6396–6405 *Nucleic Acids Research, 2008, Vol. 36, No. 20* *Published online 5 October 2008*
doi:10.1093/nar/gkn639

Regional mutagenesis of the gene encoding the phage Mu late gene activator C identifies two separate regions important for DNA binding

Yide Jiang and Martha M. Howe*

Department of Molecular Sciences, University of Tennessee Health Science Center, Memphis, TN 38163, USA

Received June 29, 2008; Revised August 29, 2008; Accepted September 16, 2008

ABSTRACT

Lytic development of bacteriophage Mu is controlled by a regulatory cascade and involves three phases of transcription: early, middle and late. Late transcription requires the host RNA polymerase holoenzyme and a 16.5-kDa Mu-encoded activator protein C. Consistent with these requirements, the four late promoters P_{lys}, P_I, P_P and P_{mom} have recognizable −10 hexamers but lack typical −35 hexamers. The C protein binds to a 16-bp imperfect dyad-symmetrical sequence element centered at −43.5 and overlapping the −35 region. Based on the crystal structure of the closely related Mor protein, the activator of Mu middle transcription, we predict that two regions of C are involved in DNA binding: a helix-turn-helix region and a β-strand region linking the dimerization and helix-turn-helix domains. To test this hypothesis, we carried out mutagenesis of the corresponding regions of the *C* gene by degenerate oligonucleotide-directed PCR and screened the resulting mutants for their ability to activate a P_{lys}-*galK* fusion. Analysis of the mutant proteins by gel mobility shift, β-galactosidase and polyacrylamide gel electrophoresis assays identified a number of amino acid residues important for C DNA binding in both regions.

INTRODUCTION

Bacteriophage Mu is a temperate phage capable of growth on many enteric bacteria, including *Escherichia coli* K-12 (*E. coli*). Upon infection, Mu DNA inserts almost randomly into host DNA and then enters either a lysogenic or lytic pathway (1–3). Expression of the Mu c repressor protein is required for maintenance of the repressed prophage in the lysogenic state. In the absence of repression or after heat induction of a *c*ts mutant prophage, phage development proceeds via a lytic pathway which is controlled by a regulatory cascade involving three phases of rightward transcription: early, middle and late (4–6). Mu transcription is catalyzed by the host RNA polymerase (RNAP) which is required throughout the lytic cycle for the production of phage particles (7). Early transcription initiates from P_e and requires neither *de novo* protein synthesis nor Mu DNA replication (6,8). Middle transcription initiates from the middle promoter P_m and requires both DNA replication and Mor, an activator protein encoded by the Mu early transcript (5,6,9). Late transcription requires the Mu C protein encoded by the middle transcript; C activates transcription from the four late promoters: P_{lys}, P_I, P_P and P_{mom} (10–15).

Consistent with their need for an activator, the Mu late promoters P_{lys}, P_I and P_P contain −10 hexamers but lack recognizable −35 hexamers (12) (Figure 1A). In P_{mom} the sequence ACCACA is proposed to serve as a −35 element, making the spacing between the −10 and −35 elements 19 bp instead of the typical 17 bp (16–18). DNAse I footprinting of P_{lys} and P_{mom} showed that binding of C protein protects the region from about −30 to about −55 (15,18,19). Deletion mapping of plasmid-borne P_{lys} showed that P_{lys} bases −60 to +8 (relative to the transcription start site, +1) are sufficient for C-dependent transactivation (20). Analysis of an extensive collection of single base-substitution mutations in P_{lys} indicated that, in addition to the −10 region, a 19-bp region from −52 to −34 containing the C footprint is required for normal levels of C-dependent promoter activity (20). Similar, more limited, analysis of the P_{mom} promoter demonstrated the importance of this region in P_{mom} function as well (21). In both promoters this region contains an imperfect dyad-symmetrical sequence element which is proposed to serve as the binding site for a C dimer (15,18,20,21). Consistent with this hypothesis, C was found to form a dimer in solution in chemical crosslinking and gel-shift experiments (19,22).

Work on the interaction of C with P_{mom} (23) and Mor with P_m (24,25) suggests that there are both similarities

*To whom correspondence should be addressed. Tel: +1 901 448 8215; Fax: +1 901 448 8462; Email: mhowe@utmem.edu
Present address:
Yide Jiang, Oncology, Genzyme Corporation, 5 Mountain Road, Framingham MA 01701-9322

