6942–6951 *Nucleic Acids Research, 2005, Vol. 33, No. 21*
doi:10.1093/nar/gki998

Transcription regulation of the EcoRV restriction–modification system

Ekaterina Semenova[1], Leonid Minakhin[1], Ekaterina Bogdanova[3], Maxim Nagornykh[3], Anatoliy Vasilov[1], Tomasz Heyduk[4], Alexander Solonin[3], Marina Zakharova[3] and Konstantin Severinov[1,2,5,*]

[1]Waksman Institute for Microbiology, [2]Department of Biochemistry and Molecular Biology, Rutgers, the State University of New Jersey, Piscataway, NJ 08854, USA, [3]Institute of Biochemistry and Physiology of Microorganisms, Russian Academy of Sciences, Puschino 142292, Russia, [4]E. A. Doisy Department of Biochemistry and Molecular Biology, St Louis University Medical School, St Louis, MO 63104, USA and [5]Institute of Molecular Genetics, Moscow, Russia

Received October 17, 2005; Revised and Accepted November 16, 2005

ABSTRACT

When a plasmid containing restriction–modification (R–M) genes enters a naïve host, unmodified host DNA can be destroyed by restriction endonuclease. Therefore, expression of R–M genes must be regulated to ensure that enough methyltransferase is produced and that host DNA is methylated before the endonuclease synthesis begins. In several R–M systems, specialized Control (C) proteins coordinate expression of the R and the M genes. C proteins bind to DNA sequences called C-boxes and activate expression of their cognate R genes and inhibit the M gene expression, however the mechanisms remain undefined. Here, we studied the regulation of gene expression in the C protein-dependent EcoRV system. We map the divergent EcoRV M and R gene promoters and we define the site of C protein-binding that is sufficient for activation of the EcoRV R transcription.

INTRODUCTION

Type II restriction–modification (R–M) systems consist of (i) a restriction endonuclease that recognizes a specific DNA sequence and introduces double-stranded breaks at or around the recognition site and (ii) a methyltransferase (methylase) that recognizes the same DNA sequence and methylates it. Methylation prevents site recognition by the endonuclease and thus protects the target DNA from cleavage. Type II R–M genes are often plasmid-encoded and can spread from one bacterial host to another, crossing species boundaries and impacting genome evolution on a global scale (1,2). While some view R–M systems as purely selfish, i.e. concerned with their own propagation through bacterial populations (2,3), a plasmid containing R–M genes can confer selective advantage by, e.g. protecting the host from bacteriophage infection, which the phage will have to overcome by acquiring specialized anti-restriction genes [such as, e.g. T7 gene 0.3, (4)].

During cell entry and establishment of a plasmid containing R–M genes, unmodified host DNA can be attacked by the endonuclease, causing host cell death. It is therefore intuitively clear that expression of R–M genes should be regulated to ensure that enough methylase is produced to methylate host DNA before endonuclease is synthesized. Since many R–M genes are found clustered on broad-range mobile genetic elements, coordinated expression of these genes should occur in different bacteria, i.e. should be independent of host regulators.

Many R–M systems, such as BamHI (5), BglII (6), Eco72I (7), EcoRV (8), Esp1396I (9), PvuII (10) and SmaI (11) rely on specialized Control (C) proteins (12) for coordinated expression. Genes coding for C proteins are usually located upstream of, and partially overlap with the endonuclease gene (13). C proteins bind to palindromic DNA sequences called C-boxes (13) and activate expression of their cognate endonuclease genes as well as their own expression. As a consequence, C protein overproduction interferes with establishment of plasmids carrying the corresponding R–M genes (3) presumably by causing premature activation of endonuclease expression.

*To whom correspondence should be addressed at Waksman Institute for Microbiology, 190 Frelinghuysen Road, Piscataway, NJ 08854, USA. Tel: +1 732 445 6095; Fax: +1 732 445 5735; Email: severik@waksman.rutgers.edu
Correspondence may also be addressed to Marina Zakharova. Tel: +7 096 773 2598; Fax: +7 095 956 3370; Email: zemskovam@mail.ru

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

© The Author 2005. Published by Oxford University Press. All rights reserved.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oxfordjournals.org

Overproduction of some C proteins was also shown to inhibit methylase gene expression *in vivo* (3,10).

All C proteins are related through common ancestry and several R–M systems can be regulated by heterologous C proteins, indicating that the regulation mechanism is evolutionarily conserved. C proteins are also related to phage helix–turn–helix (HTH) DNA-binding transcription factors, including the well-studied λ repressor. The structure of one C protein, from the *Ahd*I R–M system, was recently solved (14). As expected from the results of sequence analysis, the structure of the C.AhdI dimer is similar to the λ repressor structure. However, in contrast to the well-studied λ switch, nothing is known about the mechanism(s) of C proteins action, their mode of interaction with RNAP, or with the C-boxes. In this work, we characterized the effects of C.EcoRV, a C protein encoded by the EcoRV R–M system, on transcription from the EcoRV promoters.

MATERIALS AND METHODS

Bacterial strains and media

Escherichia coli HB101 (ATCC33694) was used as a host to study *eco*RV genes expression; *E.coli* XL1-Blue (Stratagene, USA) was used as a cloning host; *E. coli* M15[pREP4] (Qiagen, USA) was used to express recombinant C.EcoRV protein, *E. coli* Z85 (23) was used for phage restriction experiments. Cells were grown at 37°C in standard Luria–Bertani medium (broth) (LB) media with appropriate antibiotics.

Plasmids and proteins

Plasmid pEF42 was constructed by cloning a PvuII fragment containing the entire set of *eco*RV genes from a natural plasmid pLG13 (15) in a cloning vector pHSG415. Plasmid pEF42ΔC and pEF42ΔCR were constructed by site-specific PCR mutagenesis of pEF42. Plasmid pEcoRVC was constructed by cloning the entire C.EcoRV reading frame with the upstream 70 bp EcoRV sequence into the pFD51 vector. pEcoRVCS and pEcoRVCL were constructed from pEcoRVC by site-directed mutagenesis. Plasmids pM and pR were generated by inserting, in opposite orientations, the 170 bp *eco*RV spacer with truncated EcoRV•C gene into the pFD51 vector plasmid. Details of molecular cloning are available from the authors upon request. Mutations of the C-boxes in pR were made by QuikChange Site-Directed Mutagenesis Kit (Stratagene) according to manufacturer's protocol.

Purification of recombinant C.EcoRV was described previously (8).

Primer extension and 5′ RACE

For primer extension experiments, *E.coli* HB101 cells harboring corresponding plasmid(s) were harvested in the exponential phase of growth at $OD_{600} = 0.4$, and total RNA was extracted using RNeasy Mini Kit (QIAGEN) according to manufacturer's instructions including the DNase I digestion step. RNA concentrations and purity were tested by measurements of absorbance at 260/280 nm and by electrophoresis in 1% formaldehyde-agarose gel. For primer extension reaction, 20 μg of total RNA were reverse-transcribed with 100 U of SuperScript III enzyme from First-Strand Synthesis kit for RT–PCR (Invitrogen) according to the manufacturer's protocol in the presence of 1 pmol [γ-^{32}P] end-labeled primer. The reactions were treated with RNase H, precipitated with ethanol and dissolved in formamide loading buffer. As a marker for each primer extension reaction, sequencing reaction (with *fmol* DNA Cycle Sequencing kit from Promega) was performed on PCR fragment of the corresponding plasmid using the same end-labeled primer as was used for the primer extension. The reaction products were resolved on a 7% sequencing gel and revealed using PhosphorImager.

