Tuning the relative affinities for activating and repressing operators of a temporally regulated restriction-modification system

Iwona Mruk1,2 and Robert M. Blumenthal1,3,*

1Department of Medical Microbiology and Immunology, University of Toledo Health Sciences Campus, Toledo, OH 43614-2598, USA, 2Department of Microbiology, University of Gdansk, Gdansk, 80-822, Poland and 3Program in Bioinformatics and Proteomics/Genomics, University of Toledo Health Sciences Campus, Toledo, OH 43614-2598, USA

Received October 24, 2008; Revised November 26, 2008; Accepted December 3, 2008

ABSTRACT

Most type II restriction-modification (R-M) systems produce separate endonuclease (REase) and methyltransferase (MTase) proteins. After R-M genes enter a new cell, MTase activity must appear before REase or the host chromosome will be cleaved. Temporal control of these genes thus has life-or-death consequences. PvuII and some other R-M systems delay endonuclease expression by cotranscribing the REase gene with the upstream gene for an autogenous activator/repressor (C protein). C.PvuII was previously shown to have low levels early, but positive feedback later boosts transcription of the C and REase genes. The MTase is expressed without delay, and protects the host DNA. C.PvuII binds to two sites upstream of its gene: OL, associated with activation, and OR, associated with repression. Even when symmetry elements of each operator are made identical, C.PvuII binds preferentially to OL. In this study, the intra-operator spacers are shown to modulate relative C.PvuII affinity. In light of a recently reported C.Esp1396I-DNA co-crystal structure, in vitro and in vivo effects of altering OL and OR spacers were determined. The results suggest that the GACTnnnAGTC consensus is the primary determinant of C.PvuII binding affinity, with intra-operator spacers playing a fine-tuning role that affects mobility of this R-M system.

INTRODUCTION

Many bacteria and archaea possess restriction-modification (R-M) systems (1), at least in part for defense against DNA bacteriophages. PvuII, like other type II R-M systems (2), includes two separate enzymes: a restriction endonuclease (REase) that cleaves DNA at a target sequence, and a methyltransferase (MTase) that modifies the same sequence to protect it from the cognate REase (3–8).

The PvuII R-M system is naturally carried on a plasmid together with mobilization genes (9,10). If the system is to be mobile, the REase and MTase activities must be carefully balanced in a relatively host-independent manner, so as to minimize killing of new host cells that initially have completely unmethylated chromosomes. Any R-M system misregulation would lead to auto-restriction and cell death if REase activity is too high (11,12), or would weaken defense against bacteriophage invasion if MTase activity is too high (13). In addition to the evident effects of R-M systems on resistance to bacteriophages (14), the roles of restriction in modulating gene flow between bacteria are also clear (15–19). The regulation of R-M systems having only REase and MTase genes is under active investigation (20,21).

A subset of type II R-M systems contains regulatory genes in addition to the MTase and REase genes. The regulatory C (controlling) gene was first discovered in the PvuII (22) and BamHI (23) R-M systems. Subsequently, active regulatory genes have been demonstrated in several other systems (24). The C proteins are quite small [PvuII has a subunit MW of 9.4 kDa (25)] and appear to have a remarkably broad host range for their action (26,27).

C proteins, where tested, activate and repress their own transcription autogenously (28,29). In the PvuII R-M system at least (Figure 1A and B), the C protein is responsible for delayed REase appearance after entering a new host cell (30); this gives the MTase time to protect the new host’s genome. In R-M systems having a C gene, the REase gene typically does not have its own promoter

*To whom correspondence should be addressed. Tel: +1 419 383 5422; Fax: +1 419 383 3002; Email: Robert.Blumenthal@utoledo.edu

© 2009 The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
The C proteins act on conserved operator sequences called ‘C boxes’ (33,35,36). Functional C boxes have been determined for C.PvuII (33) and several other C proteins (24,29,37,38). We have previously explored the roles of the GACTnnnAGTC symmetry elements and the inter-operator TGTA spacer (28). We found that any substitution within the TGTA eliminated transcription activation by C.PvuII, and noted that the PvuII-related C-boxes fall into two classes based on this spacer (TGTA versus CGTA), with co-variation of the cognate C protein predicted recognition helices (HRTY versus DRTY) (28).

The palindromic nature of C boxes suggests that each of the two C boxes (designated as 1AB or O_L, operator left, and 2AB or O_R, operator right; Figure 1B) could each be bound by the helix–turn–helix motifs of a C protein homodimer (39). The two operators have conserved center-to-center spacing of 15 bp (Figure 1C), meaning that the two C homodimers occupy opposite faces of the DNA double helix (25,33,40).

Recently, a C protein–DNA operator co-crystal structure was solved for C.Esp1396I (41). As expected, the two C protein dimers interact with C-box DNA on opposite faces of the DNA. Surprisingly, a second symmetry pattern (41) (yellow in Figure 2A) was identified in addition to the one previously described (33) (blue in Figure 2A). The center of symmetry is shifted by 0.5 nt between the two patterns. The C.Esp1396I protein/C0 DNA co-crystal structure suggests that the second symmetry pattern is indeed recognized (41). However, the relative roles of the two symmetries are not yet clear.

C.PvuII binding to O_L activates transcription, while its binding to O_R represses (28). However C protein binding to O_L and O_R is highly cooperative, with a Hill coefficient of about 4 (28). Further, when the GACTnnnAGCT symmetry elements of O_L and O_R are made identical, activation still occurs (28). This alone might be explained by a proposed in vivo competition at O_R between C.PvuII and RpoD (σ^70) (42). However, we also find a C.PvuII binding preference for O_L over O_R in vitro, even with identical symmetry elements and in the absence of RpoD (28). This result suggests a key role for the two asymmetrical intra-operator spacer sequences (boxes in Figure 1C). In this paper we focus on the role of these spacers, their contribution to C protein affinity for the two operators and the resulting impact on R-M system establishment.
MATERIALS AND METHODS

Strains, phages and plasmids

The *Escherichia coli* K-12 strains used in this study are described below. All strains into which *pvuIIM* is introduced must lack the *mcrBC* restriction system (9,43,44). MC1061 [araD139/C1(ara, leu)7697, lacX74, galU, galK, hsdR, strA] (45) is able to transport arabinose but is deficient in its metabolism; it was used as the host for *in vivo* titrations with C.PvuII. TOP10 (Invitrogen; mcrA Δmrr-lsdRMS-mcrBC φ80lacZM15 lacX74 recA1 araD139 (ara-leu)7679 galU galK rpsL(StrR) endA1 nupG) was used for all other purposes including cloning steps and CAT assay. [Top10 with F'/lacI, Tn10(lacI)] was used as the host strain for M13 phage infections. The plasmids used are listed in Table S1. Recombinant phage M13pvuIIwt were handled as described before (30).