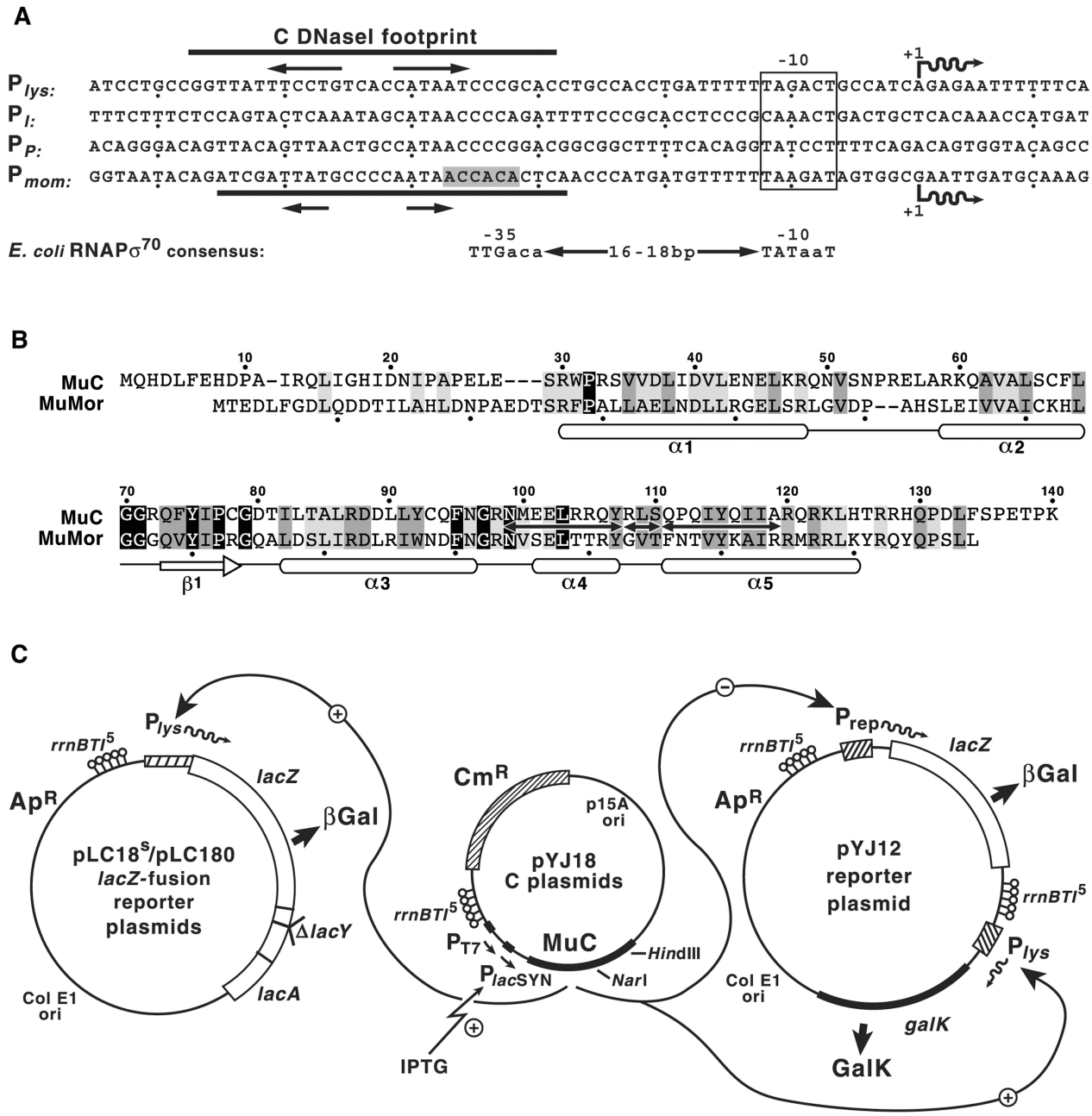

Figure 1. (**A**) Mu late promoter sequences. The sequences of the four Mu late promoters are shown aligned by their −10 hexamers (boxed). Bars show the bases protected by C protein in DNAse I footprinting experiments with P_{lys} and P_{mom} (19,56). Inverted arrows indicate the proposed dyad-symmetry elements required for C binding as determined from sequence inspection and properties of promoter mutants (20,21). The squiggly arrow represents the start (+1) and direction of transcription (rightward). The shaded hexamer in P_{mom} identifies the proposed −35 element. Below the sequences are the consensus −10 and −35 hexamers with the spacing (17 bp) for a typical *E. coli* promoter (16,17); capitalized positions are the most important for promoter strength (16,17). (**B**) Amino acid sequence alignment of Mor and C proteins. The alignment shown was derived from a 15-member family alignment (29) with black and dark gray indicating invariant and chemically conserved residues, respectively, in that alignment. Light gray shading shows additional residues that are identical or chemically conserved between C and Mor (shading has been left out from positions in which residues are conserved in the 15-member alignment, but not in C and Mor). The two-headed arrows between the sequences show the amino acids corresponding to the typical 20 amino acid HTH DNA binding motif (32). (**C**) Three plasmids used in this work (not drawn to scale). Plasmid pYJ18 is the *C* expression plasmid with P_{T7} and P_{lac}SYN driving transcription of the wild-type (pYJ18) and mutant *C* genes. The negative control plasmid pYJ38 was made by deleting the *C* DNA segment between the NarI and HindIII sites. Plasmid pYJ12 is the reporter plasmid used for screening C mutant properties; it expresses *galK* under P_{lys} control and *lacZ* under P_{rep} control. Finally, pLC18^S and pLC180 are the reporter plasmids with *lacZ* transcribed under P_{lys} control with promoter sequences −52 to +8 for pLC18^S and −60 to +8 for pLC180; they were used for quantitative β-galactosidase assays reflecting the ability of the mutant C proteins to activate P_{lys}.

and differences in their activation mechanisms. Binding of each protein to a proposed dyad-symmetrical sequence element is critical for activation as are the immediately adjacent downstream bases (20,21,25,26). Binding of C caused modest DNA bending (~20°–40°), and each protein generated hypersensitive sites in footprinting assays carried out with and without RNAP (19,21,26). In the case of Mor-dependent activation of P_m, the C-terminal regions of both the alpha (α) and sigma (σ) subunits of RNAP are needed for optimal activation (24); whereas neither is required for C-dependent activation of P_{mom} (23). The C protein plays at least two roles in activation of P_{mom}; C binding to P_{mom} leads to recruitment of RNAP and also facilitates promoter escape (27).

The Mu C protein is a small protein of 140 amino acids (16.5 kDa) which shows considerable sequence similarity to the Mu Mor protein, the activator of middle transcription (9). Previous BLASTP (28) analysis identified 13 additional Mor/C homologs, predominantly in prophage sequences in bacterial genomes, making Mor and C the founding members of a new family of transcription factors (29). Recent BLASTP analyses identified more than 40 family members (data not shown). These proteins are small (~100–150 amino acids) and exhibit a preponderance of acidic amino acids in the N-terminal half, basic amino acids in the C-terminal half, and a predicted helix-turn-helix (HTH) DNA-binding motif near the C-terminus as shown in Figure 1B (9,11,13,30–32).

The crystal structure of Mor protein for Mor amino acids 27 to 120 was determined to 2.2 Å resolution (29). The structure (Figure 2A) contains a Mor dimer with a single dimerization domain formed by the intertwining of helices α1 and α2 of both monomers to form a 4-helix bundle. Flanking that central domain are the two HTH domains, one from each monomer. In the linker between the two domains of each monomer there is a β-strand which interacts in an anti-parallel fashion with the β-strand of the other monomer. In this structure, the side chains of conserved β-strand residues Q68 and Y70 of Mor extend away from the protein (Figure 2B), and those of the hydrophobic residues V69, I71 and P72 point toward the hydrophobic interior, forming a cap on the dimerization domain. When the structural coordinates of the HTH domain of Mor were compared with those of other proteins, the TrpR protein HTH region was the most similar (r.m.s.d. 1.5), leading to the prediction that Mor, like TrpR, would use 'ends on' base recognition binding by residues in the central turn of the HTH motif and in the N-terminal region of the following recognition helix (Mor α5) to interact in the DNA major grooves. Since those HTH regions were too far apart to contact two adjacent major grooves, Kumaraswami *et al.* (29) proposed that the protein would undergo a conformational change in which the HTH domains would be rotated up away from the dimerization domain and closer together to contact bases in the DNA major groove (Figure 2D). That conformational change could bring the β-strands close to the DNA minor groove located between the two major grooves, potentially allowing interaction of side chains Q68 and Y70 with the minor groove and intercalation of one or more nearby hydrophobic residues into the minor groove.

To test whether the predicted β-strand and HTH regions of C are important for its DNA binding we used degenerate oligonucleotide mutagenesis of both the β-strand and HTH regions, screened for activation-deficient mutants, and then characterized the properties of the mutant proteins *in vivo* and *in vitro*. The results support the prediction that the regions containing both the predicted HTH motif and β-strand are important for DNA binding by C.

MATERIALS AND METHODS

Media, chemicals and enzymes

Minimal medium was M9 (33) supplemented with vitamin B1 and a carbon source, as well as amino acids and antibiotics when needed; minimal plates contained in addition 1.5% Bacto-agar (Difco laboratories, Detroit, MI, USA). Minimal medium supplemented with 0.2% casamino acids (M9CA) was used for β-galactosidase assays (33) and protein over-expression. Liquid Luria broth (LB) and LB plates (5) were used for routine cell growth purposes. MacConkey lactose or galactose plates containing only 0.5% sugar (half the normal amount) were made with 40 g Difco MacConkey Agar Base and 5 g lactose or galactose per liter.

Ampicillin (Ap; U.S. Biochemical Corp., Cleveland OH, USA) and chloramphenicol (Cm; Sigma Chemical Co., St. Louis, MO, USA) were used as necessary at 50 and 25 μg/ml, respectively, unless indicated otherwise. Isopropyl-β-D-thiogalactopyranoside (IPTG) and *o*-nitrophenyl-β-D-galactopyranoside were from American Bioorganics Inc., Niagara Falls, NY, USA. Radiolabeled compounds were purchased from DuPont NEN, Boston, MA, USA. Acrylamide, bisacrylamide, N,N, N′,N′-tetramethylethylenediamine and protein molecular mass markers were from BioRad, Hercules, CA, USA. Ammonium sulfate, polyethyleneimine (PEI), phenylmethylsulfonyl fluoride (PMSF), NP-40 and 2-deoxygalactose were from Sigma Chemical Co. Both SeaKem ME and NuSieve GTG agarose were from FMC Bioproducts, Rockland, ME, USA.