5′ RACE was performed exactly as described in Semenova *et al*. (16) and Minakhin *et al*. (17). In this method, one-half of the RNA sample is treated with tobacco acid pyrophosphatase (TAP), an enzyme that converts 5′ triphosphates of RNA into monophosphates that can be ligated to exogenously added RNA oligonucleotide by RNA ligase. Another one-half of the RNA sample is left untreated and is used as a control. After the ligation, RT and PCR amplification are used to identify DNA fragments whose abundance increases after TAP treatment (such fragments must be generated from RNA molecules that contained 5′ triphosphates, and must therefore correspond to transcription initiation start points).

Gel retardation assay

The reactions contained, in 10 μl of reaction buffer [40 mM Tris–HCl (pH 8.0), 90 mM KCl, 10 mM $MgCl_2$, 125 μg/ml BSA, 5 mM DTT and 10% glycerol], 0–1.6 μM of C.EcoRV and 2 nM of [γ-^{32}P] labeled DNA fragment containing wild-type or mutant C boxes. Reactions were incubated for 10 min at 37°C, combined with 2 μl of loading buffer (50% glycerol, 0.05% bromophenol blue and 0.1 μg/ml heparin) and immediately loaded on to an 8% native polyacrylamide gel. After electrophoresis at 400 V for 2 h at room temperature, the gels were visualized using PhosphorImager.

RESULTS

Genetic organization of the EcoRV control region

The genetic organization of the EcoRV control region is shown on Figure 1. The *eco*RV•R and *eco*RV•M genes are divergently transcribed; their initiating codons are separated by a 306 bp spacer. The spacer contains the *eco*RV•C gene, which is transcribed in the same direction as *eco*RV•R and overlaps with it by 17 bp. The *eco*RV•C and *eco*RV•R are likely co-transcribed from an upstream promoter, though the transcription start site has not been mapped. Two translational products of the *eco*RV•C gene are possible. The ATG for the longer 99 aminoacid product is located 23 bp upstream of the *eco*RV•M initiating ATG; the ATG for the shorter 75 amino acid product is 72 bp further upstream. Both proteins were overexpressed and both were reported to bind to *eco*RV C-box *in vitro* (8).

The product of the shorter *eco*RV•C open reading frame (ORF) is sufficient for stable maintenance and expression of EcoRV R–M genes *in vivo*

Plasmid pEF42 contains the entire EcoRV R–M system. *E.coli* cells carrying pEF42 restrict the growth of phage λ (Figure 2), indicating that plasmid-borne *eco*RV genes are functional and

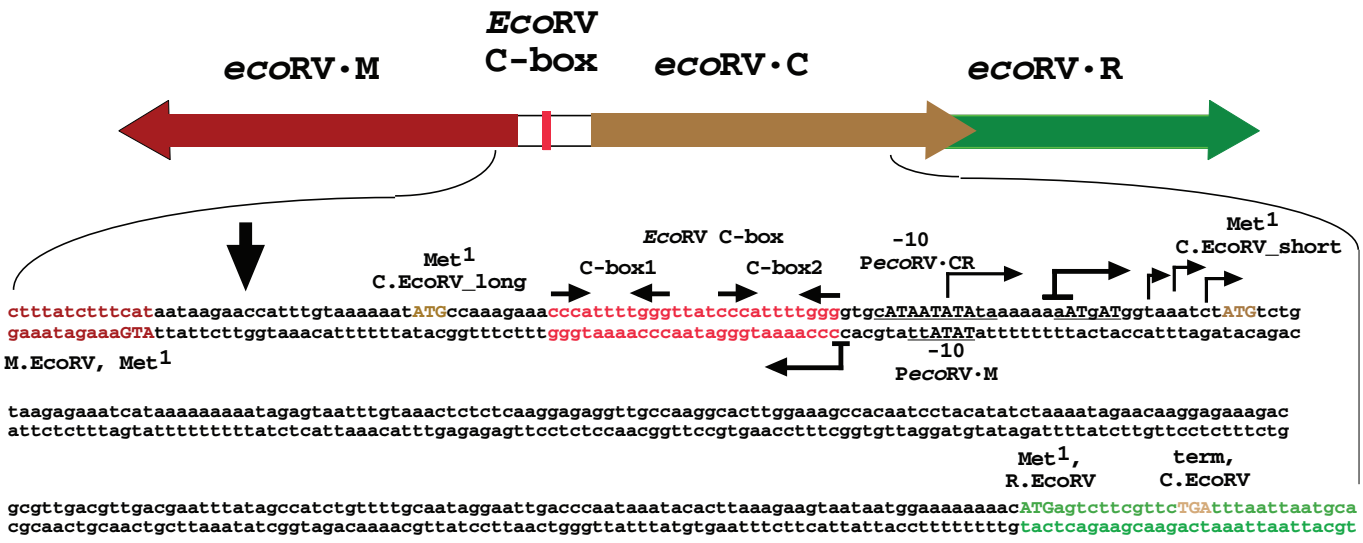

Figure 1. The EcoRV regulatory region. The *eco*RV genes are schematically shown at the top (the shorter *eco*RV•C ORF is shown in solid brown color. The sequence of the intergenic region containing the entire sequence of *eco*RV•C and the initial sequences of the *eco*RV R–M genes is expanded at the bottom. The beginnings of *eco*RV R–M ORFs are indicated by colors matching those used at the top of the figure. The P*eco*RV•M and P*eco*RV•CR start sites are shown by leftward and rightward arrows, respectively, the potential −10 promotor elements are capitilized and underlined. The EcoRV C-box is shown in red; the two sets of inverted repeats in the C-box are indicated by arrows and labeled C-box1 and C-box2. The two potential ATG codons for *eco*RV•C product and its termination codon are capitalized and color-coded.

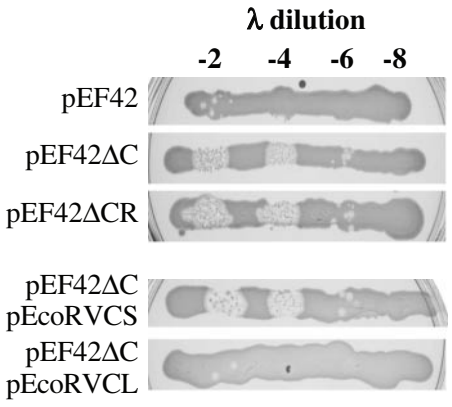

Figure 2. The shorter *eco*RV•C ORF is sufficient for EcoRV R–M system function. The horizontal lines show the overnight 37°C growth of *E.coli* Z85 strain harboring the indicated plasmids on LB plates. Cells were spotted with indicated dilutions of λ-*vir* phage lysate.

that the activities of the *eco*RV promoters in cells harboring pEF42 represent the situation during stable maintenance of the *eco*RV•R–M system, when both the methyltransferase and the restriction endonuclease genes are expressed. pEF42ΔC is a derivative of pEF42 in which the reading frame of the shorter *eco*RV•C ORF is disrupted by a single-bp deletion. Cells carrying pEF42ΔC were as ineffective in restricting λ growth as cells harboring pEF42ΔCR, a plasmid that lacks both the *eco*RV•C and *eco*RV•R genes (Figure 2). The results therefore show that no R.EcoRV synthesis occurs in the absence of C.EcoRV.

