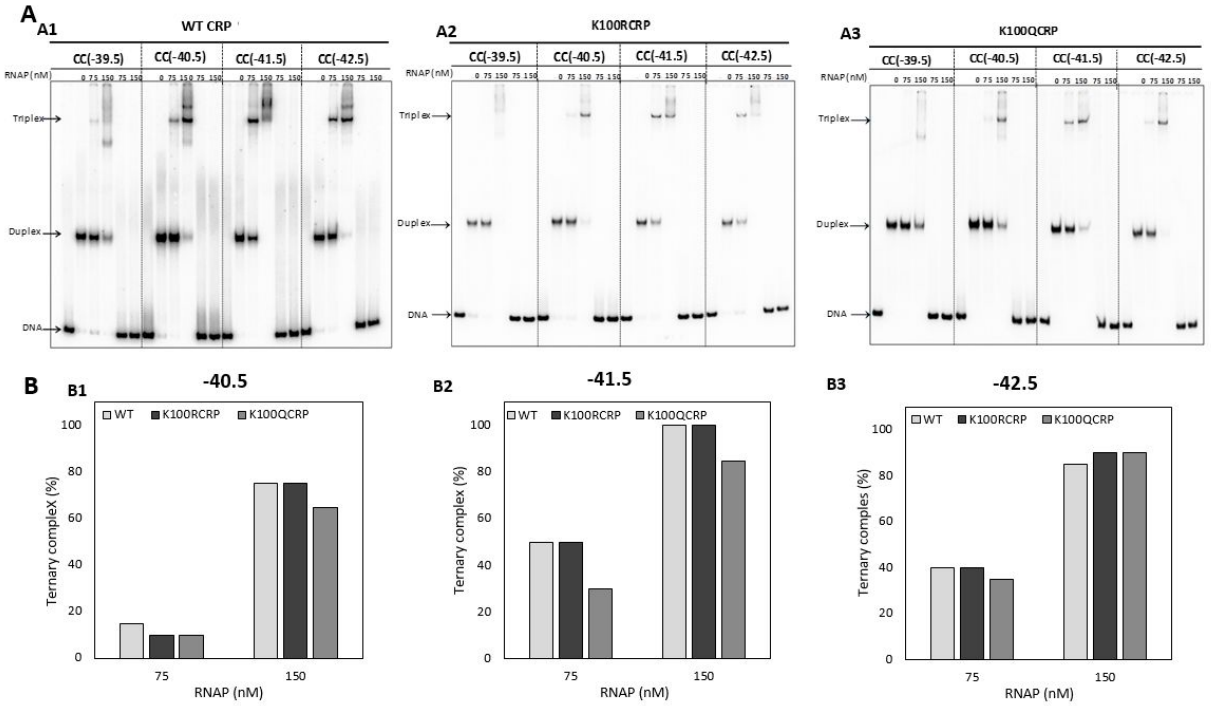


Fig. S2. A. Ternary complex formation by electrophoretic mobility shift assays. End-labelled CC(-39.5), CC(-40.5), CC(-41.5) and CC(-42.5) fragments were incubated with increasing concentrations of RNAP in the presence of 100 nM WT (**A1**), K100R (**A2**) and K100Q (**A3**) mutants of CRP. **B.** EMSA gel quantitative analysis from WT, K100R and K100Q mutants based on the CRP binding site: -40.5 bp (**B1**), -41.5 bp (**B2**) and -42.5 bp (**B3**). Each value derives from the analysis of the gel images shown above, but three independent experiments were carried out with standard deviation of inter-assay less than 10%.