Shrimp alkaline phosphatase, dNTPs, Sequenase 2.0, labeling and termination mixes were from U.S. Biochemical Corp (USB). The enzymes EcoRI, BamHI and Taq polymerase were from Boehringer Mannheim Biochemicals, Indianapolis IN, USA; T4 polynucleotide kinase was from Promega Corporation, Madison, WI, USA; other restriction enzymes were from New England BioLabs, Beverley, MA, USA. All enzymes were used according to the manufacturer's recommendations.

Bacterial strains and strain construction

The bacterial strains used are all derivatives of *E. coli* K12 strain JM109 (*mcrA* Δ*proAB-lac thi gyrA endA hsdR relA supE44 recA*/*F′ traD36 lacI*^{Q} *lacZΔM15 proAB*$^{+}$). Strain MH13312, containing a different F′ factor, F′ *pro*$^{+}$ *lacI*^{Q1} *ΔlacZY*, was the host strain used for *in vivo* β-galactosidase assays following introduction of the promoter and

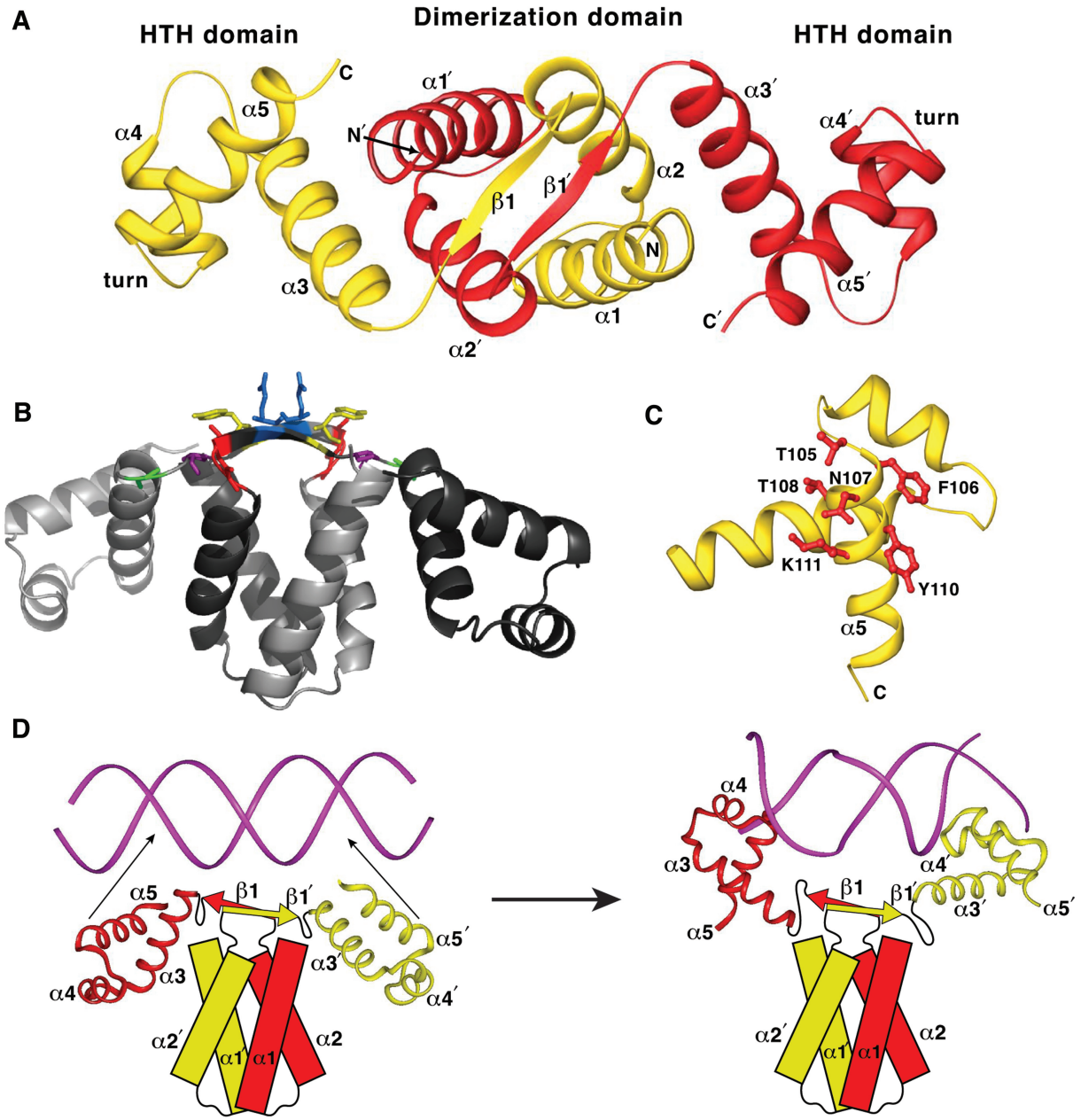

Figure 2. (**A**) The Mor dimer structure (amino acids 27–120) viewed from above with one monomer red and the other yellow. The secondary structure elements, α helices 1–5 and β-strand 1 are numbered from amino acid 27 through amino acid 120; α and β numbers with a prime are in the second monomer. (**B**) The Mor dimer structure viewed from the side. Side chains of key amino acids are shown in a 'ball and stick' representation with 65G and 66G shown in red, 68Q in blue, 70Y in yellow, 72P in magenta and 74G in green. (**C**) The structure of the HTH domain of Mor protein with side chains shown for amino acids predicted to be involved in DNA binding. (**D**) Conformational changes predicted to occur in both Mor and promoter DNA for stable DNA binding. The helices α-1 and α-2 of the Mor dimerization domain are shown as cylinders; the β-strands as arrows, and the HTH domains as ribbons. Thin black arrows point from the vicinity of the predicted DNA binding amino acids to the two adjacent major grooves typically contacted by HTH motifs (48–50). The DNA (in purple) shows the predicted curvature away from the protein upon Mor binding. Images in A, C and D are used with permission from the Mor structure paper by Kumaraswami *et al.* (29). (**E**) Mutagenesis of the predicted DNA-binding regions of C. The heavy solid line represents the 140 amino acids of C protein, with dots at 10 amino acid intervals. Upside-down triangles indicate the positions of unique pre-existing restriction sites (in parentheses below) and new sites introduced to facilitate cassette mutagenesis; from the left end and proceeding rightward these triangles indicate restriction sites NdeI, MunI, ClaI, XhoI, SalI, AflII, SacI, BpuI102I, NarI, DraIII, (BstXI), BanII, (BglII), BsshII, HindIII and (AatII). The secondary structure elements, α helix (oval) and β-strand (arrow), from the crystal structure of Mor are shown at their corresponding positions in C as shown in Figure 1B; the dashed oval represents an additional α helix in C that was predicted by the protein structure prediction program PHD (57). The HTH near the C-terminus indicates the positions of the helix-turn-helix DNA-binding motif as predicted by the method of Dodd and Egan (30,31). Brackets with Roman numerals at the top of the figure show the regions mutagenized in the three different cassettes. The arrows and numbers below the solid line identify the amino acid positions with single amino acid changes; the letters indicate the wild-type amino acid and mutant amino acid at that position, respectively, using the single letter amino acid code.