In *in vivo* titration of C.PvuII and M.PvuII

PCR-amplified genes for C.PvuII or M.PvuII were cloned downstream of the arabinose-inducible P_{BAD} promoter in vector pBAD24 (46) yielding plasmids pIM1 (28) and pBadMTwt, respectively. Arabinose induction yielded

---

Figure 2. Testing the two alternative symmetry patterns in the C-box region. (A) Two alternative symmetry patterns. The top portion shows the 'AGTC' consensus [in blue, (33)] comprising the C-box elements themselves, the actual sequence from PvuII R-M system, and the 'TATA' consensus [in yellow, (41)]. The centers of symmetry are shown. The numbers at the right indicate the extent of match to each consensus, and whether a characteristic repression complex forms in *in vivo* at ≤500 nM C.PvuII. The lower portion shows several previously tested C.PvuII operator variants (28). Magenta shading indicates mutations introduced in each case, in the context of symmetrized C-boxes (underlined positions). (B) Binding effects of 'TATA' consensus alteration. A series of 126-bp dsDNA-binding targets were prepared by PCR amplification, and included in each binding reaction at 20 nM. The DNAs contained WT or variant C boxes flanked on either side by 50 bp of native PvuII sequence. EMSA reactions with concentrations of C.PvuII increasing from 0–500 nM were processed as outlined in ‘Materials and methods’ section. PvuSym indicates wild-type spacers in the context of symmetrized C-box sequences (see sequence in A). Other variants are in the same sequence context, but with different intra-operator spacers as indicated (O_1/O_2). Reactions were resolved on 10% native polyacrylamide gels, and DNA was visualized by staining with ethidium bromide. Numbers at right indicate the expected numbers of bound C.PvuII dimers. (C) *In vivo* titration of intra-operator spacers variants with C.PvuII. Cells, carrying *pvuIIC* under the control of P_{BAD}, were grown in minimal media with 0.2% glucose and the indicated concentration of arabinose, as described before (28). Expression from variants (labeled as in B) was measured via transcriptional fusion of the variant C-boxes/P_{pvuIICR} to reporter gene lacZ. Each point represents β-galactosidase specific activity, determined by linear regression of the plot of LacZ activity (modified Miller units) versus optical density of the culture, as previously (49). Each point represents the slope of a regression from at least 3 points; in all cases R^2 > 0.97. The beginning of the curve (0), indicates values obtained in glucose with no arabinose. Black diamonds represent LacZ activity from cells with the WT symmetrized PvuII C-boxes with CAT/CAA intra-operator spacers (pIM8). The 'TATA' consensus variant (TAT/TAT) is shown as blue circles, and the 'TATA'-disrupted variant (CGC/CGC) is shown as red diamonds. As negative control vector plasmid with no *pvuIIC* was used (open circles).

---

Nucleic Acids Research, 2009, Vol. 37, No. 3 985
graded gene expression, from near 0 to above the physiological wild-type level (46). Experiments were performed in MOPS-minimal medium (Teknova) with 0.2% glucose as carbon source (47) as described before (28). Briefly, single colonies were used to inoculate overnight cultures in MOPS media and appropriate antibiotics. These cultures were diluted 1:50 into the same medium but without antibiotics, and grown with shaking to an OD_{600 nm} of 0.2–0.3. The cells were then gently pelleted, resuspended and divided among flasks containing MOPS-minimal media with varied concentrations of L-arabinose.

**β-Galactosidase assay**

The LacZ assays were based on hydrolysis of o-nitrophenyl-β-D-thiogalactoside (48) as described (49). Briefly, β-galactosidase activity and culture density were measured at 20–30-min intervals during exponential growth. The units for this assay were calculated by dividing the measured A_{420 nm} (released nitrophenol) by the time the assay was done by fitting relative mobilities to the quadratic function

$$y = ax^2 - bx + c,$$

where $a = -b = 2r(1 \cos z)$ and $z$ is the bend angle to be determined (52).

**Western blot analysis**

This analysis was carried out exactly as described (28). Briefly, equal volumes of culture were centrifuged, supernatants were removed and the cell pellets stored at −80°C. Pellets were resuspended in 1× SDS buffer (Novagen), and lysed by heating at 98°C for 10 min. Equal amounts of protein were loaded onto a 4–12% Bis–Tris NuPAGE Novex gradient gel (Invitrogen) and electrophoresed at 100 V in 1× NuPAGE MES buffer (Invitrogen). Proteins were then electroblotted to PVDF membranes and detected by fluorescence using the ECL-plus Western Blotting Detection System (GE Health Sciences) with 1:5000 dilution of rabbit anti-C.PvuII polyclonal serum (Strategic Biosolutions), and a 1:25000 dilution of horse-radish peroxidase-conjugated goat anti-rabbit IgG. Protein bands were visualized on an Omega Molecular Imaging System (UltraLum). The prestained MW markers used were SeeBluePlus (Invitrogen).

**Efficiency of transformation assay**

Efficiency of transformation (EOT) is defined in this study as the relative number of transformants obtained from a given preparation of competent cells, using a given amount of plasmid DNA. EOT is calculated from the ratio of transformants with a test plasmid relative to those with a control plasmid (usually pBR322). This term is equivalent to ‘relative transformation efficiency’. For plasmid transformation the standard CaCl$_2$-heat shock method was used (53).

**Relative restriction activity assay**

The restriction activity of *E. coli* cells carrying PvuII R-M system and its variants was measured through the plaquing efficiency of phage λvir. The efficiency of plaquing (EOP) of λvir was calculated as the ratio of plaques formed on *E. coli* TOP10 containing plasmid pBR322 to those formed on the same strain containing a plasmid with the PvuII R-M system or its variants.
In vitro kinetics of chloramphenicol acetyltransferase (CAT) production following M13pvuIIwt infection

M13pvuIIwt stock and infection procedures were used as previously optimized and detailed (30). Briefly, E. coli TOP10F’ cells (Invitrogen) were grown in LB supplemented with tetracycline (to maintain the F’ episome). At OD600 nm of 0.25, E. coli cultures containing plasmids with promoter/operator variants fused to a promoterless cat gene were infected with M13pvuIIwt (30) at a multiplicity of infection (MOI) of 15. Duplicate samples were collected for each culture over a 100 min postinfection time course. One sample was immediately used to determine culture density, while the other was pelleted and stored at –20°C until used for CAT assay. Two separate experiments were performed on consecutive days. CAT production was quantified colorimetrically based on a sandwich ELISA method (CAT ELISA kit, Roche) following the previously described protocol (30).