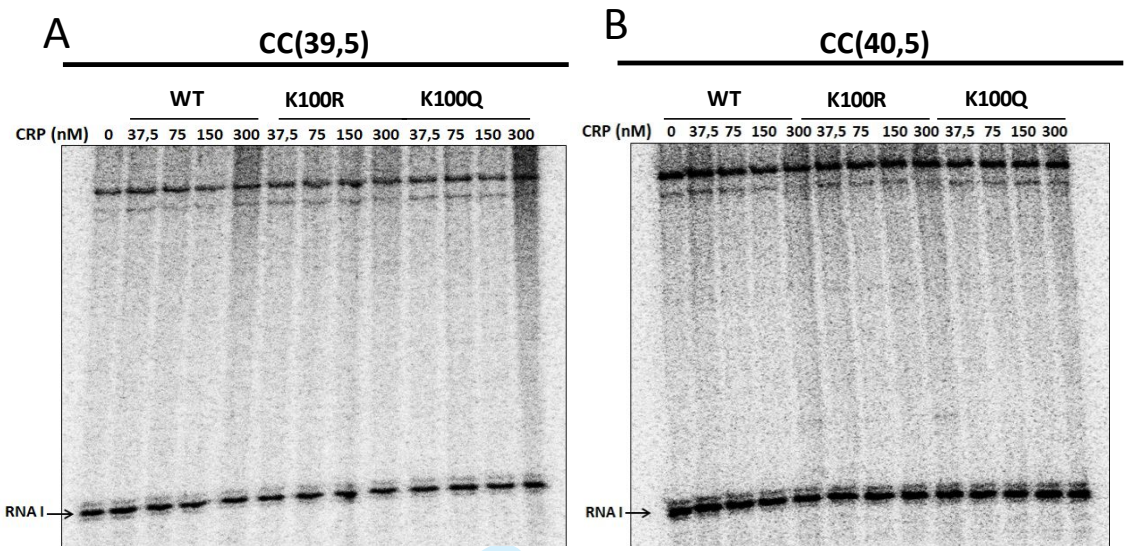


Fig. S3. Autoradiography of gels used for in vitro transcription. pSR plasmids containing semi-synthetic promoter fragments, CC(-39.5), CC(-40.5) were incubated with WT and K100 mutants CRP in the presence of 50 nM RNAP. RNAI transcript, encoded by the pSR vector, was used as an internal control.

Table S1. Class II CRP-dependent promoters with validated CRP binding sequence (Salgado et al. 2013).