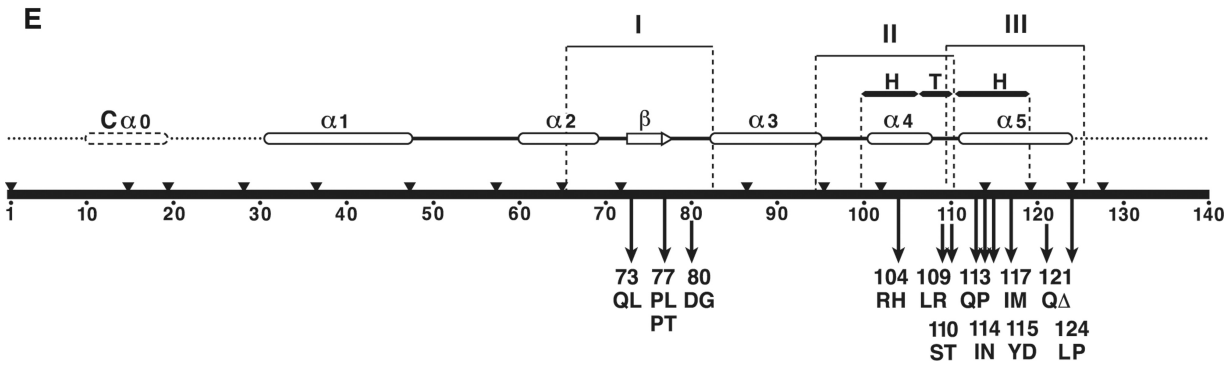

Figure 2. Continued.

protein expression plasmids (26). Strain MH12802 is JM109 containing the pLC3 plasmid, which encodes C protein under the control of a P_{lacUV5} promoter (20). Strains MH13708 and MH14607 were made by transformation of MH13312 with plasmids $pLC18^S$ or pLC180, respectively. Strain MH13355 is a derivative of JM109(DE3), containing a different F′ factor, F′ *pro*$^+$ *lacI*^{Q1} *ΔlacZY* and the lambda DE3 prophage encoding phage T7 RNAP under P_{lacUV5} control (26). A spontaneous *galK* mutant derivative of MH13355, designated MH13881, was selected by its resistance to 2-deoxygalactose (34) by plating on minimal plates containing 0.2% each of 2-deoxygalactose and glycerol and then screening of resistant mutant colonies on MacConkey galactose plates for a Gal$^-$ mutant (white colony) that became Gal$^+$ (red colony) upon transformation with plasmid pYJ12 containing a wild-type *galK* gene. The derivative of MH13881 carrying pYJ12, MH13906, was used in assays of promoter activation by mutant C proteins.

Plasmids and plasmid construction

Plasmid pLC1 is a *lacY* derivative of the promoterless *lacZ* fusion vector pRS415 (35) generated by deletion of the SnaBI fragment in *lacY* (20). Cloning of a promoter into the EcoRI-SmaI-BamHI linker in pLC1 or pRS415 generates a promoter-*lacZ* fusion suitable for analysis of promoter activity. Plasmid vector pIA12 is a derivative of pLC1 with an additional unique HindIII site located just upstream of the EcoRI site (26). Plasmid pLC180 (Figure 1C) is a new derivative of pLC1 with the *lacZ* gene under the control of the Mu P_{lys} promoter (−60 to +8). Plasmid $pLC18^S$ (made previously and designed to be identical to pLC180) was recently found to contain only P_{lys} sequences −52 to +8. Comparison of the two plasmids showed only minor differences:

(i) induction of wild-type C *in vivo* resulted in only slightly (~23%) higher promoter activity from pLC180 (9506 ± 604) than from $pLC18^S$ (7705 ± 821), and
(ii) gel shift assays with both promoter sequences using crude extract and purified C protein gave shifted bands of similar intensity (19).

The transactivator plasmid pZZ13, derived from pACYC184 (36,37), has the Mu *C* gene under the control of both P_{lacUV5} and P_{T7} promoters (19). Plasmid pZZ41 is similar to pZZ13 except that a more repressible synthetic P_{lac} promoter, designated P_{lacSYN}, was substituted for P_{lacUV5} (19).

Unique silent restriction sites were introduced into the *C* gene to facilitate subsequent cassette mutagenesis. The approach involved multi-step PCR with oligonucleotides containing site-directed mutations, followed by restriction digestion and cloning into existing NdeI, BstXI and BamHI sites in or flanking the *C* gene in pZZ13. Then, the NdeI-BamHI fragment containing the entire *C* gene was cloned into pZZ41 replacing the wild-type *C* gene to generate the final *C* expression vector pYJ18 (Figure 1C). DNA sequence analysis confirmed that pYJ18 contains the desired restriction sites and encodes C protein with the wild-type C amino acid sequence. A control experiment showed that plasmids pYJ18 and pZZ41 promoted similar levels of P_{lys}-*lacZ* transactivation *in vivo* as shown by relative β-galactosidase units of 1000:861 for pYJ18:pZZ41 as assayed in strain MH13708 containing the $pLC18^S$ reporter plasmid. Oligonucleotide sequences, restriction sites and details of the construction will be provided upon request. The negative control plasmid, pYJ38, which is missing most of the *C* gene, was derived from pYJ18 by deleting the NarI-HindIII fragment from *C*.

Plasmid pYJ12 was constructed to facilitate the screening of mutant C protein activity by colony color on MacConkey plates; it contains two reporter systems: a P_{lys}-*galK* fusion to assay promoter activation by C and a P_{rep}-*lacZ* fusion to assay DNA binding by C by its ability to repress the constitutive synthetic promoter P_{rep}. First, the 1686-bp PstI-EcoRI fragment from the P_{rep}-*lacZ* plasmid pYJ3 containing five copies of the TI terminator was cloned into the promoterless *galK* transcription fusion vector pKO4 (38,39) to generate pYJ6. Then, P_{lys} was introduced between the terminator and *galK* gene by cloning the $pLC18^S$ P_{lys} EcoRI-BamHI fragment into corresponding sites of pYJ6 to generate pYJ7. The unique restriction sites BsmBI and DraII in pYJ7 were changed to unique SalI and XbaI sites by cloning an adapter. To prepare the P_{rep}-*lacZ* plasmid pYJ3 to receive the P_{lys}-*galK* fusion, an XbaI linker was cloned into the SnaBI site of pYJ3 to create pYJ8. Finally, the 2480-bp XbaI-SalI fragment from pYJ11 was cloned into pYJ8 to generate pYJ12.

Oligonucleotide synthesis

Oligonucleotides were synthesized by the University of Tennessee Molecular Resource Center on an Applied Biosystems (Foster City, CA, USA) DNA synthesizer (Model 380B) using the phosphoramidite method (40). For each degenerate oligonucleotide primer (20,41), degeneracy was introduced into a 45–47 nt targeted region and synthesis was accomplished with a mixture of 98% of the correct phosphoramidite and 2% of a 1:1:1:1 mixture of dA, dC, dG and dT phosphoramidites to yield on average 0.8 mutations per DNA strand. For site-specific mutagenesis the desired nucleotides were used for synthesis at the appropriate positions. When necessary, restriction sites for cloning were added at the 5′-ends of primers and were preceded by three to six bases for efficient digestion. Oligonucleotides were purified by ether extraction and ethanol precipitation (20) prior to use. The sequences of oligonucleotides will be provided upon request.