M.PvuII initiator mutants

Transcriptional (operon) fusions of C-box variants to lacZ were generated by site-directed mutagenesis (Quick Change, Stratagene) of the WT P<sub>prad</sub><sup>AC/L</sup>–lacZ fusion in plasmid pDK435 (28). For translational fusions, a HindIII–SpeI Klenow-filled fragment from the WT PvuII R-M system was inserted into vector pLex3B (54) that had been cleaved with XhoI and Smal and Klenow filled. This yielded pLex-met1, in which lacZ is preceded by 26 <i>pvuII</i> codons in the same reading frame. Individual M→L replacements were introduced by site-directed mutagenesis. All generated mutations were sequence confirmed. LacZ assays were performed as described in earlier section.

MTase activity assay

The M.PvuII gene was amplified by PCR (primers: ACTT TGAATCTCAGACTATGAG and GTTGCTGCACT ACGAACTCAA), cleaved with PstI and cloned after the araBAD promoter of pBAD24 (cleaved with NcoI, Klenow-filled then cleaved with PstI). The resulttant pBadMTwt was used as a template for site-directed mutagenesis, to obtain pBadMT2 and pBadMT3 (Table S1). To replace the plasmid’s origin of replication and antibiotic marker, all three derivatives were separately cleaved with ClaI and ApaLI, then Klenow-filled, and a fragment encompassing <i>P<sub>bad</sub></i>–<i>pvuII</i> was ligated to an Nhel–ApaLI Klenow-filled fragment from plasmid pIM4 (Table S1). This fragment contains the p15A replication origin and a gene for kanamycin resistance, making the plasmids compatible with the R.PvuII substrate plasmid pUC × 7PvuII. Plasmids pBadMTwt-kan and its variants [pBadMT2-kan (N45/S10R) and pBadMT3-kan (N45/S8L/S10R) (Table S1)] were introduced into <i>E. coli</i> TOP10 together with high copy pUC × 7PvuII plasmid (Table S1) carrying seven PvuII sites. The dual transformants were grown in MOPS-glucose media as described above. After 4-h induction with different arabinose concentrations (from 0 to 0.2%), cells were harvested and their plasmid DNA was isolated. XhoI enzyme was added to linearize pBad derivatives, which have no PvuII site. Equal amounts of DNA (400 ng) digested with 10 u of PvuII restriction enzymes (NEB) were run on 1% agarose TAE gels to compare the methylation status of the substrate plasmid.

RESULTS

Intra-operator spacers differ between O<sub>L</sub> and O<sub>R</sub>

Recently sequenced bacterial genomes are revealing many putative C protein-like open reading frames, and their upstream C-box operator sequences. We have analyzed the C box/promoter regions for 21 PvuII-like C proteins (25,28,33) and found that the boxes resemble two pairs of gapped symmetrical inverted repeats linked by a central spacer: GACTnnnAGTC(TGTA)GACTnnnAGTC (Figure 1). The WT PvuII C box region differs from the symmetrical pattern at just two of 16 positions: in boxes 1A (GACT→TACT) and 2A (GACT→GATT) (Figure 1B and C). Logo analysis (55) for C proteins having a HRTY recognition helix revealed that the central 4-nt TGTA spacer is more highly conserved than the symmetry elements themselves (Figure 1C) (25,28). We found that all four bases of the ‘inter’-operator TGTA spacer are required for C.PvuII-dependent activation (28). The 3-nt ‘intra’-operator spacers within O<sub>L</sub> and O<sub>R</sub> (‘nnn’ above) are also well conserved; for PvuII the actual spacers are CAT and CAA, respectively (Figure 1B and C). These spacers form part of an alternative symmetry among C-boxes that was recently found in interaction of E. coli with its C-box DNA (41).

We have found that C.PvuII has higher in vitro affinity for O<sub>L</sub> than for O<sub>R</sub>, even when the same symmetrical C-box core sequence is used in both O<sub>L</sub> and O<sub>R</sub> (GACTeatAGTCtgtaGACTteaaAGTC) (‘PvuSym’ in Figure 2A) (28). This result suggested that the two 3-bp intra-operator spacer sequences (and/or their associated alternative symmetry) might determine the relative affinity of C.PvuII for the two operators. Logo analysis reveals that the preferred sequence for both operator spacers is TAT, with particularly high conservation of the A position at the center (Figure 1C). This is consistent with the E. coli C-box co-crystal structure (41), but does not explain conservation of the inverted repeats (Figure 1C). The spacer consensus is the same for O<sub>L</sub> and O<sub>R</sub> (Figure 1C), but an analysis of 21 C-box regions from the C.PvuII subfamily (28) reveals that only two of 21 (<10%) have the same spacer in both operators (Figure S1). We sought to explore the roles of the intra-operator spacers and the alternative symmetry of which they are a part.

Altering operator spacers does not prevent C.PvuII tetramer formation, and still yields an activation-repression in vivo

Our first aim was to investigate the relative importance of the two alternative symmetry elements for C.PvuII binding and C gene regulation: the AGTC C-boxes themselves (33) (blue in Figure 2A), as well as the recently reported intra-operator spacer-associated ‘TATA’-based symmetry (41) (yellow in Figure 2A). We used site-directed
mutagenesis to generate C-box intra-operator spacer variants. In two cases, both operator spacers \((O_L/O_R)\) were identical: TAT/TAT (pIM57) to perfectly match the ‘TATA’ symmetry based on C.Esp1396I structure data (41) (Figure 2A) or CGC/CGC (pIM68) to disrupt that pattern. We also made variants in which one operator had TAT and the other had CGC: TAT/CGC (pIM58) or CGC/TAT (pIM67).

We first carried out \textit{in vitro} EMSA. C.PvuII formed apparent tetramer complexes in each case, even when the ‘TATA’ consensus was replaced in both operators by CGCA; binding was weaker, but the low-mobility complex formed (Figure 2B). Interestingly, the CGC/TAT variant showed reduced cooperativity with an intermediate EMSA species appearing (Figure 2B). In contrast, our previous data from variants in the C-boxes themselves (28) (summarized in Figure 2A) revealed the absence of low-mobility complexes when just box 2A of \(O_R\) was reversed, even though the ‘TATA’ consensus was intact [Figure 2A, see also Figure 3B in (28)]. Together, these findings suggest that the AGTC symmetry plays a substantially larger role than the ‘TATA’ symmetry for C.PvuII \textit{in vitro} binding.