Promoter	Distance [CRP-+1][1]	bp spacing[2]	Promoter sequence
deoCp2	-40.5	21	GATTTCCTTAAT**TGTGA**TGTGTA**TCGAA**GTGTGTTGCGGAGTAGATGTTTAGAATACTAACAAACTCGCAAGGTGAATTTTA
focAp1	-40.5	22	AGCCAGGCGAGA**TATGA**TCTATA**TCAAT**TTCTCATCTATAATGCTTTGTTAGTATCTCGTCGCCGACTTAATAAAGAGAGA
fucPp	-40.5	21	CTAGCTAATAAG**TGTGA**CCGCCG**TCATA**TTACAGAGCGTTTTTTATTTGAAAATGAATCCATGAGTTCATTTCAGACAGGC
gatYp	-40.5	22	ATTGTCGTTTT**TGTGA**TCGTTA**TCTCG**ATATTTAAAAACAAATAATTTCATTATATTTTGAAATCGAAAACAAACGACAG
glpABCp	-40.5	22	AATGTTCAAAAT**GACGC**ATGAAA**TCACG**TTTCACTTTCGAATTATGAGCGAATATGCGCGAAATCAAACAATTCATGTTTT
mhpRp1	-40.5	22	ACTCGGACAAAA**TGTCG**TTGCGC**GCACA**GTACAGCGCAACTTATTTTGTTAAAAACATGTAAATGATTTTTTATTGTGCGC
nupGp	-40.5	21	TTGCAATTATT**TGCCA**CAGGTA**ACAAA**AAACCAGTCCGCGAAGTTGATAGAATCCCATCATCTCGCACGGTCAAATGTGC
tsxp2	-40.5	21	AATGATAGAAC**TGTGA**AACGAA**ACATA**TTTTTGTGAGCAATGATTTTTATAATAGGCTCCTCTGTATACGAAATATTTAG
galEp1	-41.5	22	GATTCCACTAA**TTTAT**TCCATG**TCACA**CTTTTCGCATCTTTGTTATGCTATGGTTATTTCATACCATAAGCCTAATGGAGC
glpTQp	-41.5	22	ATTTAATAATG**TGTGC**GGCAAT**TCACA**TTTAATTTATGAATGTTTTCTTAACATCGCGGCAACTCAAGAAACGGCAGGTTC
malXp	-41.5	22	TCGTTGCGTAA**TGTGA**TTTATG**CCTCA**CTAAAATTTGATAAAACGTTTTATCTTCTCGCGCAATTTACTGAATCCAGATTG
melRp	-41.5	22	AGGGTGAAAAC**CGTGC**TCCCAC**TCGCA**GTCATCCTCCCTCACTCCTGCCATAATTCTGATATTCCAGGAAAGAGAGCCATC
mglBp	-41.5	22	CGCTTTCAATC**TGTGA**GTGATT**TCACA**GTATCTTAACAATGTGATAGCTATGATTGCACCGTTTTAACGTTGTAACCCGTA
rpoHp5	-41.5	22	GCATTGAACT**TGTGG**ATAAAA**TCACG**GTCTGATAAAACAGTGAATGATAACCTCGTTGCTCTTAAGCTCTGGCACAGTTG
udpP	-41.5	22	ATTTGCGTCAT**GGTGA**TGAGTA**TCACG**AAAAAATGTTAAACCCTTCGGTAAAGTGTCTTTTTGCTTCTTCTGACTAAACCG
ychHp	-41.5	22	AGGGTTGTAAT**TGTGA**TCACGC**CCGCA**CATAACCCACTGGGTGTTGTCTATACTTTACACATAAGGAAGAGGGGTATTCCC
araFp	-42.5	22	AATTCTGCGA**TGTGA**TATTGC**TCTCC**TATGGAGAATTAATTTCTCGCTAAAACTATGTCAACACAGTCACTTATCTTTTAG
cddp	-42.5	22	GCATAATTAA**TGAGA**TTCAGA**TCACA**TATAAAGCCACAACGGGTTCGTAAACTGTTATCCCATTACATGATTATGAGGCAA
cyaRp	-42.5	22	TGGAAAATTC**TTAGA**AACCGA**TCACA**TACAGCTGCATTTATTAAGGTTATCATCCGTTTCGCTGAAAAACATAACCCATAA
ppiAp2	-42.5	23	CATTTTAAGA**GGTGA**TTTTGA**TCACG**GAATAAAAAGTGATCGTCAGGTTACATATATTTCAGATACGTAAAATTAGGTAAA
yhfAp	-43.5	23	GCACGGTAA**TGTGA**CGTCCT**TTGCA**TACATGCAGTACATCAATGTATTACTGTAGCATCCTGACTGTTTTAGCATAGCTTT
dadAp1	-43.5	22	TCAGGGAGA**TGTGA**GCCAGC**TCACC**ATAAAAAAGCCGCATGTTGAATAATATTTTCAACTGAGTTATCAAGATGTGATTAG
malIp	-43.5	22	AAATTTTAG**TGAGG**CATAAA**TCACA**TTACGCAACGATAATAGCGGGTATAAGATAAATAAAGGTAAAACGTTTTATCTGT
tdcAp	-43.5	23	AAGTTAATT**TGTGA**GTGGTC**GCACA**TATCCTGTTCATTTCATTTTGATACACTTCATGCCGTCAATGAGGTAATTAACGTA
cpdBp	-43.5	23	TGCGCCAAC**TGTGA**TAGTGT**CATCA**TTTTCAAAGCGTAAAATTGTGGCATTCTTCACTGTTCTATAAGTAAGACGTTTATT

[1] Distance from the TSS (+1) to the central position of CRP site.

[2] bp spacing between the CRP site (shaded in grey and in bold) and the A2 of the -10 hexamer (underlined), highlighted in light green shading.