Mutagenesis and mutation identification

Three regions of the *C* gene in pYJ18 were mutagenized by cloning three separate cassettes made by PCR with oligonucleotides containing degeneracy in specific targeted regions (20,41). The primers (YJ24, YJ26 and YJ27 for regions I, II and III, respectively) were used with convenient wild-type opposing primers in separate PCR reactions, and the products were cloned into pYJ18 by using restriction enzymes whose sites flanked the cassette. Strain MH13906 containing the P_{lys}-*galK*, P_{rep}-*lacZ* reporter plasmid (pYJ12) was transformed with each ligation mixture, and transformants were selected by plating on MacConkey galactose indicator plates containing ampicillin and chloramphenicol but lacking IPTG. White Gal^- colonies containing *C* plasmids defective for P_{lys} activation were then assayed for C-dependent repression of the constitutive promoter P_{rep} by streaking on MacConkey lactose (with only 0.5% lactose) plates with and without 10^{-2} mM IPTG to induce moderate levels of C production. By design, colonies that were red on both plates (with and without IPTG) were to contain plasmids with candidate DNA binding-defective mutations in *C*; in practice this repression assay was found to be unreliable, so the next assay used for mutant characterization was SDS-PAGE (42) to determine the size of the over-expressed mutant C protein, and examination of the mutants for DNA binding was deferred to a later step.

Plasmid templates for sequencing were isolated using a QIAprep spin purification kit (Qiagen, Valencia, CA, USA) or a standard alkaline denaturation DNA miniprep procedure (43). Templates were denatured in 0.2 M NaOH for 10 min at 65°C, neutralized, ethanol precipitated and sequenced with Sequenase (USB) and dideoxynucleotides using standard USB protocols. The mixtures were run on 6% polyacrylamide sequencing gels, which were fixed, dried and exposed to X-OMAT AR film (Kodak, Rochester NY, USA) to visualize the band pattern.

Characterization of mutant proteins

Protein production and detection. Derivatives of strain MH13881 carrying pYJ18 or its *C*-mutant forms were grown to $\sim 2 \times 10^8$ cells/ml at 37°C in LB medium and induced with 1 mM IPTG for 90 min. Cells from 0.5 ml of culture were collected by centrifugation (9000*g* for 10 min), resuspended in 100 µl buffer H (25 mM HEPES, pH 7.5 at room temperature; 0.5 mM EDTA, 0.5 mM dithiothreitol, 0.5 mM $MgCl_2$, 1 mM $CaCl_2$, 50 mM NaCl, 5% glycerol, 0.1% NP40) and lyzed by boiling in 125 µl of sample buffer (42). Then, 40-µl portions were subjected to electrophoresis on 15% polyacrylamide gels containing sodium dodecyl sulfate (SDS−PAGE) and stained with Coomassie blue to detect the sizes and levels of the mutant C proteins.

For preparation of crude extracts for gel retardation assays, overnight LB cultures of the above strains were diluted 1:33 into 50 ml M9CA minimal medium supplemented with 25 µg/ml Cm and grown at 37°C to an A_{600} of 0.6. After addition of IPTG to 1 mM, the cells were grown for 90 min at 37°C, collected by centrifugation (9000 *g* for 10 min), resuspended in 5 ml buffer H and lyzed by sonication. Cell debris was removed by centrifugation at 9000*g* for 10 min, and the supernatant, designated crude extract, was distributed into aliquots, and stored at −70°C. The relative concentration of C protein in each extract was estimated by electrophoresis in 15% PAGE with SDS in Tris−glycine buffer for about 4 h, followed by staining with Coomassie blue (42), visual comparison of C protein staining intensity, and dilution of the extracts to achieve comparable amounts of C protein in each sample.

Analysis of DNA binding activity. The ability of C proteins to bind to P_{lys} was determined by a gel retardation assay performed essentially as described by Carey (44) utilizing a probe containing P_{lys} sequence −52 to +8 made by PCR from $pLC18^s$ with ^{32}P-labeled oligonucleotides LIL1 and LIL2 (20). Probes were purified by using a QIAquick spin PCR purification kit (Qiagen), and their concentrations were determined on an ethidium bromide stained agarose-mini gel (45). Binding reactions contained 10–20 ng of probe and were conducted in 40 µl of modified buffer H containing 1% NP40 and 7% glycerol for 15 min at room temperature. Reactions were loaded on 10% acrylamide gels (29:1), 0.5× TBE, and run for 4 hr at 8°C at 15 V/cm. Gels were exposed overnight to X-OMAT AR film (Kodak) without drying.

Quantitative assay for transcription activation. Plasmid pYJ18 and its *C*-mutant derivatives were transformed into strain MH13708 containing the P_{lys}-*lacZ* reporter plasmid $pLC18^s$ for P_{lys} activation assays (Figure 1C). Cells were grown in 10 ml of M9CA to A_{600} 0.3–0.5 at 37°C, induced for C production by addition of IPTG to 2 mM and grown at 37°C for 1 hr. Uninduced controls and a wild-type *C* control culture were included in each set of assays. Assays for β-galactosidase were performed as described by Miller (33) with minor modifications (20). The β-galactosidase units were calculated and normalized

relative to those of the parallel induced wild-type control (set to 1000 U) to minimize the effect of day to day variation; values over 10 days for activation of the reporter pLC18^{s} by wild-type C ranged from 6256 to 9185 Miller units with an average of 7705 ± 821.

RESULTS

Design of mutagenesis and reporter plasmids

Two plasmids were constructed, one (pYJ18) to allow efficient production of *C* mutants by cassette mutagenesis and the other (pYJ12) to facilitate detection of the DNA binding and transactivation abilities of the mutant C proteins. The cassette mutagenesis plasmid pYJ18 contains a modified *C* gene, which was altered by introduction of a number of unique 'silent' restriction sites as described in Materials and methods section (Figure 2E). These sites are called 'silent' because the C protein encoded by pYJ18 has the wild-type C amino acid sequence; pYJ18 produces slightly better than normal transactivation activity *in vivo* (see Materials and methods section). Plasmid pYJ18 carries this altered *C* gene under both P_{lacSYN} and P_{T7} control. Induction of P_{lacSYN} with IPTG leads to production of relatively physiological levels of C protein for assays of activation *in vivo*. In a host containing λ DE3, which carries the T7 RNA polymerase gene under P_{lacUV5} control, IPTG induction leads to substantial overproduction of C which is lethal to the host cell but suitable for preparation of crude extracts for gel-shift analysis, estimation of C protein size and quantity, and purification of C protein.

The second plasmid, pYJ12, carries two reporter systems: a P_{lys}-*galK* fusion to assay promoter activation by C and a P_{rep}-*lacZ* fusion to assay binding of C to DNA by its ability to repress the constitutive promoter P_{rep}. P_{rep} is a synthetic promoter which contains a perfectly symmetrical strong C binding site overlapping the −10 and +1 region of the promoter such that binding of C represses transcription. As designed, introduction of mutagenized derivatives of pYJ18 into a host strain carrying pYJ12 (MH13906) would allow for fast and convenient assays for C DNA binding and transactivation functions by the colony color on MacConkey galactose and MacConkey lactose indicator plates. Transactivation-defective mutants would give white colonies on MacConkey galactose plates; whereas, DNA binding-defective mutants would give red colonies on MacConkey lactose plates. In practice, the screening assay for transactivation worked very well, allowing detection of mutants with varying levels of transactivation by the range of colony colors on MacConkey galactose plates. In contrast, the P_{rep} repression assay was sufficiently leaky that all colonies eventually turned red, so its use was limited to early stages of mutant isolation.