To test the effects of these changes \textit{in vivo}, the same spacer sequence variants were cloned in front of \(lacZ\). The resulting transcriptional fusions were used to

\[\text{Figure 3. Effect of exchanging or duplicating the native intra-operator spacers. (A) C.PvuII titration experiment performed as in Figure 2C. Closed symbols represent PvuIIICR–lacZ fusions with, respectively, the native PvuII C-boxes with CAT/CAA intra-operator spacers (filled diamonds, pDK435), the symmetrized C-box variant with native spacers (filled circles, pIM6), the symmetrized variant with CAA/CAA spacers (open circles, pIM13), with CAT/CAT spacers (filled squares, pIM10) and with CAA/CAT spacers (open squares, pIM14). The nonrepressing C-box mutant (pWWWR) was added as control (open diamonds). As C.PvuII negative control pBAD24 was used with pDK435 (triangles). (B) C.PvuII concentrations giving peak expression. Same data as shown in (A), but normalized to the highest expression value (LacZ modified Miller units) for each variant. (C) Same plasmids used in (A), but with C.PvuII delivered at a physiological level from the WT C-box/promoter in the intact R-M system (pPvuRM3.4, black bars), or elevated level from a nonrepressing C-box mutant of R-M system (pIM6, gray bars). A plasmid with no C.PvuII gene was used as a negative control (pBR322, white bars). (D) Arabinose-induced WT C.PvuII production visualized by western blotting. Plasmid pIM1 (also used in A) was in \(E.\ coli\) cells growing in the indicated arabinose concentration. The level of C.PvuII produced from pPvuRM3.4 and its nonrepressing variant (pIM6) were also determined. The asterisk indicates the C.PvuII level at 0.02% arabinose induction, corresponding to the highest level of \(P_{pVUICR}–lacZ\) transcription [peak for black diamonds in (A)].}\]
quantitate activation/repression profiles in cells where intracellular C.PvuII protein levels (from a P_{BAD} arabinose inducible promoter) could be varied from near 0 to above the physiological wild-type level (28). The TAT/TAT variant was not greatly affected relatively to WT-Sym operators, but appeared to enter the repression phase at a slightly lower C.PvuII concentration (Figure 2C). The radical CGC/GCG variant is poorly expressed, but still shows significant activation and apparent onset of repression by C.PvuII. Again, in comparison, reversal of C-box 2B resulted in complete loss of repression (next section and Figure 3A).

**Altered operator spacers can lead to repression at lower intracellular levels of C.PvuII**

We next investigated the role of the native intra-operator spacers (CAT in O_L and CAA in O_R) in the context of fully symmetrical C-boxes. We used site-directed mutagenesis to generate spacer variants upstream of a promoterless lacZ gene. In two cases, both operator spacers (O_L/O_R) were identical: CAT/CAT (pIM10) or CAA/CAA (pIM13). In the other two cases the spacers were exchanged (CAA/CAT, pIM14), or remained as in WT (CAT/CAA, pIM8). We also tested WT spacers in the context of the natural nonsymmetrical C-box (pDK435) (Table S1).

We titrated C.PvuII in these P_{pvuIIC/R-lacZ} transcriptional fusion strains using a compatible P_{BAD-pvuIIC} plasmid. Western blot analysis of the resulting ramp in C.PvuII levels is shown in Figure 3D. For WT C-boxes, as before (28), activation was seen with a peak 1.5-fold above baseline at about 0.02% arabinose, followed by strong repression at 0.2% arabinose to the level of the negative control (vector, no _pvuIIC_; Figure 3A). All tested intra-operator spacer variants, except for the reciprocally exchanged spacers (pIM14), showed very similar transcription profiles, though peak heights varied. The WT C-box fusions gave the highest maximal activation of the tested variants. The reciprocally exchanged spacer variant (CAA/CAT; pIM14) showed the same general profile with respect to inducer concentration, but with substantially decreased expression (about 10-fold less at peak) than its WT symmetrical variant at peak (pIM8).

A most striking finding was that the peak of expression for tested spacer variants depended on the spacer sequence in O_L, but not in O_R. We normalized the data shown in Figure 3A to the highest expression for each variant, to facilitate their comparison (Figure 3B). The two variants with CAA in O_L (regardless of the O_R spacer sequence) gave peak expression at 0.002% arabinose, in contrast to variants with CAT in O_L, for which the peak was at 0.02% arabinose. The ascending part of the expression profiles at C.PvuII concentration presumably correspond to the relative binding affinity of the activating operator O_L with CAA, which may be higher than with CAT spacer.

We next tested the levels of reporter gene transcription fused to the various C-box spacer variants in the context of the entire WT PvuII R-M system carried on the pBR322 vector (9), producing the physiological C.PvuII concentration in the cells growing exponentially at 37°C in MOPS-glucose media (as described in ‘Materials and methods’ section). The WT C.PvuII level is close to that produced in 0.008% arabinose (Figure 3A) as also shown by western blot (Figure 3D). The expression of each variant (black bars; Figure 3C) was compared to the position on the titration profile (Figure 3A). As a negative control, we used vector without the _pvuIIC_ gene (white bars, Figure 3C), and saw minimal expression due to C.PvuII-independent promoter activity. As a positive control we supplied the _pvuIIC_ gene in the context of the WT R-M system (pIM6; gray bars, Figure 3C), but with mutation of O_R C-box 2B that prevents repression [open diamonds in Figure 3A; (28)]. C.PvuII is elevated in these cells (Figure 3D) and we saw the expected drop in expression for all variants except the nonrepressing fusion itself (pIM9) (Figure 3C).

At the physiological C.PvuII level, the variants with two CAA or two CAT spacers are indistinguishable (black bars, Figure 3C), but the rest of the titration data indicate that the CAT/CAT operator is still in the transcription activation phase, while the CAA/CAA operator is undergoing repression (Figure 3A). These data suggest that the activation/repression circuit is not greatly disturbed by the CAT/CAA variants (Figure 3A), but just modulated.

**The O_L intra-operator spacer contributes more to net C.PvuII-binding affinity than the O_R spacer**

As noted in the previous section, the C-box region of PvuII and similar R-M systems includes two symmetry patterns. One (blue in Figure 2B) has two equivalent gapped palindromes referred to as O_L and O_R (33). From this information alone, one might expect the two intra-operator spacers to have equivalent effects with respect to C.PvuII binding. We asked if the intra-operator spacers might be involved in the determination of relative C.PvuII affinity to O_L and whether the _in vivo_ observations might be due to differential affinity binding.

We tested this in two ways, via EMSA and competition EMSA. We used 126-bp PCR products, amplified from various plasmids, at a fixed concentration of 20 nM. The 26-bp C boxes were located at the center of the amplification product, flanked on each side by 50 bp of DNA. For each variant we used 0–300 nM C.PvuII. As with the _in vivo_ lacZ fusions described above, we exchanged the two native intra-operator spacers, or made them equivalent. The EMSA results for all intra-operator spacer variants show the same distinct low-mobility complex obtained for WT C-boxes [28 and Figure S3A]. C.PvuII has substantially greater affinity for the fully symmetrical C-box variants than for the WT C-boxes, based on disappearance of the unshifted complex (Figures 4A and S1A). We tested this result further via competition EMSA, using unlabeled variant DNA fragments in 1-, 2.5-, 5- or 10-fold molar excess with a fixed amount of biotin-labeled WT C-box DNA. The results (Figures 4B and S4B) indicate significantly higher C.PvuII affinity for O_L with a CAA spacer than with a CAT spacer. With O_L = CAA, 10-fold excess of competitor resulted in nearly 90% loss of the shifted band; in contrast,
We next asked whether the O_L spacer effects might result from C.PvuII bends DNA within C-boxes/promoter region (Figure S5). None of these variations significantly altered cooperativity, as judged by Hill coefficients were identical. None of these variations significantly affected the temporally controlled balance between PvuII modification and restriction. We used site-directed mutagenesis to place the same operator spacer sequences as used in Figures 3 and 4 in the context of the entire pBR322 (Table 1). The nonrepressing plasmid (pIM6), but nonrepressing variant (pIM6) (28). Our previous data showed that lack of repression increases amounts of R-Ease relative to MTase, presumably increasing the risk of damage to the new host’s chromosome.