Plasmid pEcoRVC contains the longer *eco*RV•C ORF and 70 bp of the upstream EcoRV sequence. pEcoRVC is compatible with pEF42 and its derivatives. To determine which of the two possible C.EcoRV proteins is sufficient for restriction of λ growth, *E.coli* cells harboring pEF42ΔC were transformed with two pEcoRVC derivatives. The first derivative, pEcoRVCS, harbored a frame-shift mutation identical to that engineered in pEF42ΔC (affects both the shorter and the longer *eco*RV•C ORF); the second derivative, pEcoRVCL, harbored a 1 bp deletion that only affected the longer ORF. As can be seen from Figure 2, cells harboring the second derivative efficiently restricted λ growth, while cells harboring the first one did not, as expected. Therefore, we conclude that the shorter C.EcoRV is sufficient for production of enough R.EcoRV to restrict λ growth. Based on the efficiency of restriction by cells harboring pEcoRVCS and pEF42ΔC (compared to that observed with cells harboring pEF42 only), the longer C.EcoRV, if it is produced, does not significantly contribute to R.EcoRV synthesis.

Determination of the EcoRV transcription start sites

The key to understanding transcription regulation in the EcoRV system is to relate the site of C.EcoRV-binding, the C-box, with transcription initiation start points and rates of utilization of divergent promoters P*eco*RV•CR and P*eco*RV•M in the presence and in the absence of C.EcoRV. Total RNA was isolated from exponentially growing *E.coli* HB101 cells harboring pEF42 or pEF42ΔC and primer extension reactions were performed to identify divergent EcoRV transcripts initiated close to the C-box. As can be seen from Figure 3, left panel, a single primer extension product reflecting leftward (P*eco*RV•M) transcription from pEF42 or pEF42ΔC was detected. The primer extension product band obtained with RNA prepared from cells harboring pEF42 had a diffuse appearance on the denaturing gel, suggesting the presence of several closely spaced 5′ ends. The primer extension band obtained with RNA from cells harboring pEF42ΔC was less diffuse. The amounts of primer extension product did not

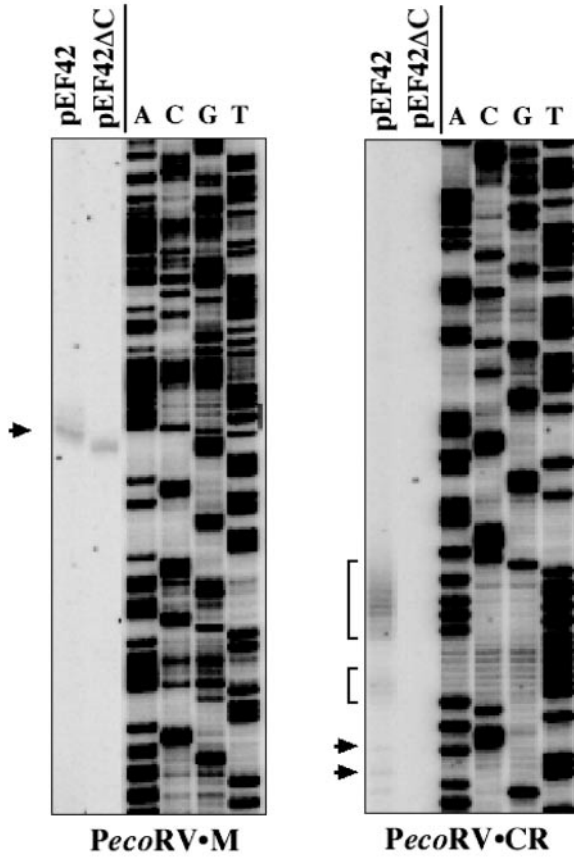

Figure 3. Primer extension analysis of *eco*RV promoters. RNA was purified from the *E.coli* HB101 cells harboring the indicated plasmids and primer extension reactions were performed to reveal 5′ end of divergent RNAs transcribing from the EcoRV regulatory region. The sequencing reactions' marker lanes were prepared with the EcoRV DNA and primers used for primer extension.

appear to be significantly different in cells harboring pEF42 or pEF42ΔC, suggesting that C.EcoRV does not have a strong effect on steady-state levels of P*eco*RV•M transcripts.

The 5′ end of primer extension products were localized to the run of 4 Gs at the rightmost edge of the C-box (for the orientation presented in Figure 1). The 5′ end of primer extension products obtained with RNA from cells harboring pEF42ΔC were 2 nt downstream of 5′ end of primer extension products obtained with RNA from cells harboring pEF42. The primer extension end points are preceded by two overlapping sequences, TATtAT and TATAtT, which are similar to the σ^{70}-10 promoter element consensus sequence TATAAT, 17–19 bases further upstream, there is a TaGAtt sequence similar to the −35 promoter element consensus sequence TTGACA.

As an alternative way to determine P*eco*RV•M transcription start site(s), 5′ RACE was performed. A single PCR fragment was observed, and its abundance increased upon the treatment of the RNA preparation with TAP, indicating that the PCR fragment was due to amplification of transcripts initiated by RNAP rather than due to processed transcripts' amplification (see Materials and Methods for details). The PCR fragment obtained with RNA prepared from cells harboring pEF42 was cloned and the EcoRV insert sequence was determined in 27 recombinant plasmids. The results are summarized in Table 1 and overall confirm the results of primer extension mapping. Interestingly, the most frequent clone contained 5 Gs at the end of the EcoRV sequence (9 plasmids), while five clones contained 4 Gs. Since EcoRV DNA only contains 4 Gs in the area of the P*eco*RV•M start site, at least one C residue present at the 5′ end of the majority of P*eco*RV•M transcripts is not template-encoded. This must be due to transcript slippage at the run of G residues in the template strand of the initial transcribed sequence of P*eco*RV•M. Plasmids containing EcoRV inserts with 2 and 1 Gs at the end of the EcoRV sequence were also found (1 each). Since transcript slippage requires at least three identical residues in the DNA template (18) the most likely transcription initiation start point for most P*eco*RV•M is at the two rightmost Gs of the C-box. In addition to clones corresponding to transcription initiation events at the stretch of Gs, rare clones containing EcoRV inserts that began downstream and upstream of this region were also detected (Table 1).

5′ RACE was also performed with RNA purified from cells harboring pEF42ΔC and 23 plasmids containing EcoRV inserts were sequenced (Table 1). In this case, no slippage products at the stretch of Gs was observed and most clones corresponded to transcripts containing 1,2 or 3 Gs at their 5′ end (5 plasmids each). No plasmids corresponding to upstream transcription initiation/processing events were present, while plasmids with downstream boundaries were found at the same frequency as in the samples prepared from cells harboring pEF42 (Table 1).

The locations of rightward (P*eco*RV•CR) transcripts start points were also determined. As can be seen, primer extension products were only observed when functional C.EcoRV was present (Figure 3, right panel). The major group of primer extension products mapped to an AT-rich area upstream of an uninterrupted run of 8 As in the EcoRV DNA. Additional minor primer extension products mapped in the beginning of the run of As and also downstream of it. The most downstream product's 5′ end was located only 2 nt upstream of the initiating ATG of the shorter *eco*RV•C ORF. It should be noted that no primer extension products corresponding to RNA 5′ end that originated from within the shorter *eco*RV•C ORF were detected (data not shown), indicating that the *eco*RV•C and *eco*RV•R ORFs are indeed co-transcribed.