Mutagenesis

The mutagenesis strategy was to target mutations specifically to the β-strand and HTH regions of the *C* gene using three degenerate oligonucleotide primers. PCR with pairs of one wild-type and one degenerate primer (20,46) was used to generate the three populations of mutant cassettes which were cloned into the *C* gene in the expression vector pYJ18, replacing the wild-type sequence in that region. The primers were synthesized under conditions predicted to result in a 1.5% mis-incorporation rate per nucleotide over the mutagenized region, producing primer populations that on average should have ∼44% with wild-type sequence, ∼36% with single mutations, ∼15% with double mutations and ∼5% with more than two mutations (46).

The mutagenized plasmid libraries were transformed into MH13906 (containing pYJ12; Figure 1C) and transformants were screened for transactivation of the P_{lys}-*galK* fusion on MacConkey galactose plates. The transformants were next assayed for the size and quantity of C protein produced by using SDS-PAGE and Coomassie staining of proteins from IPTG-induced cells lyzed by boiling in sample buffer. As shown in Figure 3A, induction with IPTG resulted in high level expression of wild-type or mutant C protein from the T7 promoter, as demonstrated by an intense band corresponding to C protein on the gel. Some mutants gave over-expressed C protein with the same migration as wild-type C, for example, Figure 3A—mutant 77. Others produced protein which appeared to be larger or smaller than wild-type C, for example, Figure 3A—mutants 148 and 254. Others showed no over-expressed protein band at all (data not shown), suggesting that the C protein made was unstable and rapidly degraded. For the mutants that produced an intense protein band with a migration similar to that of wild-type C, the *C* gene was sequenced on both strands within and beyond the entire cloned cassette. Figure 2E shows the locations of the mutations and identity of mutant proteins with single amino acid changes.

Gel retardation assay

A gel mobility shift assay was used to test the DNA-binding ability of the mutant proteins with single amino acid substitutions. Crude cell extracts containing roughly similar amounts of overproduced C protein were assayed for binding to a wild-type promoter DNA fragment containing P_{lys}^{S} (−52 to +8). All of the mutant proteins (identified in Figure 2E) were defective in DNA binding as reflected by the absence of a detectable shifted band. Figure 3B shows representative results from gel-shift assays of two such mutants. Most mutant proteins gave results similar to those for protein 110ST (Figure 3B), that is, no detectable DNA binding. Approximately one-third gave results similar to those for protein 115YD (Figure 3B), which we interpret as weak unstable binding.

Transactivation assay

A more quantitative assessment of the transactivation ability of the mutant C proteins was obtained by transformation of the mutant plasmids into *E. coli* strain MH13708 containing the P_{lys}- *lacZ* reporter plasmid pLC18^{S} and performing liquid β-galactosidase assays. All assays were done in parallel with strains containing pYJ18 (wild-type *C*) and pYJ38 (*C* gene deleted) as positive and negative controls, respectively. These results showed that all the DNA-binding defective mutants are severely defective in transactivation, producing less than 5% of wild-type levels of transactivation (data not shown).

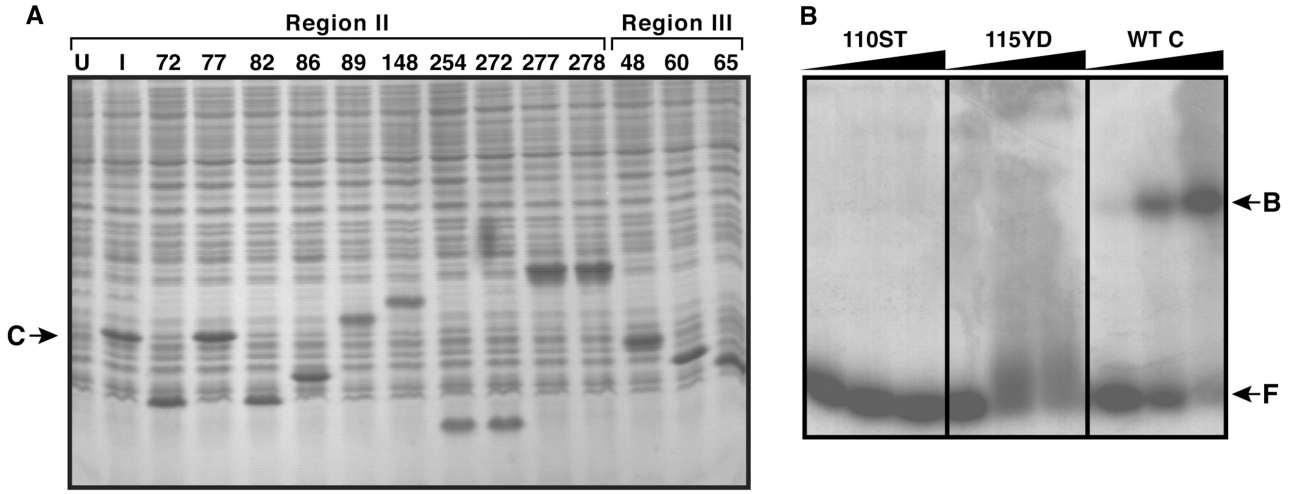

Figure 3. (**A**) Mutant protein screening by SDS-PAGE. Mutant C proteins were over-expressed by IPTG induction of T7 RNAP expression in cells containing *C* mutant derivatives (numbered above each lane) of the expression vector pYJ18. Cell extracts were subjected to electrophoresis in 15% PAGE in Tris-glycine buffer for ~4 hr and then stained with Coomassie blue R250. Lanes 'U' and 'I' contain extracts from uninduced and induced cells, respectively, making wild-type C protein. The arrow points to a C protein band of wild-type length. (**B**) Representative gel shift assay for DNA binding. Two-fold serial dilutions of crude extract containing similar amounts of mutant or wild-type C protein were incubated for 15 min at 25°C with a ^{32}P-labeled PCR fragment containing P_{lys} sequence from −52 to +8, then subjected to electrophoresis in the cold for ~3 hr. Free probe (F) and bound complexes (B) were detected by autoradiography. All the panels came from a single gel, but three lanes between 115YD and WT C were removed.

DISCUSSION

The results presented here demonstrate that the β-strand and HTH regions of C protein are important for its DNA binding and transactivation functions. Mutations in regions II and III are located within or just downstream of the predicted HTH DNA-binding motif characteristic of many transcription regulators [for reviews see (32,47–50)]. In the crystal structure of Mor, the analogous residues make up helices α4 and α5 in the three-helix bundle of the HTH domain (29). In the well characterized 20 amino acid HTH motif (30–32), amino acids at positions 4, 8, 10, 15 and 18 are usually hydrophobic, and both Mor and C have hydrophobic residues at these positions. For C these residues are 103L, 107Y, 109L, 114I and 117I. Thus the severe DNA-binding defects caused by mutations 109LR and 114IN are consistent with their predicted roles in forming the hydrophobic core of the HTH domain. In many HTH proteins, amino acids in the turn and final helix (α5 of Mor) of the HTH make important contacts within the two successive DNA major grooves, leading to its designation as the recognition helix (32). The structural similarity between the Mor HTH domain and that of TrpR led to the prediction that the Mor HTH would contact the DNA in an 'ends-on' manner, using amino acids in the turn and N-terminus of α5 for these contacts (29). The analogous C amino acids correspond to 110S–116Q; their involvement is supported by the DNA binding defects caused by mutations 110ST, 113QP, 114IN and 115YD. Taken together, these results provide strong support for the prediction that the C-terminal one-third of C contains an HTH motif which is required for DNA binding.