To study the effects of intra-operator spacers on establishment in new cells, plasmids carrying the intact R-M system or its variants were introduced into cells and the efficiency of transformation (EOT) was determined. Two sets of competent cells were prepared: one strain carried the pvuIM gene (M+; plasmid pPvuM1.9-ACYC), so the cells were fully protected before the R-M system was introduced, while the other strain lacked pvuIM (M-; pACYC184). The relative EOTs were expressed as the fraction of colony-forming units (CFUs) for M- versus M+, and then normalized to the values for vector control controls, we used the WT PvuII R-M system plasmid (pPvuIRM3.4; O_L = CAT; O_R = CAA) and its activated but nonrepressing variant (pIM6) (28). Our previous data (28) showed that lack of repression increases amounts of R-Ease relative to MTase, presumably increasing the risk of damage to the new host’s chromosome.

We compared bend angle calculated for WT C-boxes and the symmetrical variant with CAA spacer in O_L (Figure 5). We did not find a major difference in the mobility, with DNA being bent by 44.5 ± 1.5° for WT and 48.0 ± 0.5° for O_L spacer = CAA. We cannot, however, rule out the possibility that this slight difference might be enough to affect C.PvuII affinity (57).

**C.PvuII bends DNA within C-boxes/promoter region**

We next asked whether the O_L spacer effects might result from structural changes in the DNA. Other C proteins bend C box DNA, causing local distortion at both operator sites, as previously demonstrated for C.AhdI (40) and C.EcoO109I (38). To estimate the DNA bending angle, we cloned C-box DNA fragments into the center of the pBend2 plasmid (51), flanked by a variety of restriction sites. Restriction cleavage yielded a set of constant length DNA fragments, but with different relative location of the C-boxes. The electrophoretic mobility of DNA complexes with C.PvuII allow estimation of average bending (52,56). We did not expect a large net bending angle, because the 15-bp center-to-center spacing of the operators places them on opposite faces of the DNA helix, and bends at each operator should partially cancel one another. In fact, for the C.Esp1396I tetramer, the overall bend angle calculated from crystal structure analysis is about 43° and results from the two individual bends imposed by each dimer related by 150° rotation (41).

We compared bend angle calculated for WT C-boxes and the symmetrical variant with CAA spacer in O_L (Figure 5). We did not find a major difference in the mobility, with DNA being bent by 44.5 ± 1.5° for WT and 48.0 ± 0.5° for O_L spacer = CAA. We cannot, however, rule out the possibility that this slight difference might be enough to affect C.PvuII affinity (57).

**Higher affinity for O_L impairs establishment and maintenance of the PvuII R-M system**

When O_L contained a CAA spacer, reporter gene fusions revealed higher transcription from P_pvuIICR at lower C.PvuII concentrations, than when the WT CAT spacer was present (Figure 3B). This result corresponds to higher affinity in vitro, as revealed by EMSA (Figure 4B). We next determined how the change in C.PvuII affinity affected the temporally controlled balance between PvuII modification and restriction. We used site-directed mutagenesis to place the same operator spacer sequences as used in Figures 3 and 4 in the context of the entire PvuII R-M system. This yielded plasmids pIM17 (spacer O_L = O_R = CAT); pIM18 (spacer O_L = O_R = CAA) and pIM19 (spacer O_L = CAA; O_R = CAT) (Table S1). As controls, we used the WT PvuII R-M system plasmid (pPvuIRM3.4; O_L = CAT; O_R = CAA) and its activated but nonrepressing variant (pIM6) (28). Our previous data (28) showed that lack of repression increases amounts of R-Ease relative to MTase, presumably increasing the risk of damage to the new host’s chromosome.

To study the effects of intra-operator spacers on establishment in new cells, plasmids carrying the intact R-M system or its variants were introduced into cells and the efficiency of transformation (EOT) was determined. Two sets of competent cells were prepared: one strain carried the pvuIM gene (M+; plasmid pPvuM1.9-ACYC), so the cells were fully protected before the R-M system was introduced, while the other strain lacked pvuIM (M-; pACYC184). The relative EOTs were expressed as the fraction of colony-forming units (CFUs) for M- versus M+, and then normalized to the values for vector control transformations. The wild-type PvuII R-M system showed no significant difference in EOT from the vector control pBR322 (Table 1). The nonrepressing plasmid (pIM6), pIM20 (WT spacers, fully symmetrical C boxes) and pIM17 (spacer O_L = O_R = CAT) showed similar about...
3-fold reductions in EOT. The strongest effect on EOT was observed for two variants having O_L = CAA (pIM18, O_R = CAA; and pIM19, O_R = CAT). Both showed drastic >100-fold reductions in comparison to WT PvuII R-M (Table 1). These plasmids appear to have had selectively lethal effects on M.PvuII – cells, as their transformation of M.PvuII + cells was efficient.

We next tested whether this apparently lethal effect for O_L = CAA is due to a crucial temporal perturbation in R-M system establishment leading to premature REase expression. Cells carrying both a C-box operator variant plasmid and the compatible plasmid pPvuM1.9-ACYC (M+) were transformed with pACYC-kan (M–). One of the plasmids was expected to be lost during growth without antibiotics to permit plasmid segregation, since pPvuM1.9-ACYC (M+) and pACYC-kan (M–) belong to the same incompatibility group (p15A) (58).

Transformants were then plated with selection for both the R-M system plasmid and pACYC-kan (M –). The question was whether established R-M system variants still required the second MTase+ plasmid. Spacer variants with O_L = CAT, including WT and WT-Sym R-M system variants grew well without pPvuM1.9-ACYC (M+) (Table 1, Figure S2). In contrast, additional MTase activity was apparently indispensable when O_L spacer = CAA, as M– segregants were not found. This implies an inability of the variant R-M system to be maintained, even after establishment (Table 1), suggesting that elevated (rather than, or in addition to premature) REase expression is responsible for reduced EOT.