5′ RACE was performed to identify the P*eco*RV•CR transcripts initiation points. One PCR fragment was observed, however its abundance did not change upon TAP treatment. The PCR fragments obtained with and without TAP treatment were treated as described above, and the EcoRV sequence was determined in 36 individual recombinant plasmids (18 each from samples with and without TAP treatment). Analysis of plasmid sequences showed no differences in distribution of different clones in samples with or without TAP treatment (data not shown). Therefore the combined (with and without TAP) results are presented in Table 2. As can be seen, in addition to transcripts containing 1, 2, 3, 7 and 8 As and corresponding to transcription initiation within the run of 8 As in the EcoRV DNA, transcripts containing 9, 10, 11, 12 and 16 As were present (Table 2). Thus, as is the case with the P*eco*RV•M transcripts, the majority of P*eco*RV•CR transcripts are the products of transcript slippage. It should be noted that though the results of 5′ RACE do not identify

Table 1. Determination of the P*eco*RV•M transcript start points by 5′ RACE

No. of clones	
	pEF42
1	TAATACGTGGGGTTTTACCCTATTGGGTTTTACCCAAAGAAACCGTATAAAAAATGTTTACCAAGAATAATACTTTCTATTTCAAAAACAAGGCG
1	AATACGTGGGGTTTTACCCTATTGGGTTTTACCCAAAGAAACCGTATAAAAAATGTTTACCAAGAATAATACTTTCTATTTCAAAAACAAGGCG
9	**G**GGGGTTTTACCCTATTGGGTTTTACCCAAAGAAACCGTATAAAAAATGTTTACCAAGAATAATACTTTCTATTTCAAAAACAAGGCG
5	GGGGTTTTACCCTATTGGGTTTTACCCAAAGAAACCGTATAAAAAATGTTTACCAAGAATAATACTTTCTATTTCAAAAACAAGGCG
1	GGTTTTACCCTATTGGGTTTTACCCAAAGAAACCGTATAAAAAATGTTTACCAAGAATAATACTTTCTATTTCAAAAACAAGGCG
1	GTTTTACCCTATTGGGTTTTACCCAAAGAAACCGTATAAAAAATGTTTACCAAGAATAATACTTTCTATTTCAAAAACAAGGC
1	TTTTACCCTATTGGGTTTTACCCAAAGAAACCGTATAAAAAATGTTTACCAAGAATAATACTTTCTATTTCAAAAACAAGGC
1	TACCCTATTGGGTTTTACCCAAAGAAACCGTATAAAAAATGTTTACCAAGAATAATACTTTCTATTTCAAAAACAAGGC
2	AAAAAATGTTTACCAAGAATAATACTTTCTATTTCAAAAACAAGGC
1	AAAAATGTTTACCAAGAATAATACTTTCTATTTCAAAAACAAGGC
1	AAAATGTTTACCAAGAATAATACTTTCTATTTCAAAAACAAGGC
1	AAAAACAAGGC
2	AACAAGGC
	pEF42ΔC
2	GGGGTTTTACCCTATTGGGTTTTACCCAAAGAAACCGTATAAAAAATGTTTACCAAGAATAATACTTTCTATTTCAAAAACAAGGCG
5	GGGTTTTACCCTATTGGGTTTTACCCAAAGAAACCGTATAAAAAATGTTTACCAAGAATAATACTTTCTATTTCAAAAACAAGGCG
5	GGTTTTACCCTATTGGGTTTTACCCAAAGAAACCGTATAAAAAATGTTTACCAAGAATAATACTTTCTATTTCAAAAACAAGGCG
5	GTTTTACCCTATTGGGTTTTACCCAAAGAAACCGTATAAAAAATGTTTACCAAGAATAATACTTTCTATTTCAAAAACAAGGCG
1	**TT**TTTTACCCTATTGGGTTTTACCCAAAGAAACCGTATAAAAAATGTTTACCAAGAATAATACTTTCTATTTCAAAAACAAGGCG
2	AAAAAATGTTTACCAAGAATAATACTTTCTATTTCAAAAACAAGGCG
1	AAAAATGTTTACCAAGAATAATACTTTCTATTTCAAAAACAAGGC
1	AAAATGTTTACCAAGAATAATACTTTCTATTTCAAAAACAAGGC
1	AAAAACAAGGC

The EcoRV sequences of plasmids obtained after cloning the 5′ RACE product obtained after TAP treatment of RNA prepared from cells harboring pEF42 or pEF42ΔC are shown. Residues not encoded by the EcoRV DNA are indicated in boldface.

Table 2. Determination of the P*eco*RV•CR transcript start points by 5′ RACE

No. of clones	
1	**AAAAAAAA**AAAAAAAATGATGGTAAA
3	**AAAA**AAAAAAAATGATGGTAAA
4	**AAA**AAAAAAAATGATGGTAAA
3	**AA**AAAAAAAATGATGGTAAA
7	**A**AAAAAAAATGATGGTAAA
6	AAAAAAAATGATGGTAAA
1	AAAAAAATGATGGTAAA
2	AAATGATGGTAAA
1	AATGATGGTAAA
3	ATGATGGTAAA

The EcoRV sequences of plasmids obtained after cloning the 5′ RACE product obtained with and without TAP treatment of RNA prepared from cells harboring pEF42 are shown. Residues not encoded by the EcoRV DNA are indicated in boldface.

P*eco*RV•CR initiation point (due to the lack of response to TAP treatment), the presence of long A runs at P*eco*RV•CR 5′ end clearly indicates that P*eco*RV•CR transcription initiates within the run of 8 As. The products of slippage may account for some of the longer primer extension products seen in Figure 3. Because of the large scatter in the positions of 5′ end of the P*eco*RV•CR transcripts initiated in the run of As, it is difficult to unambiguously determine the positions of promoter elements.

To understand the EcoRV system establishment and maintenance in a naïve host, the activities of P*eco*RV•CR and P*eco*RV•M need to be compared. To this end, two plasmids, pR and pM, in which the EcoRV spacer with truncated *eco*RV•C is cloned in opposite orientations in a vector plasmid, were constructed. Since the same oligonucleotide primers can be used to detect EcoRV transcripts from pM and pR, and since the same amounts of plasmid DNA are present in pM and pR-containing cells (as judged by visual inspection of ethidium bromide-stained agarose gels), pairwise comparisons of steady-state levels of transcripts from P*eco*RV•CR and P*eco*RV•M in the presence or in the absence of C.EcoRV provided by compatible pEF42 becomes possible.

Total RNA was isolated from exponentially growing *E.coli* HB101 cells harboring either pR or pM alone, or pR (or pM) and pEF42 (or pEF42ΔC), and primer extension reactions were performed (Figure 4). Primer extension products' 5′ end obtained with pR and pM-specific primers matched those obtained using the pEF42 primers (Figure 3). Importantly, comparisons of the amounts of primer extension products in the absence or in the presence of pEF42 showed that in the presence of C.EcoRV, the steady-state levels of P*eco*RV•CR and P*eco*RV•M transcripts are comparable.

