JOURNAL OF BACTERIOLOGY, Apr. 1989, p. 2019–2027
0021-9193/89/042019-09$02.00/0

Vol. 171, No. 4

Role of Bacteriophage Mu C Protein in Activation of the *mom* Gene Promoter

MICHAEL BÖLKER, F. GREGORY WULCZYN, AND REGINE KAHMANN*

Institut für Genbiologische Forschung Berlin GmbH, Ihnestrasse 63, D-1000 Berlin, Federal Republic of Germany

Received 15 August 1988/Accepted 27 December 1988

The phage Mu C gene product is a specific activator of Mu late gene transcription, including activation of the *mom* operon. Fusion of the *C* gene to the efficient translation initiation region of the *Escherichia coli atpE* gene allowed significant overproduction of C protein, which was subsequently purified and assayed for DNA binding by gel retardation and nuclease footprinting techniques. C protein binds to a site immediately upstream of the −35 region both of the *mom* promoter and the related phage D108 *mod* promoter. The location of the *mom* promoter has been determined by primer extension. Upstream deletions extending more than 3 base pairs into the C-binding site abolished activation of the *mom* promoter in vivo. In vitro binding of C was not significantly affected by DNA methylation. A second, C-dependent promoter was identified just downstream of the C coding region; comparison with the *mom* promoter revealed common structural elements.

In the life cycle of temperate bacteriophage Mu, the switch from early- to late-gene transcription requires the *C* gene product. Gene *C* is located between early and late genes on the Mu map, and its expression depends on Mu DNA replication (reviewed in reference 15). The *C* gene has been sequenced (11, 25), and from the derived amino acid sequence, the molecular mass of C is predicted to be 16.5 kilodaltons (kDa). Because of some sequence similarities to sigma factors, it has been speculated that C may represent a phage-specific sigma factor (25). At least four promoters are activated by C; they are located between *C* and *lys*, *G* and *I*, in *N*, and between *gin* and *mom* (see accompanying paper by Margolin et al. [26]).

The *mom* gene of phage Mu encodes an unusual DNA modification function (41, 43) which is subject to complex transcriptional as well as posttranscriptional control (reviewed in reference 18). The requirement for C is only one of the facets of transcriptional regulation at the *mom* promoter. In addition, methylation of three GATC sites in region I (Fig. 1) by the Dam methylase is a prerequisite for *mom* promoter activity in the wild-type situation. When these three sites are mutated or deleted, the Dam requirement ceases (17, 39), although activation by C is maintained. This indicates that the *cis*-acting regulatory elements are separate units. Current models suggest that an *Escherichia coli* protein may act as a repressor, turning off *mom* gene transcription by binding to region I when the GATC sites are un- or hemimethylated; this in turn could prevent access of C or RNA polymerase (9). For a more direct understanding of protein-DNA interactions in the regulation of *mom* gene transcription, characterization of individual components is necessary. In this study we describe a plasmid for the overproduction of the Mu C protein which has greatly facilitated C purification. We show, furthermore, that C is a specific DNA-binding protein.

MATERIALS AND METHODS

Bacteria and bacteriophage strains. All strains are derivatives of *E. coli* K-12. C600 carries *thr leu* (1); C600(P1Cm) carries a P1Cm prophage that is hsd_{P1}^{+} (43); GM33 is prototrophic W3110 *dam-3* (27); JM83 is *ara* Δ(*lac-pro*) *rpsL thi* (φ80d*lacZ*ΔM15) sup^{0} (44); CSH50 is *ara* Δ(*lac-pro*) *strA thi*; DK1066 is CSH50(Mu $cts62_{445\text{-}5}$), and due to a substitution the prophage has a Mom^{-} phenotype (5, 20). JR500 is C600 *hsdR lon-146* and was kindly provided by J. Rouviere-Yaniv. C600(Mu *cts*62 pAp1) has been described before (22); the prophage is com^{+} mom^{+}. C600(Mu *cts*62 pAp Δ*com*9) carries a 9-base-pair (bp) deletion in the *com* gene which results in a Mom^{-} phenotype (46). MH3105 *thr leu tonA lac supII* (Mu *cts*62 *C*am2005) was kindly provided by M. M. Howe.