While we could not test variants with O_L = CAA without additional M+ activity present (as just described), we tested whether the observed ~30% reduction in EOT for

### Table 1. The effect of altered spacers in O_R or O_L on PvuII R-M system establishment and maintenance

<table>
<thead>
<tr>
<th>C box intraoperator spacers O_L/O_R</th>
<th>Relative EOT (%)</th>
<th>Ability to replace co-existing pPvuM1.9-ACYC (M+) with pACYC177-kan (M–)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No PvuII R-M (pBR322)</td>
<td>100 ± 7</td>
<td>+</td>
</tr>
<tr>
<td>CAT/CAA, WT</td>
<td>98 ± 13</td>
<td>+</td>
</tr>
<tr>
<td>(pPvuRM3.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAT/CAA, Sym (pIM20)</td>
<td>30 ± 3</td>
<td>+</td>
</tr>
<tr>
<td>CAT/CAA, Sym (pIM6)</td>
<td>34 ± 3</td>
<td>ND</td>
</tr>
<tr>
<td>CAT/CAT, Sym (pIM17)</td>
<td>26 ± 7</td>
<td>ND</td>
</tr>
<tr>
<td>CAA/CAT, Sym (pIM18)</td>
<td>0.2 ± 0.1 d</td>
<td>–</td>
</tr>
<tr>
<td>CAA/CAT, Sym (pIM19)</td>
<td>0.7 ± 0.4 d</td>
<td>–</td>
</tr>
</tbody>
</table>

*Equal amounts of plasmid DNAs with PvuII R-M system and its C-box spacer variants were used to determine the efficiency of transformation (EOT) in each of two host strains. These plasmids were introduced into competent E. coli TOP10 cells that already carried either the gene for the PvuII MTase (pPvuIM; plasmid pPvuM1.9-ACYC) or a vector control (pACYC177). Relative EOT was determined as the fraction of M.PvuII – transformants obtained relative to the number of transformants for the M.PvuIIM+ strain, and then normalized to the pBR322 EOT ratio. The standard deviation is indicated.

*The same competent cells, having indicated plasmid and pPvuM1.9-ACYC (M+), were transformed with pACYC-kan (M+) and grown without antibiotics for 2h to allow plasmid segregation. As they belong to the same incompatibility group, either pPvuM1.9-ACYC or pACYC-kan should be lost unless the plasmid is essential. Potential segregants were plated with selection for R-M system plasmid (Amp+) and pACYC-kan (M+) (Kan+). Lack of transformants after overnight incubation indicates a requirement of additional MTase expression for R-M system maintenance.

*ND, not determined.

*Smaller colonies, viability problems.

Figure 5. DNA bending by C.PvuII. WT C-boxes (A) and O_L spacer variant with CAA (B). With the indicated restriction enzymes 150-nt DNA fragments (60-ng each) were released, and contain the C-box in the center (EcoRV) or close to either end (MluI or BamHI). C.PvuII (200 ng) was added and EMSA was carried out (as in Figure 3). (C) Flexure displacement analysis was as described (52). The net DNA bending average angle was calculated.
C-box variant spacers O_L = CAT was associated with elevated restriction activity. For this purpose, we determined the efficiency of \( \lambda \)vir phage plaquing (EOP) to estimate relative restriction. We measured restriction relative to a nonrestricting reference strain and to the WT PvuII R-M system. The data in Table 2 show comparable relative restriction for cells with WT (pPvuRM3.4) or WT-Sym (pLM20) PvuII R-M systems, despite their 3-fold difference in EOT values (Table 2). Similarly, EOT values for WT-Sym (pLM20) and its nonrepressing variant of O_R (pLM6) were very close, despite relative restriction being more than 5-fold higher for the nonrepressing variant. In vivo restriction is a complex process, and its relationship to the number of target restriction sites and the relative amounts of REase and MTase have not been shown to have a predictable linear nature. However, the fact that restriction by the poorly established variants on plasmids pLM17 and pLM20 is marginally less than in the WT PvuII R-M system (Figure S3) does not support the model that their reduced establishment is due to increased restriction.

**Table 2. Relative restriction activity of C-box spacer variants estimated from efficiency of plaquing (EOP) of \( \lambda \)vir phage**

<table>
<thead>
<tr>
<th>C box spacer (plasmid name)</th>
<th>PFU( ^a )</th>
<th>Restriction( ^b ) relative to R( ^- )M( ^+ )</th>
<th>Restriction( ^c ) relative to WT R( ^- )M( ^+ )</th>
</tr>
</thead>
<tbody>
<tr>
<td>no PvuII R-M (pBR322)</td>
<td>(3.2 ± 0.8) ( \times 10^5 )</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>CAT/CAA, WT</td>
<td>(8.8 ± 0.2) ( \times 10^5 )</td>
<td>3.6 ( \times 10^5 )</td>
<td>1.00</td>
</tr>
<tr>
<td>(pPvuRM3.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAT/CAA, Sym (pLM20)</td>
<td>(1.0 ± 0.2) ( \times 10^5 )</td>
<td>3.1 ( \times 10^5 )</td>
<td>0.86</td>
</tr>
<tr>
<td>CAT/CAA, Sym( ^d ) (pLM6)</td>
<td>(2.0 ± 1.0) ( \times 10^5 )</td>
<td>1.6 ( \times 10^6 )</td>
<td>4.40</td>
</tr>
<tr>
<td>CAT/CAT, Sym (pLM17)</td>
<td>(7.4 ± 1.1) ( \times 10^4 )</td>
<td>4.3 ( \times 10^4 )</td>
<td>0.12</td>
</tr>
</tbody>
</table>

\( ^a \)PFU (plaque-forming units).

\( ^b \)Calculated as PFU obtained for control, nonrestricting strain [E. coli TOP10 (pBR322)] divided by PFU obtained for tested strain.

\( ^c \)Normalized to PFU for WT R-M system.

\( ^d \)Nonrepressing mutant.

Effects of O_L spacer on kinetics of C.PvuII-activated in vivo gene expression

The drastically decreased efficiency of transformation for C-box variants having a CAA spacer in O_L might be due to a disturbed temporal control of R-M genes by activation at lower C.PvuII concentration, with premature REase expression, in comparison to O_L = CAT.

We first determined whether the kinetics of C protein translation initiation (30) differs when the O_L spacer is CAA versus CAT. We placed C box variants in front of a promoterless cat (chloramphenicol acetyltransferase) reporter gene. CAT production can be easily measured by a sensitive colorimetric assay based on ELISA sandwich method (see ‘Materials and methods’ section). The pvuIIC gene was introduced into these cells via the WT PvuII R-M system present on M13 bacteriophage (Figure 6A), in a fairly synchronized manner (30). We used three spacer variants tested previously by in vivo C.PvuII titration (Figure 3), all in the context of symmetrized C-boxes: WT (CAT/CAA; pDK178), CAA/CAA (pLM22) and nonrepressing (box 2B reversed, pWWWR) as a control. The first two variants have identical, symmetrical GACTnnnAGTC cores and differ only by a single nt in the intra-operator O_R spacer. A control plasmid without C-box sequence (pKK232-8) was also tested. Host cell cultures showed comparable growth (Figure 6B). The cultures were infected with recombinant M13pvuIIwt at an MOI of 15 in duplicate. Postinfection CAT production was measured at different times up to 100 min (Figures 6C and 6D).