The *eco*RV•R promoter-proximal C-box half-site does not contribute to *eco*RV•C regulation

The C.EcoRV protein is homologous to λ repressor, a prototypical molecular switch that cooperatively interacts with several closely spaced DNA-binding sites (operators). Different patterns of transcription from divergent λ promoters that are regulated by the repressor binding and whose activities determine the lysis-lysogeny decision are achieved depending on the occupancy of repressor binding sites (19). Inspection of the C-box sequence reveals that it contains two pairs of inverted repeats (Figure 5A). Assuming that only one pair of inverted repeats is recognized by the C.EcoRV protein

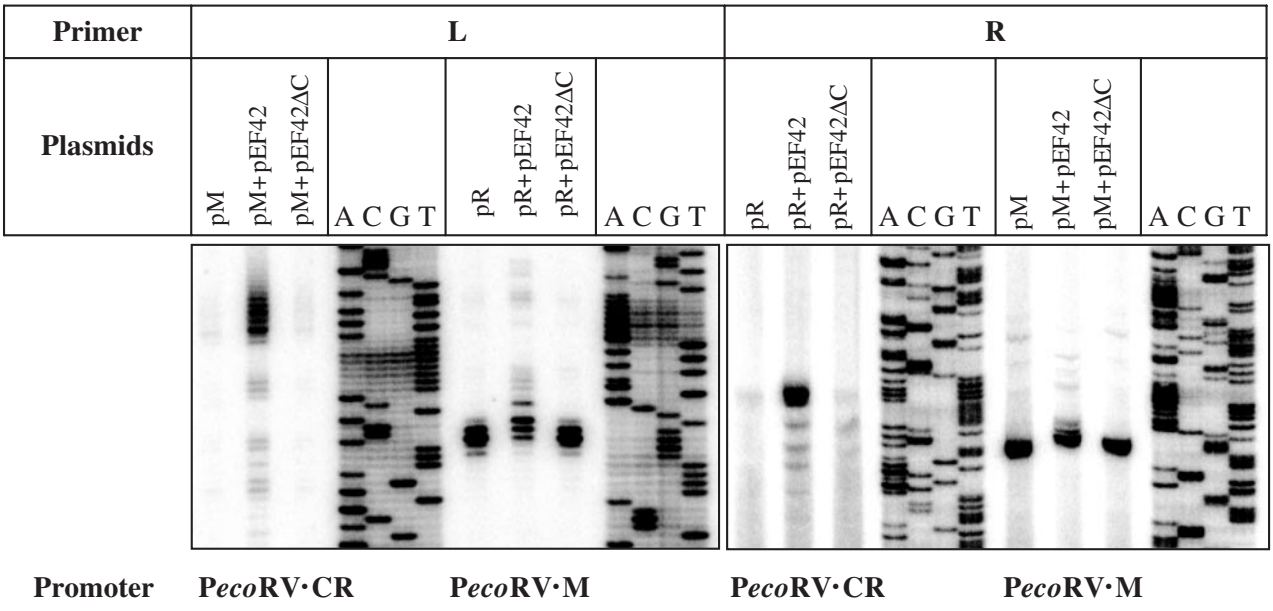

Figure 4. Steady-state levels of P*eco*RV•CR and P*eco*RV•M transcripts in the presence or in the absence of C.EcoRV. RNA was purified from the *E.coli* HB101 cells harboring the indicated plasmids and primer extension reactions were performed to reveal 5′ end of divergent RNAs arising from the EcoRV regulatory region. The R primer allows the detection of rightward transcripts from pR or pM (P*eco*RV•CR or P*eco*RV•M transcripts, correspondingly). The L primer allows the detection of leftward transcripts from pR or pM (P*eco*RV•M or P*eco*RV•CR transcripts, correspondingly). The sequencing reactions' marker lanes were prepared using the pM or pR plasmids and primers used for primer extension.

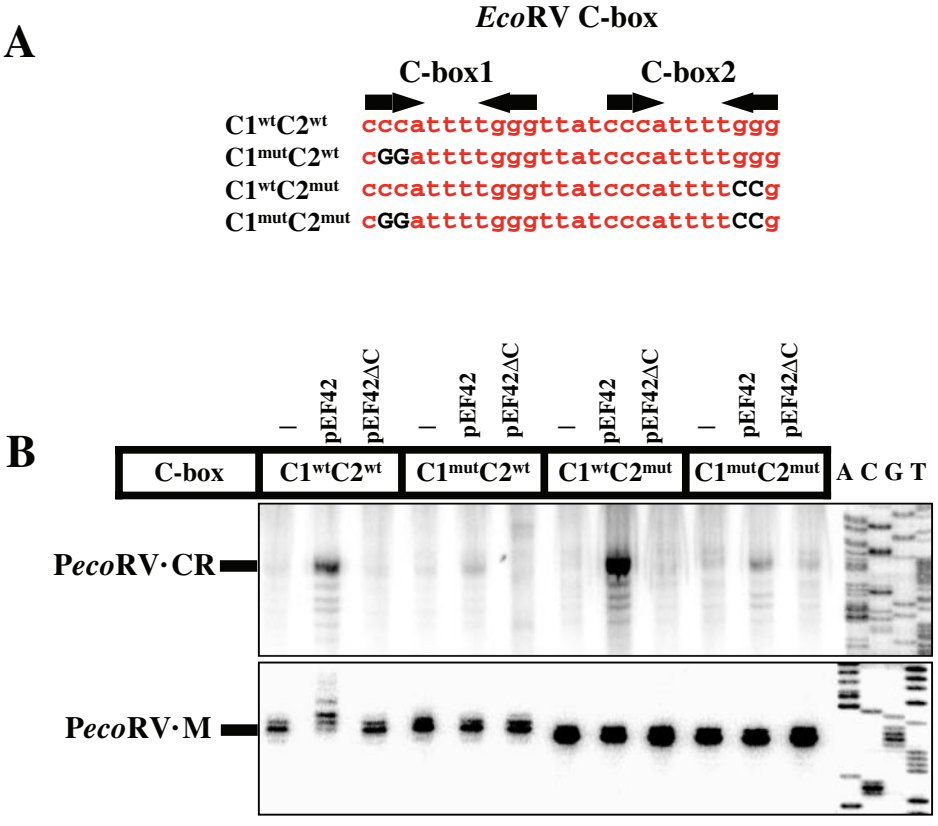

Figure 5. Effect of the EcoRV C-box mutations on steady-state levels of *eco*RV promoters' transcripts. (**A**) Sequences of wild-type and mutant C-boxes used in experiments are shown. Two sets of inverted repeats above the alignment are indicated by arrows and labeled C-box1 and C-box2. (**B**) RNA was purified from the *E.coli* HB101 cells harboring the pR plasmid with or without mutations in the EcoRV C-box and compatible plasmids pEF42 (the source of C.EcoRV) or pEF42ΔC (no functional C.EcoRV produced) and primer extension reactions were performed to determine the amounts of P*eco*RV•CR and P*eco*RV•M transcripts. The sequencing reactions' marker lanes were prepared using the pR plasmid and primers used for primer extension.