Plasmids. pBR325 was used (2); pMuAS1C1 is a pBR322 derivative in which the rightmost 1,107 bp of Mu containing all of region I, *com*, and *mom* have been cloned (39). pMu1034 contains the rightmost 1,024 bp of Mu cloned into the *Cla*I-*Bam*HI sites of pBR322; region I is missing but *com* and *mom* are intact (17). pMuPH6R contains the Mu *C* gene and flanking sequences cloned as a 2.1-kilobase (kb) *Taq*I fragment into the *Cla*I site of pBR322 (11). In pJLA502 a promoter cassette consisting of λ p_L and λ p_R is followed by the TIR region of *atpE*, a linker region containing sites for *Nco*I and *Bam*HI, the λ *c*I857 repressor gene, and Ap^{r} as a selectable marker (38). pMLB1034 contains a truncated *lacZ* gene; the first eight codons, promoter, and ribosome-binding site are missing (40). Plasmid pMLys-Lac was constructed by cloning an *Hpa*II fragment (positions 10396 to 10530) from pMuPH6R into pMLB1034. To obtain an in-frame fusion to the *lacZ* gene, the *Hpa*II fragment (see Fig. 6) was first cloned into the *Cla*I site of the πVX polylinker (24), excised as an *Eco*RI fragment, and then inserted into the *Eco*RI site of plasmid pMLB1034 in such an orientation as to direct synthesis of a *lys-lacZ* fusion protein. The junction sequence starting at the *Hpa*II site (position 10531; see Fig. 6) is as follows: C CGG ATT GCC GAT GAT AAG CTG TCA AAC ATG AGA ATT CCC GGG GAT CCC GTC. The last GTC corresponds to codon 9 of *lacZ* in pMLB1034. pD108-175 was constructed by cloning the rightmost *Cla*I fragment of D108 DNA including about 0.5 kb of host DNA into the *Cla*I site of pBR322. This fragment contains the D108 *com* and *mod* genes but lacks D108 region I (42); it is analogous to the Mu fragment inserted in pMu1034.

General methods. Growth of bacteria and preparation of phage lysates were done as described before (4, 30). Plasmid isolation, restriction enzyme digestion, labeling of 5′ ends with polynucleotide kinase and [γ-^{32}P]ATP, and ligation with

* Corresponding author.

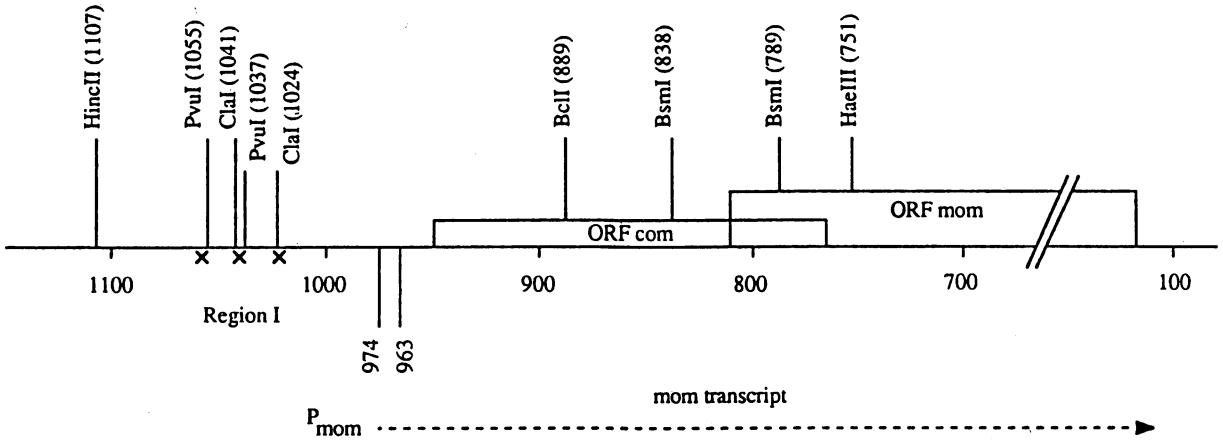


FIG. 1. Organization of the Mu *mom* operon. ORFs for *com* and *mom* are indicated by open bars. Numbering is from right to left according to Kahmann and Kamp (19). Only restriction sites used in this study are indicated. In pMuAS1C1, the *Hin*cII cleavage site has been converted to a *Bam*HI cleavage site. Asterisks indicate the three GATC sites in region I. Positions 974 and 963 indicate the different mRNA starts that were found by S1 nuclease mapping (9, 33). The *mom* transcript is indicated by a broken line; the arrow marks the direction of transcription.