Cells infected with M13pvuII carried WT C-box operator spacers (CAT/CAA, symmetrical or nonrepressing variant) showed similar exponential normalized CAT production rate until about 50 min postinfection, after which the nonrepressing variant produced CAT at a higher rate (Figure 6C). In contrast, in the O_L = CAA variant, CAT production rose until about 50 min postinfection, more rapidly than O_L = CAT, but subsequent accumulation slowed, eventually yielding the lowest levels among tested variants (Figure 6C). The different activation profile for O_L = CAA thus appears to be associated, at least in part with temporal misregulation.

**Translation initiation of pvuIIM is not affected by tested pvuIIC C-box operator variants**

A second possible explanation for the strong effects of O_L = CAA spacer variants is that the coding sequence of MTase gene on the opposite strand is also changed. PvuII MTase has four ATG codons near its N terminus (Figure 7), so we began by determining the start of translation. The purified, overexpressed and active M.PvuII is made in two forms (38.3 and 36.9 kDa) due to alternative translation initiators that would correspond to difference in size about 13 amino acids beginning at the first and fourth methionines ([3,22]; Figure 7). The M.PvuII shorter form is about 20-fold more abundant than the larger form (9). The spacer in O_L = CAT (WT) overlaps the third methionine codon (59), and mutation to CAA generates a Met→Leu MTase change (UUG on MTase strand; Figure 7).

We generated translational fusions in which lacZ followed 26 PvuII MTase codons. Variants with AUG (M) sites replaced by UUG (L) were generated to estimate the contribution of each methionine codon as a translation initiation start. Our data indicate that the second AUG is indispensable for translation (Figure 7). The M→L change opposite the O_L spacer did not detectibly affect pvuIIM translation.

**Effect of M→L substitution on M.PvuII MTase activity**

We next tested whether PvuII MTase kinetics or activity is changed by the M→L substitution at the third methionine, opposite the O_L spacer. WT \( \lambda \)vir \( \text{pvuIIM} \) and its variants were cloned downstream of the pBAD promoter, so its expression is modulated by arabinose. The cells also contained a second plasmid bearing seven PvuII sites. A rise in MTase expression would result in greater number of methylated PvuII sites within the second plasmid, detectible by digestion with cognate PvuII REase. For these experiments we also tested M.PvuII substitution
associated with the C-box symmetrization. Tested plasmids with M.PvuII variants, were: pBadMT2 (N4S/S10R, symmetrized C-box) and pBadMT3 (N4S/M8L/S10R, symmetrized and with CAA sequence in O_L spacer). The degree of DNA protection under the same conditions of increasing MTase expression levels was compared for the three variants (Figure 8). The DNA was clearly protected by lower levels of WT M.PvuII than for the two other MTase variants (compare lane 7 for each MTase, Figure 8). However, the somewhat lower
establishment of the PvuII R-M system with symmetrized C-boxes (Table 1) might be associated with reduced protection by M.PvuII (N4S/S10R). No difference in protection is obvious between variants N4S/S10R and N4S/M8L/S10R. This result suggests that the O_L spacer change has no detectible effect on MTase activity in the context of the symmetrized C-box sequence.

**DISCUSSION**

What does C.PvuII recognize?

C-boxes are binding sites that are believed to be recognized by the HTH motifs of C proteins from R-M systems. C-boxes of the PvuII family (28) have two GACT(N3)AGTC (consensus) palindromes, separated by a 4-bp inter-operator spacer (Figure 1C). The consensus is recognized with high cooperativity by C.PvuII (28) and C.AhdI (29,42). C.PvuII has greater binding affinity for the symmetrical C-box consensus than to the WT C-boxes (Figure 3A), as also seen with symmetrical consensus operators for *lac* repressor (60) and phage λ Cro repressor (61).

A recent co-crystal structure of C.Esp1396I with its C-box DNA (41) suggested recognition of an alternative symmetry element within the C-box region (Figure 2A). Both symmetries, abbreviated ‘AGTC’ and ‘TATA’ (the latter spanning the intra-operator sequence), are well conserved in C.PvuII family C-boxes, though with slightly less conservation of the ‘TATA’ match within O_R than within O_L (Figure 1C). The C.Esp1396I structure reveals a center of symmetry shifted by 0.5 nt relative to the ‘AGTC’ consensus, with the center of the ‘TATA’ pattern at the ‘T’ of the CGTG central spacer (41). The ‘TATA’ recognition seems to involve indirect readout, with specific contacts limited to the C-box flanks (conserved TG/CA at outer edges in Figure 1C) and within the inter-operator spacer (41).

To assess the role of the ‘TATA’ symmetry, we studied the impacts of C-box variants on C.PvuII with TATA→CGCA mutations in both O_L and O_R. Tetramer complexes still formed in vitro, though at higher concentrations of C.PvuII (Figure 2B). More importantly, *in vivo* assays showed that transcriptional activation and repression still occurred (Figure 2C), though at lower peak expression levels than in the WT-Sym C.PvuII variant. Collectively, this evidence suggests that the ‘AGTC’ consensus is more important than the ‘TATA’ symmetry for binding, at least for C.PvuII.

The C.Esp1396I structure may not be fully representative, in that the C box is very unusual in having identical intra-operator spacers in O_R and O_L that both perfectly match the ‘TATA’ symmetrical consensus (there are also two hypothetical systems in the PvuII family with this
property; Figure S1). Remarkably, the majority of each HTH recognition helix in the C.Esp1396I structure makes no DNA contacts. It will be interesting to see if co-crystal structures for C-boxes lacking this perfect ‘TATA’ match show the same positioning. While C.PvuII and C.Esp1396I may simply differ in their DNA recognition, it is also possible that the tetrameric repression complex solved in (41) has substantially different DNA–protein interactions than a single dimer bound to O_L, with the latter perhaps relying on HTH amino acid contacts to the ‘AGCT’ consensus.

Complexity of the C.PvuII C-box region

In the PvuII R-M system, several regulatory elements overlap one another (Figure 1A and B). The PvuII C-boxes are located amid two overlapping promoters for pvuIICR transcription: a weak C-independent promoter, and a strong C-activated promoter. The pvuIM and pvuIICR promoters do not overlap one another, and C.PvuII has little if any effect on pvuIM transcription (30). In addition, the translation initiation area for pvuIM is on the opposite strand (arrows in Figure 1A) (25,37), with the codons for the MTase N-terminus overlapping the C.PvuII-binding region. Generating symmetrical consensus C-boxes resulted in two single-aa changes to the MTase.