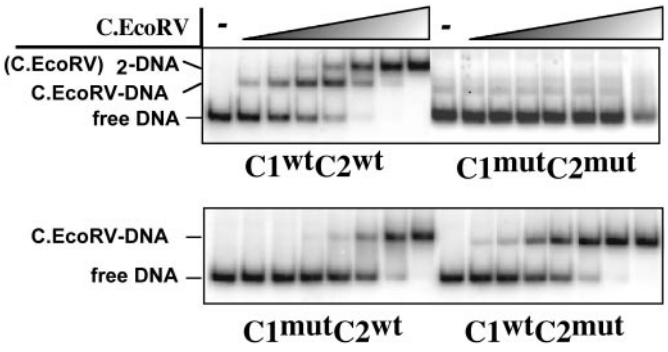

Figure 6. Interaction of C.EcoRV protein with mutant EcoRV C-box DNA. DNA fragments containing mutations in EcoRV C-box were combined with increasing amounts of C.EcoRV (the product of the shorter *eco*RV•C ORF) and reaction products were resolved by native PAGE and revealed by autoradiography.

dimer, a question arises about the role of each of the two C-box operators in regulation of EcoRV transcription. To answer this question, pR derivatives with substitutions in the first, second or both C-box operators were created and tested *in vivo* in the presence or in the absence of C.EcoRV provided by pEF42. In each C-box operator, two outermost G:C base pairs were substituted by the C:G base pairs (Figure 5A). The mutations break the dyad-axis symmetry of C-box operators and should prevent C.EcoRV-binding (this expectation is validated in experiment shown in Figure 6, see below). As can be seen from Figure 5B, substitution in the second, P*eco*RV•CR-proximal operator (C-box2) had no deleterious effect on C.EcoRV-dependent activation of the *eco*RV•CR promoter transcription. In fact, the fold activation of transcription from P*eco*RV•CR (compared to basal level) was stronger when C-box2 was mutated. In contrast, mutation of the first, P*eco*RV•CR-distal operator (C-box1) abolished the activation of P*eco*RV•CR and was indistinguishable from the double mutation changing both operators. We therefore conclude that the binding of the C.EcoRV protein to P*eco*RV•CR-distal C-box1 is required (and sufficient) for activation of *eco*RV•R expression.

Both the C-box2 mutation and the double mutation stimulated transcription from P*eco*RV•M and also shifted the transcription initiation start point 2 nt downstream. Since the C-box2 mutation alters the initial transcribed sequence of P*eco*RV•M, the reasons for observed effects may be complex. One attractive possibility is that changing the C-box2 sequence stimulates transcription by interfering with transcript slippage. Mutation in C-box1 had little effect on the overall level of transcription from P*eco*RV•M but inhibited C.EcoRV-dependent transcript slippage. It is therefore tempting to speculate that slippage results from a combination of appropriate initial transcribed sequence (a run of 4 Gs) and from a clash between RNAP that escapes the promoter and C.EcoRV bound at C-box1.

C.EcoRV binds more tightly to the P*eco*RV•CR-distal C-box1

EcoRV fragments containing wild-type or mutant C-boxes were used in gel mobility shift assays with purified recombinant C.EcoRV (the product of the shorter ORF). The results are presented in Figure 6. As can be seen, the addition of increasing concentrations of C.EcoRV shifted the wild-type EcoRV DNA fragment. Two distinct shifted bands were apparent. One band with intermediate mobility was present at low concentrations of C.EcoRV. At higher concentrations of C.EcoRV, an additional band of lower mobility appeared. The results are consistent with a view that at high concentrations, two C.EcoRV dimers simultaneously interact with the C-box DNA, producing a complex with the lowest electrophoretic mobility. This interpretation is reinforced by results obtained with DNA fragments harboring mutations in C-box1 or C-box2, where only the intermediate mobility band was present. No shifted bands were observed with the DNA fragment harboring mutations in both C-box operators.

The results of the binding experiments presented in Figure 6 were quantified and subjected to nonlinear regression analysis to extract the values of equilibrium dissociation constants and to assess potential cooperativity of C.EcoRV interactions with the two C-box operators. The results of such analysis for DNA fragments containing a single functional C-box, i.e. harboring mutations either in C-box1 or C-box2, are presented in Figure 7A and B, respectively. Experimental data were fitted to a simple binding model involving a formation of a 1:1 complex between C.EcoRV dimer and each C-box operator. The data fitted well to this simplest model, therefore more complex models (e.g. involving monomer–dimer C.EcoRV association linked to DNA-binding) were not considered. Recovered equilibrium dissociation constants were 502 ± 170 nM (binding to a C-box2 only) and 115 ± 17 nM (binding to a C-box1 only). Figure 7C illustrates analysis of binding to the wild-type EcoRV fragment. In this case, C.EcoRV can interact with each C-box operator either in the presence or in the absence of the protein bound to the other operator. The ratio of equilibrium dissociation constants for these two binding scenarios provides information regarding potential cooperativity of binding to the two operators. When the protein binds to the two sites independently, this ratio is expected to be 1. Deviations from the ratio of 1 are indicative of either positive (ratio <1) or negative cooperativity (ratio >1). Fitting of the data for the wild-type EcoRV DNA fragment to a model involving formation of 1:1 C.EcoRV dimer–DNA complexes (protein bound either to C-box1 or C-box2) and 2:1 C.EcoRV dimer–DNA complexes (protein bound to both C-box1 or C-box2) resulted in a ratio value of 0.8 ± 0.3, i.e. was not significantly different from 1. Based on these quantitative analyses, we conclude that C.EcoRV binds to the two C-boxes independently without significant cooperativity and with ~5-fold higher affinity to C-box1 compared to C-box2.

DISCUSSION

For host cell survival, it is essential that sufficient amounts of the M.EcoRV enzyme are produced before R.EcoRV synthesis commences. Our results indicate that during the establishment of a plasmid harboring the EcoRV genes in a naïve host (a situation mimicked in cells harboring pM or pR plasmids in the absence of pEF42) the steady-state amount of P*eco*RV•CR transcripts is low, while P*eco*RV•M levels are high, which should result in preferential synthesis of M.EcoRV and modification of host DNA.