A number of HTH and non-HTH DNA-binding proteins use more than one motif for DNA binding, with the secondary motif often interacting in the DNA minor groove (48–50). The mutations in region I correspond to residues within or flanking the two β-strands observed in the crystal structure of the Mor dimer (Figure 2A and B; 29) and may identify such a secondary binding motif in the Mor/C family of proteins (29). In the Mor structure this region serves as a linker between the dimerization and HTH domains and contains five invariant amino acids and two highly conserved hydrophobic residues. These correspond to invariant C amino acids 70G, 71G, 75Y, 77P and 79G as well as conserved hydrophobic residues at 74F and 76I (Figure 1B). Since the N-terminal residues of Mor α5 that are predicted to bind in two successive major grooves of the DNA are too far apart to reach them (Figure 2D), Kumaraswami *et al.* (29) proposed that the conserved glycines provide pivot points for conformational changes that move the HTH domains from their original positions beside the dimerization domain to new positions above it in order for the HTH motifs to contact the DNA major grooves (Figure 2D). This change would bring the β-strand and nearby amino acids of C into close proximity with the DNA minor groove, explaining why the glycine residues are invariant and potentially allowing interaction of the 73Q and 75Y residues with the minor groove (Figure 2D). Consistent with this hypothesis, footprinting with the minor groove-specific chemical nuclease 1, 10-phenanthroline copper showed that interaction of C with its binding site in P_{mom} prevented minor groove cleavages seen in the absence of C (22). Binding of both Mor and C proteins produces a ~40° bend in the DNA (21,26) and Kumaraswami *et al.* (29) predicted that the DNA will bend away from the protein. Such bends are often generated by intercalation of one or more hydrophobic amino

acid side chains between the base pairs (51–55). The bend angle is dependent on the size and length of the amino acid side chain and the depth of its insertion into the DNA minor groove, generating bends as small as 20° and as large as 180° (51–55). The β-strand in Mor has 68Q and 70Y side chains extending out from the surface of the protein (Figure 2B) and 69V, 71I and 72P pointing into the protein, forming a cap on the hydrophobic dimerization domain (29). This cap may be retained in the DNA-bound Mor or one or more of its residues may intercalate into the minor groove. These residues correspond to C amino acids 73Q and 75Y pointing out and hydrophobic residues 74F, 76I and 77P, respectively. Taken together, these arguments explain the serious DNA binding defects caused by mutations 73QL, 77PL and 77PT.

ACKNOWLEDGEMENTS

Oligonucleotides were synthesized by the University of Tennessee Molecular Resource Center on an Applied Biosystems DNA synthesizer (Model 380B) using the phosphoramidite method (40).

FUNDING

Van Vleet Chair of Excellence in Virology and College of Medicine, University of Tennessee Health Science Center (partial); National Science Foundation Grants (MCB-9305924, MCB-9604653 and MCB-0418108). Funding for open access charge: Van Vleet Chair of Excellence in Virology.

Conflict of interest statement. None declared.

REFERENCES

1. Symonds,N.A., Toussaint,A., van de Putte,P. and Howe,M.M. (1987) *Phage Mu*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
2. Howe,M.M. (1997) Bacteriophage Mu. In Busby,S.J.W., Thomas,C.M. and Brown,N.L. (eds), *Molecular Microbiology*. Spring-Verlag, Heidelberg, pp. 65–80.
3. Paolozzi,L. and Ghelardini,P. (2006) The Bacteriophage Mu. In Calendar,R. (ed.), *The Bacteriophages*. Oxford University Press, New York, NY, pp. 469–496.
4. Goosen,N. and van de Putte,P. (1987) Regulation of transcription. In Symonds,N., Toussaint,A., van de Putte,P. and Howe,M.M. (eds), *Phage Mu*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 41–52.
5. Stoddard,S.F. and Howe,M.M. (1989) Localization and regulation of bacteriophage Mu promoters. *J. Bacteriol.*, **171**, 3440–3448.
6. Marrs,C.F. and Howe,M.M. (1990) Kinetics and regulation of transcription of bacteriophage Mu. *Virology*, **174**, 192–203.
7. Toussaint,A. and Lecocq,J.-P. (1974) Sensitivity of bacteriophage Mu-1 development to rifampicin and streptolydigin. *Mol. Gen. Genet.*, **129**, 185–188.
8. Wijffelman,C. and van de Putte,P. (1974) Transcription of bacteriophage Mu: an analysis of transcription pattern in the early stage of phage development. *Mol. Gen. Genet.*, **135**, 327–337.
9. Mathee,K. and Howe,M.M. (1990) Identification of a positive regulator of the Mu middle operon. *J. Bacteriol.*, **172**, 6641–6650.
10. Hattman,S., Ives,J., Margolin,W. and Howe,M.M. (1985) Regulation and expression of the bacterio-phage Mu *mom* gene: mapping of the transactivation (Dad) function to the C region. *Gene*, **39**, 71–76.
11. Heisig,P. and Kahmann,R. (1986) The sequence and mom-transactivation function of the C gene of bacteriophage Mu. *Gene*, **43**, 59–67.
12. Margolin,W., Rao,G. and Howe,M.M. (1989) Bacteriophage Mu late promoters: four late transcripts initiate near a conserved sequence. *J. Bacteriol.*, **171**, 2003–2018.
13. Margolin,W. and Howe,M.M. (1986) Localization and DNA sequence analysis of the *C* gene of bacteriophage Mu, the positive regulator of Mu late transcription. *Nucleic Acids Res.*, **14**, 4881–4897.
14. Margolin,W. and Howe,M.M. (1990) Activation of the bacteriophage Mu *lys* promoter by Mu C protein requires the σ70 subunit of *Escherichia coli* RNA polymerase. *J. Bacteriol.*, **172**, 1424–1429.
15. Bolker,M.F., Wulczyn,G. and Kahmann,R. (1989) Role of bacteriophage Mu C protein in activation of the *mom* gene promoter. *J. Bacteriol.*, **171**, 2019–2027.
16. Harley,C.B. and Reynolds,R.P. (1987) Analysis of *E. coli* promoter sequences. *Nucl. Acids Res.*, **15**, 2343–2361.
17. Hawley,D.K. and McClure,W.R. (1983) Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucl. Acids Res.*, **11**, 2237–2255.
18. Balke,V., Nagaraja,V., Gindlesperger,T. and Hattman,S. (1992) Functionally distinct RNA polymerase binding sites in the phage Mu mom promoter region. *Nucleic Acids Res.*, **20**, 2777–2784.
19. Zhao,Z. (1999) Effects of P_{lys} promoter sequence on transcription activation by the C protein of bacteriophage Mu. *Ph.D. Thesis*. The University of Tennessee, Memphis.
20. Chiang,L.W. and Howe,M.M. (1993) Mutational analysis of a C-dependent late promoter of bacteriophage Mu. *Genetics*, **135**, 619–629.
21. Sun,W., Hattman,S. and Kool,E. (1997) Interaction of the bacteriophage Mu transcriptional activator protein, C, with its target site in the *mom* promoter. *J. Mol. Biol.*, **273**, 765–774.
22. Ramesh,V. and Nagaraja,V. (1996) Sequence-specific DNA binding of the phage Mu C protein: footprinting analysis reveals altered DNA conformation upon protein binding. *J. Mol. Biol.*, **260**, 22–33.
23. Sun,W., Hattman,S., Fujita,N. and Ishihama,A. (1998) Activation of bacteriophage Mu *mom* transcription by C protein does not require specific interaction with the carboxyl-terminal region of the alpha or sigma-70 subunit of *Escherichia coli* RNA polymerase. *J. Bacteriol.*, **180**, 3257–3259.
24. Artsimovitch,I., Murakami,K., Ishihama,A. and Howe,M.M. (1996a) Transcription activation by the bacteriophage Mu Mor protein requires the C-terminal regions of both alpha and sigma70 subunits of *Escherichia coli* RNA polymerase. *J. Biol. Chem.*, **271**, 32343–32348.
25. Artsimovitch,I., Kahmeyer-Gabbe,M. and Howe,M.M. (1996b) Distortion in the spacer region of P_m during activation of middle transcription of phage Mu. *Proc. Natl. Acad. Sci. USA*, **93**, 9408–9413.
26. Artsimovitch,I. and Howe,M.M. (1996) Transcription activation by the bacteriophage Mu Mor protein: analysis of promoter mutations in P_m identifies a new region required for promoter function. *Nucleic Acids Res.*, **24**, 450–457.
27. Chakraborty,A. and Nagaraja,V. (2006) Dual role for transactivator protein C in activation of *mom* promoter of bacteriophage Mu. *J. Biol. Chem.*, **281**, 8511–8517.
28. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
29. Kumaraswami,M., Howe,M.M. and Park,H.W. (2004) Crystal structure of the Mor protein of bacteriophage Mu, a member of the Mor/C family of transcription activators. *J. Biol. Chem.*, **279**, 16581–16590.
30. Dodd,I.A. and Egan,J.B. (1987) Systematic method for the detection of potential λ Cro-like DNA-binding regions in proteins. *J. Mol. Biol.*, **194**, 557–564.
31. Dodd,I.A. and Egan,J.B. (1990) Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. *Nucleic Acids Res.*, **18**, 5019–5026.
32. Brennan,R.G. and Matthews,B.W. (1989) The helix-turn-helix DNA binding motif. *J. Biol. Chem.*, **264**, 1903–1906.
33. Miller,F.H. (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