We found no effect of our C-box alterations on MTase translation initiation (Figure 7). Coupled in vitro transcription-translation previously indicated that M.PvuII is made in two forms, corresponding to translation starts from the first and fourth AUG, and that the shorter form is about 20-fold more abundant than the larger form (9). Interestingly, substitution of the fourth AUG codon reduced translation by half using lacZ translational fusions. The second Met codon was absolutely required for pvuIM translation initiation. The translation initiation region of pvuIM has very limited similarity to canonical ribosome binding sites (62), regardless of which translation initiator is analyzed (Figure 7). However the initiating ATG codon is followed by an AC sequence that can strongly enhance translation (63). Importantly, the third Met codon (changed in O_L intra-operator spacer mutants) is not at all involved in translation initiation (Figure 7).

MTase activity was tested by in vivo titration, which revealed slightly reduced activity for the variant having two substitutions (N4S/S10R; due to the symmetrized C-box) (Figure 8). This may contribute to less-efficient establishment of relevant PvuII R-M system variants (Table 1).

Interestingly, these results suggest that, if both pvuIM initiators are in fact used, the longer form of the MTase plays an important role in PvuII R-M system establishment (EOT for all symmetrized variants decreased to 30%, Table 1), but not necessarily in maintenance (p1M20, Table 2). Furthermore, clones producing only the shorter form of M.PvuII are unstable (3). During R-M system introduction into a new host cell, MTase is expressed first, and endonuclease expression rises after a delay [about 8 min for PvuII, (30)]. One possibility is that the longer MTase form is produced first and provides greater initial protection activity, while later the shorter, more abundant MTase form acts as a maintenance MTase. This possibility deserves to be tested.

Higher relative affinity of C.PvuII for O_L

Even when the operator sequences themselves (GACTnnnAGTC ‘and’ their spacers are identical, there is still preferential C.PvuII binding to O_L (Figures 3C and 4B). We used in vivo assays with four variants having identical copies of the two intra-operator spacers. The four spacer variants were TAT or CGC (Figure 2C) and CAT or CAA (Figure 3A). Normally, the level of pvuIICR transcription rises with the C.PvuII level due to autogenous activation, but repression results at higher levels of C.PvuII (28). All four variants having identical spacers retained transcription profiles showing activation and subsequent repression, indicating formation of an O_L complex prior to formation of the tetrameric O_L–O_R repression complex (Figure 2A). What explains this O_L preference?

It has been proposed that competition with RNA polymerase reduces binding of C proteins to O_R (29). While that may be a contributing factor in vivo, it cannot explain the O_L dominance in the determination of net affinity that we observe in vitro (in the absence of RNAP) (Figure 4B). By inference, we might predict that flanking sequences and/or the central inter-operator sequence have an impact on C-protein regulation. There is weak conservation in the flanking sequences, except for the presence of TG/CA at the edges (Figure 1C). The C.Esp1396I tetramer co-crystal structure clearly shows direct interaction of conserved Arg35 from the recognition helix (H/D, RTY) of subunit A with the thymine base in TG, and an equivalent contact of subunit D with the symmetry-related TG on the opposite strand (41). However, this interaction appears to be symmetrical and it is not clear how it could explain the O_L preference.

The central 4 nt (TGTA in the C.PvuII C-box) is the only clearly nonsymmetrical element. In the C.Esp1396I structure, there is a second-position contact to the inter-operator spacer in its C-box (CGTG), from the C protein Arg35 of subunit B. No equivalent contact is made on the opposite DNA face via subunit C (41). Again, however, this structure is of the tetrameric repression complex, and may not reflect the structure of the initial dimer–O_L complex. The basis for O_L preference remains unclear.

Modulatory role of the intra-operator spacers

The intra-operator spacers cannot by themselves explain the O_L preference, but they may contribute to it. The analysis of 21 C-box regions from the C.PvuII subfamily (28) reveals that <10% of them have the same spacer sequence within both operators (Figure S1). We investigated the role of the spacers naturally occurring in the PvuII C-boxes: CAT in O_R and CAA in O_L, in the fully symmetrical C-box sequence context. We find that a single third position change in the O_L spacer (CAT → CAA)
results in a ~30% rise in net C.PvuII affinity, regardless of the O_R spacer sequence (Figure 4B).

Uncontacted DNA bases can still affect protein affinity. For example, operators O_R1 and O_R3 for repressor 434 differ by 3 nt, where two of them are within an uncontacted spacer yet substantially affect relative repressor affinity (64). Detailed studies of 434 and P22 c2 repressors reveal that uncontacted bases at the center of binding sites are overwound, and the minor groove is substantially compressed (65–67). The TAT intra-operator spacers of C.EspI396I are also associated with a much-narrowed minor groove (41). Compressibility of the 434 operator minor groove depends on the absence of purine N2-NH2 groups (68). This might explain the rarity of guanosine in naturally occurring intra-operator spacers of C boxes [Figure 1C, (28)] and much weaker binding of C.PvuII to the variant CGC intra-operator spacers (Figure 2B).

The balance of operator affinities appears to affect temporal control of the R-M system. We observed viability problems for cells carrying a plasmid with CAA in O_L, as indicated by strongly reduced plasmid transformability (Table 1). We propose that this defect results from altered C.PvuII binding to its C-boxes and premature REase expression. Temporal control is critical during R-M system establishment in a new host cell (30). We tested establishment of WT (O_L spacer = CAT) and its variant (O_L spacer = CAA) in new host cells (Figure 6). C protein appears to prematurely activate expression of a reporter gene downstream of the CAA O_L variant relative to WT DNA. In a full R-M system, this could have lethal effects on the host cell.

Sensitivity to changes in spacer sequences may be a powerful evolutionary tool for fine-tuning regulatory circuits, because it expands the ‘tunable’ sequence beyond just directly contacted sites. Further, the effects of sequence changes can be transmitted over 20 bp through a DNA-mediated allosteric ‘domino effect’ (69). The 434 phage repressor, which governs the lysogeny versus lysis decision, is sensitive to nucleotide variation in noncontacted spacers in its three operators. This repressor and others appear to detect DNA elastic potentials, so as to distinguish among operators without altering specific contacts (64,70–72). For mobile, temporally regulated R-M systems, natural variation in such fine-tuning may be especially important for adaptation to a particular host’s requirements, in controlling the potentially lethal endonuclease during R-M system establishment.

NOTE ADDED IN PROOF

After this manuscript was accepted, Sorokin et al. reported a detailed analysis of C boxes (73). Their study complements this one; the symmetry shown in Figure 1B corresponds to their “Motif 2”.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Isabel Novella and Tadeusz Kaczorowski for critically reviewing the manuscript, Sadik Khuder for advice on statistical analyses and Joseph Pasquinella for excellent technical assistance.

FUNDING

The National Science Foundation (Grant No. 0516692). Funding for open access charge: the National Science Foundation under Grant No. 0516692.

Conflict of interest statement. None declared.

REFERENCES