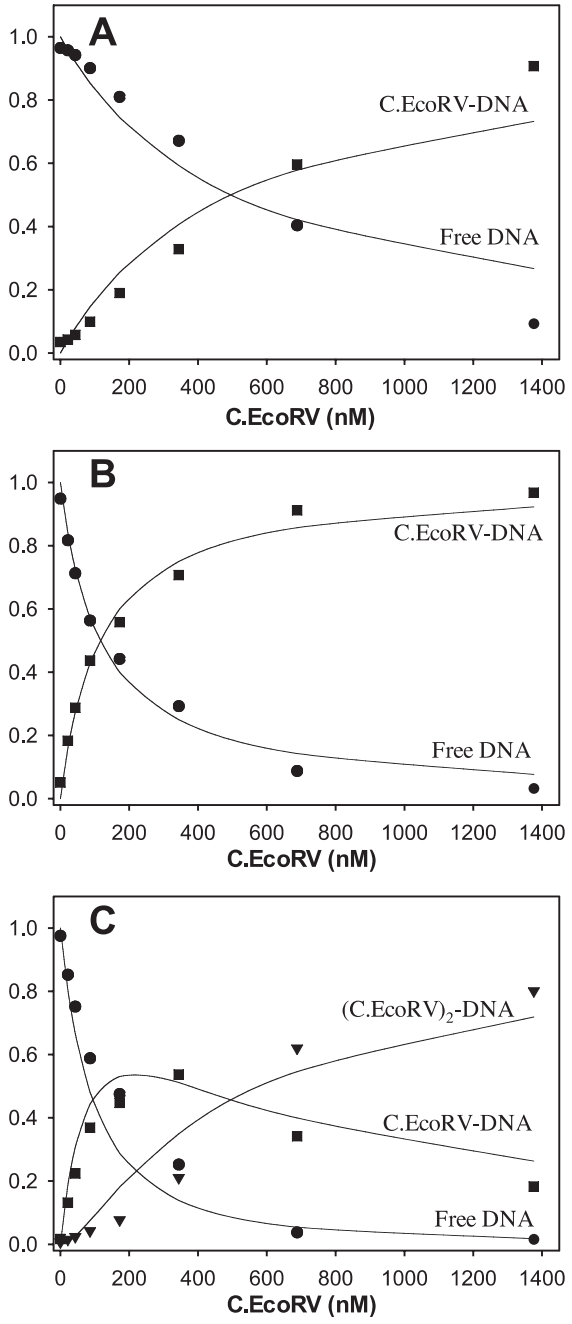

Figure 7. Quantitative analysis of C.EcoRV interaction with *eco*RV DNA. (**A**) Binding of C.EcoRV to DNA harboring mutation in C-box1, (**B**) binding of C.EcoRVC to DNA harboring mutation in C-box2, (**C**) binding of C.EcoRV to DNA containing wt C-box1 and wt C-box2. Data shown correspond to the average of 2–3 independent gel mobility shift experiments. Fractions of free DNA and protein–DNA complexes at each protein concentration were calculated as a ratio of band intensity of the band of interest to the sum of intensities of all bands in the lane. Solid lines represent the best nonlinear regression fit of experimental data to an appropriate binding model, as described in the text. Nonlinear regression was performed using SCIENTIST (Micromath Scientific Software, Salt Lake City, UT).

Weak transcription from the *eco*RV•CR promoter leads to gradual accumulation of the C.EcoRV protein. C.EcoRV binds strongly to P*eco*RV•CR-distal C-box1 leading to very strong activation of the *eco*RV•CR promoter and further increase in C.EcoRV (and R.EcoRV) synthesis. The runaway synthesis of C.EcoRV (and R.EcoRV) is regulated by C.EcoRV-binding to low-affinity C-box2, which attenuates activated transcription from P*eco*RV•CR. The resulting level of transcription from P*eco*RV•CR becomes comparable to transcription from P*eco*RV•M. The negative effect of C.EcoRV-binding on overall levels of the *eco*RV•M promoter transcription appears to be minimal, at least at our conditions.

The principles of regulation that emerge from our analysis of the EcoRV system should be generally applicable to other R–M systems that rely on C proteins for expression of their genes. These principles are significantly different from those operating in other R–M systems, where transcription from the initially strong methyltransferase promoter is strongly down-regulated either by direct binding of methyltransferase to promoter region (20–22) or by covalent modification (methylation) of promoter DNA (23) which indirectly leads to activation of restriction endonuclease promoter.

Transcript slippage occurs during transcription initiation on both P*eco*RV•M and P*eco*RV•CR and is a consequence of repetitive DNA sequences at transcription initiation start points of both promoters. While C.EcoRV changes the level of transcript slippage on P*eco*RV•M, the overall level of transcription from this promoter is the same in the absence or in the presence of the C protein, suggesting that the slippage does not significantly contribute to regulation of EcoRV genes' expression at our conditions. It should be noted, however, that we can not exclude that P*eco*RV•M transcripts initiated in the presence of and in the absence of C.EcoRV are translated with different efficiency, i.e. the levels of functional P*eco*RV•M transcripts may change in a C.EcoRV-dependent manner. Moreover, slippage may conceivably interfere with productive initiation from both *eco*RV promoters at conditions that alter the intracellular concentrations of CTP (for P*eco*RV•M) or ATP (for P*eco*RV•CR). Regulatory transcript slippage has been described for several promoters expressing genes whose products are involved in nucleotide biosynthesis [see, e.g. Cheng *et al*. (18)]; the significance, if any, of such a regulation for EcoRV is not known. We note that the appearance of primer extension product bands from the P*pvu*IIR (10) a promoter that is also a subject of activation by a C protein, is consistent with multiple 5′ end of RNA generated by transcript cleavage. Knowle *et al*. (24) had argued that multiple 5′ end of PvuII transcripts may be an artifact of primer extension by reverse transcriptase. We think this is unlikely in our case since (i) our protocol results in the appearance of different non-template-encoded nucleotides for P*eco*RV•CR and p*Eco*RV•M transcripts and (ii) no non-template-encoded residues were detected when the same protocol was used to identify transcription initiation start points from multiple phage promoters (16,17). Thus, transcript slippage may be a general phenomenon for C protein-regulated promoters.

The P*eco*RV•CR-distal C-box1 is centered ~33 bp upstream of the run of 8 As where most of the *eco*RV•CR transcription originates. Due to transcript slippage, it is impossible to unambiguously determine the P*eco*RV•CR transcription start site. However, since transcripts containing as little as 1–3 As on their 5′ end can be detected, it seems plausible that transcription from P*eco*RV•CR initiates at the end of the A run. This translates into a ~40 bp distance between C-box1 center and the P*eco*RV•CR transcription

initiation start site. A similar distance between the site of activator binding and transcription initiation start point is found in bacterial promoters activated by protein–protein contacts with σ^{70} region 4 or α subunit N-terminal domain (25,26). *In vitro* experiments with appropriate RNAP mutants will be necessary to establish the contact that is used for transcription activation by C.EcoRV and other C proteins.

The distance between the proximal C-box2 and the P*eco*RV•CR transcription start site is ~24 bp which is too short for transcription activation through most common mechanisms found in bacteria. However, a protein bound at such a distance from the transcription initiation site is likely to inhibit transcription, as is indeed observed. Thus, the relative locations of the P*eco*RV•CR transcription start site and C.EcoRV-binding sites are consistent with the general view of P*eco*RV•CR regulation outlined above.

The C proteins are related to λ repressor, a prototypical regulator that operates a molecular switch during the lysis-lysogeny decision. The C.EcoRV-operated switch is clearly simpler than the λ switch: it contains only two operators (three in λ), there is no cooperativity between operator-bound C.EcoRV molecules (cooperative binding of λ repressors is essential for an all-or-none behavior of the λ switch), and only one of the *eco*RV promoters, P*eco*RV•R, is regulated by C.EcoRV (λ repressor binding to its operators regulates two divergent promoters). Despite these differences, the molecular mechanism of transcription activation by C.EcoRV and λ repressor may be identical. Site-directed mutagenesis of C protein residues homologous to λ repressor residues that when mutated result in *p*ositive *c*ontrol (*pc*) phenotypes (19) will be required to test this conjecture.

Both C-box half-sites are located downstream of the P*eco*RV•M transcription start site. The distal half-site is centered ~21 bp downstream, so it is not surprising that C protein-binding has little effect on P*eco*RV•M transcription. The center of the proximal site is only 6 bp downstream of the P*eco*RV•M transcription start site, and so an inhibitory effect of C.EcoRV-binding to this site is expected. However, since the proximal half-site binding is relatively weak, the effect observed *in vivo* is small, possibly because not enough C protein is produced at our conditions for complete occupancy of this site. It is also possible that P*eco*RV•M transcription is truly C.EcoRV-independent. This is supported by comparative analysis of other R–M systems that are known to be regulated by C proteins. For example, the *ahd*I R and M genes are convergent and the C-box is only found in front of the co-transcribed *ahd*I•C and *ahd*I•R messages (27). The *ahd*I•M transcription must therefore be independent of C.AhdI and is likely constitutive.