34. Nikaido,H. (1961) Galactose-sensitive mutants of Salmonella. I. Metabolism of galactose. *Biochim. Biophys. Acta*, **48**, 460–469.
35. Simons,R.W., Houman,F. and Kleckner,N. (1987) Improved single and multicopy *lac*-based cloning vectors for protein and operon fusions. *Gene*, **53**, 85–96.
36. Chang,A.C. and Cohen,S.N. (1978) Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J. Bacteriol.*, **134**, 1141–1156.
37. Rose,R.E. (1988) The nucleotide sequence of pACYC177. *Nucleic Acids Res.*, **16**, 356–359.
38. McKenney,K., Shimatake,H., Court,D., Schmeissner,U., Brady,C. and Rosenberg,M. (1981) A system to study promoter and terminator signals recognized by *Escherichia coli* RNA polymerase. In Chirikjian J.G. and Papas T.S. (eds), *Gene Amplification and Analysis*, Vol. II. Elsevier/North-Holland, New York, N.Y, pp. 383–415.
39. Sutcliffe,J.G. (1979) Complete nucleotide sequence of the *Escherichia coli* plasmid pBR322. *Cold Spring Harb. Symp. Quant. Biol.*, **43**, 77–90.
40. Beaucage,S.L. and Caruthers,M.H. (1981) Studies on nucleotide chemistry V. Deoxynucleoside phosphoramidites—a new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Lett.*, **22**, 1859–1862.
41. Hermes,J.D., Parekh,S.M., Blacklow,S.C., Koster,H. and Knowles,J.R. (1989) A reliable method for random mutagenesis: the generation of mutant libraries using spiked oligodeoxyribonucleotide primers. *Gene*, **84**, 143–151.
42. Laemmli,U.K. (1970) Cleavage of structural proteins during the assembly of the head of the bacteriophage T4. *Nature*, **227**, 680–682.
43. Sambrook,J., Fritsch,E.F. and Mantiatis,T. (1989) *Molecular Cloning A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
44. Carey,J. (1991) Gel retardation. *Methods Enzymol.*, **208**, 103–117.
45. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
46. Chiang,L.W., Kovari,I. and Howe,M.M. (1993) Mutagenic oligonucleotide-directed PCR amplification (Mod-PCR): an efficient method for generating random base substitution mutations in a DNA sequence element. *PCR Methods Appl.*, **2**, 210–217.
47. Aravind,L., Anantharaman,V., Balaji,S., Babu,M.M. and Iyer,L.M. (2005) The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol. Rev.*, **29**, 231–262.
48. Wintjens,R. and Rooman,M. (1996) Structural classification of HTH DNA-binding domains and protein-DNA interaction modes. *J. Mol. Biol.*, **262**, 294–313.
49. Jones,S., van Heyningen,P., Berman,H.M. and Thornton,J.M. (1999) Protein-DNA interactions: a structural analysis. *J. Mol. Biol.*, **287**, 877–896.
50. Luscombe,N.M., Austin,S.E., Berman,H.M., and Thornton,J.M. (2000) An overview of the structures of protein-DNA complexes. *Genome Biol.*, **1**, REVIEWS001.
51. Rice,P.A., Yang,S., Mizuuchi,K. and Nash,H.A. (1996) Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell*, **87**, 1295–1306.
52. Bewley,C.A., Gronenborn,A.M. and Clore,G.M. (1998) Minor-groove binding architectural proteins: structure, function, and DNA recognition. *Annu. Rev. Biophys. Biomol. Struct.*, **27**, 105–131.
53. Klass,J., Murphy,F.V. IV, Fouts,S., Serenil,M., Changela,A., Siple,J. and Churchill,M.E.A. (2003) The role of intercalating residues in chromosomal high-mobility-group protein DNA binding, bending and specificity. *Nucleic Acids Res.*, **31**, 2852–2864.
54. Chen,C.-Y., Ko,T.-P., Lin,T.-W., Chou,C.-C., Chen,C.-J. and Wang,A.H.-J. (2005) Probing the DNA kink structure induced by the hyperthermophilic chromosomal protein Sac7d. *Nucleic Acids Res.*, **33**, 430–438.
55. Stott,K., Tang,G.S.F., Lee,K-B. and Thomas,J.O. (2006) Structure of a complex of tandem HMG boxes and DNA. *J. Mol. Biol.*, **360**, 90–104.
56. Gindlesperger,T.L. and Hattman,S. (1994) In vitro transcriptional activation of the phage Mu *mom* promoter by C protein. *J. Bacteriol.*, **176**, 2885–2891.
57. Rost,B. (1996) PHD: predicting one-dimensional protein structure by profile-based neural networks. *Methods in Enzymol.*, **266**, 525–539.