The overlapping nature of methyltransferase and restriction endonuclease promoters in EcoRV and PvuII significantly complicates mapping of promoter elements in these systems, since most genetic alterations will affect both promoters. Therefore the deletion mapping results obtained using promoter fusions (21) may be a complex mixture of direct (alteration in intrinsic promoter strength of a promoter studied) and indirect (effects on competition between divergent promoters) effects. It therefore appears that further studies of simpler systems such as AhdI will be needed to uncover the molecular mechanism of transcription activation by C proteins.

ACKNOWLEDGEMENTS

We are grateful to Dr S. Nechaev for critical reading of the manuscript. This work was supported by NIH GM grant R01 59295 and NIH FIRCA Research Grant R03 TW07145 to (KS), and NIH GM R01 50514 (to T.H.). E.S. was partially supported by Charles and Johanna Busch postdoctoral fellowship. Funding to pay the Open Access publication charges for this article was provided by NIH R01 grant GM59295.

Conflict of interest statement. None declared.

REFERENCES

1. Jeltsch,A. and Pingoud,A. (1996) Horizontal gene transfer contributes to the wide distribution and evolution of type II restriction-modification systems. *J. Mol. Evol.*, **42**, 91–96.
2. Kobayashi,I. (2001) Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res.*, **29**, 3742–3756.
3. Nakayama,Y. and Kobayashi,I. (1998) Restriction-modification gene complexes as selfish gene entities: roles of a regulatory system in their establishment, maintenance, and apoptotic mutual exclusion. *Proc. Natl Acad. Sci. USA*, **95**, 6442–6447.
4. Tock,M.R. and Dryden,D.T. (2005) The biology of restriction and anti-restriction. *Curr. Opin. Microbiol.*, **8**, 466–472.
5. Ives,C.L., Nathan,P.D. and Brooks,J.E. (1992) Regulation of the BamHI restriction-modification system by a small intergenic open reading frame, bamHIC, in both *Escherichia coli* and *Bacillus subtilis*. *J. Bacteriol.*, **174**, 7194–7201.
6. Anton,B.P., Heiter,D.F., Benner,J.S., Hess,E.J., Greenough,L., Moran,L.S., Slatko,B.E. and Brooks,J.E. (1997) Cloning and characterization of the BglII restriction-modification system reveals a possible evolutionary footprint. *Gene*, **187**, 19–27.
7. Rimseliene,R., Vaisvila,R. and Janulaitis,A. (1995) The eco72IC gene specifies a *trans*-acting factor which influences expression of both DNA methyltransferase and endonuclease from the Eco72I restriction-modification system. *Gene*, **157**, 217–219.
8. Zheleznaya,L.A., Kainov,D.E., Yunusova,A.K. and Matvienko,N.I. (2003) Regulatory C protein of the EcoRV modification-restriction system. *Biochemistry (Mosc)*, **68**, 105–110.
9. Cesnaviciene,E., Mitkaite,G., Stankevicius,K., Janulaitis,A. and Lubys,A. (2003) Esp1396I restriction-modification system: structural organization and mode of regulation. *Nucleic Acids Res.*, **31**, 743–749.
10. Vijesurier,R.M., Carlock,L., Blumenthal,R.M. and Dunbar,J.C. (2000) Role and mechanism of action of C. PvuII, a regulatory protein conserved among restriction-modification systems. *J. Bacteriol.*, **182**, 477–487.
11. Heidmann,S., Seifert,W., Kessler,C. and Domdey,H. (1989) Cloning, characterization and heterologous expression of the SmaI restriction-modification system. *Nucleic Acids Res.*, **17**, 9783–9796.
12. Tao,T., Bourne,J.C. and Blumenthal,R.M. (1991) A family of regulatory genes associated with type II restriction-modification systems. *J. Bacteriol.*, **173**, 1367–1375.
13. Bart,A., Dankert,J. and van der Ende,A. (1999) Operator sequences for the regulatory proteins of restriction modification systems. *Mol. Microbiol.*, **31**, 1277–1278.
14. McGeehan,J.E., Streeter,S.D., Papapanagiotou,I., Fox,G.C. and Kneale,G.G. (2005) High-resolution crystal structure of the restriction-modification controller protein C.AhdI from Aeromonas hydrophila. *J. Mol. Biol*, **346**, 689–701.
15. Zakharova,M.V., Kravets,A.N., Klimashina,N.V. and Solonin,A.S. (1991) Properties of the replicator region of the natural plasmid pLG13 containing genes of the EcoRV restriction-modification system. *Mol. Biol. (Mosk.)*, **25**, 1615–1625.
16. Semenova,E., Djordjevic,M., Shraiman,B. and Severinov,K. (2005) The tale of two polymerases: transcription profiling and gene expression strategy of bacteriophage Xp10. *Mol. Microbiol.*, **55**, 764–777.
17. Minakhin,L., Semenova,E., Liu,J., Vasilov,A., Severinova,E., Gabisonia,T., Inman,R., Mushegian,A. and Severinov,K. (2005) Genome and gene expression of *Bacillus anthracis* bacteriophage Fah. *J. Mol. Biol.*, **354**, 1–15.

18. Cheng,Y., Dylla,S.M. and Turnbough,C.L.,Jr (2001) A long T. A tract in the upp initially transcribed region is required for regulation of upp expression by UTP-dependent reiterative transcription in *Escherichia coli*. *J. Bacteriol*., **183**, 221–228.
19. Ptashne,M. (2005) Regulation of transcription: from lambda to eukaryotes. *Trends Biochem. Sci*., **30**, 275–279.
20. Som,S. and Friedman,S. (1994) Inhibition of transcription *in vitro* by binding of DNA (cytosine-5)-methylases to DNA templates containing cytosine analogs. *J. Biol. Chem*., **269**, 25986–25991.
21. Karyagina,A., Shilov,I., Tashlitskii,V., Khodoun,M., Vasil'ev,S., Lau,P.C. and Nikolskaya,I. (1997) Specific binding of sso II DNA methyltransferase to its promoter region provides the regulation of sso II restriction-modification gene expression. *Nucleic Acids Res*., **25**, 2114–2120.
22. Christensen,L.L. and Josephsen,J. (2004) The methyltransferase from the LlaDII restriction-modification system influences the level of expression of its own gene. *J. Bacteriol*., **186**, 287–295.
23. Zakharova,M., Minakhin,L., Solonin,A. and Severinov,K. (2004) Regulation of RNA polymerase promoter selectivity by covalent modification of DNA. *J. Mol. Biol*., **335**, 103–111.
24. Knowle,D., Lintner,R.E., Touma,Y.M. and Blumenthal,R.M. (2005) Nature of the promoter activated by C.PvuII, an unusual regulatory protein conserved among restriction-modification systems. *J. Bacteriol*., **187**, 488–497.
25. Dove,S.L., Darst,S.A. and Hochschild,A. (2003) Region 4 of sigma as a target for transcription regulation. *Mol. Microbiol*., **48**, 863–874.
26. Niu,W., Kim,Y., Tau,G., Heyduk,T. and Ebright,R.H. (1996) Transcription activation at class II CAP-dependent promoters: two interactions between CAP and RNA polymerase. *Cell*, **87**, 1123–1134.
27. Streeter,S.D., Papapanagiotou,I., McGeehan,J.E. and Kneale,G.G. (2004) DNA footprinting and biophysical characterization of the controller protein C.AhdI suggests the basis of a genetic switch. *Nucleic Acids Res*., **32**, 6445–6453.