T4 DNA ligase were performed according to standard procedures (24). The C and G+A sequencing reaction followed the conditions of Maxam and Gilbert (28); DNA was sequenced after fragments were subcloned in M13-derived vectors by the method of Sanger et al. (37). Total cell extracts for sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) were prepared as described before (29). SDS-PAGE was done on 15% gels as described by Laemmli (21); staining and destaining followed the procedure of Weber and Osborn (45).

Determination of the Mom phenotype. Plasmids to be tested for their Mom phenotype were introduced into DK1066. After thermal induction of the resident prophage, C is provided and can activate the *mom* promoter on the plasmid. Plating efficiencies of phage lysates were determined on C600 and C600(P1Cm). An efficiency of plating of $<10^{-4}$ indicates Mom$^-$, and an efficiency of $>10^{-2}$ indicates Mom$^+$ (10, 17, 43). Values in between are considered to indicate partial expression of the *mom* gene.

Primer extension. Isolation of RNA from C600(Mu *c*ts62 pAp1) and C600(Mu *c*ts62 pAp Δ*com*9) was done as described previously (46). The primer was a *Bsm*I restriction fragment isolated from the *mom* gene of plasmid pMu1034, extending from positions 837 to 789 (on the anti-sense strand). For primer hybridization, 25 μg of RNA was precipitated with 10 ng of restriction fragment and treated as described before (16). For the primer extension reaction 5 μl of the primer-template mix was incubated with 5 U of reverse transcriptase (Bio-Rad), 1 μl each of dATP, dGTP, and dTTP (2 mM), 2 μl of [α-^{32}P]dCTP (3,000 Ci/mmol; Amersham), 1 μl of dCTP (50 μM), and 3 μl of reverse transcriptase buffer (50 mM Tris hydrochloride [pH 8.0], 60 mM NaCl, 6 mM magnesium acetate, 10 mM dithiothreitol [DTT]) for 1 h at 42°C; 2 μl of dNTP mix (250 μM) was added, and incubation was continued for 15 min. Probes were prepared for electrophoresis as described before (16).

Purification of C protein. JR500(pJLC16) (1.2 liters) was grown in dYT medium containing ampicillin (60 μg/ml) to a cell density of 6×10^8/ml at 28°C. The temperature was raised to 42°C, and incubation continued for 3 h. Cells were collected by centrifugation, resuspended in buffer 1 (10% [vol/vol] glycerol, 20 mM Tris hydrochloride [pH 7.5], 1 mM EDTA, 1 mM DTT) containing 50 mM NaCl, and frozen at −80°C. After thawing, phenylmethylsulfonyl fluoride was added to a final concentration of 100 μg/ml. Cells were disrupted by two passages through a French pressure cell at 12,000 lb/in^2 (crude extract). Cell debris was pelleted by centrifugation at 40,000 rpm at 4°C in an SW55Ti rotor for 40 min. To the supernatant ammonium sulfate was added to a final concentration of 33%. Mixing at 0°C for 30 min was followed by centrifugation for 10 min at 15,000 × *g*. The pellet was suspended in 500 μl of buffer 1 containing 200 mM NaCl (fraction I). The sample was loaded on a Sephacryl S200 column (16 ml) equilibrated with the same buffer and eluted with the same buffer at a flow rate of 0.15 ml/min, and 750-μl fractions were collected. C protein eluted in fractions 17 to 20, which were combined (fraction II). The NaCl concentration was reduced to 100 mM by addition of buffer. Heparin-Sepharose (1.2 ml) equilibrated with buffer 1 containing 100 mM NaCl was added and gently mixed at 4°C for 1 h. The suspension was poured into a column (1 cm inner diameter) and washed with buffer 1 containing 100 mM NaCl until absorbance at 277 nm stayed constant. The bound material was eluted by a gradient of buffer 1 containing 100 to 300 mM NaCl (8 ml), followed by a wash with three column volumes of buffer 1 containing 500 mM NaCl. C eluted at an NaCl concentration of ca. 200 mM. The most concentrated fractions were pooled, dialyzed against 200 mM NaCl–20 mM Tris hydrochloride (pH 7.5)–1 mM EDTA–1 mM DTT–50% glycerol, and stored at −80°C (fraction III). Protein concentrations were determined according to Bradford (3). The yield in successive purification steps is depicted in Table 1.

Electrophoresis of protein-DNA complexes. The C protein-DNA complexes were studied by the technique of Fried and Crothers (7) with slight modifications. About 10 ng of 5′-end labeled fragment was incubated with 0.3 to 70 ng of purified

TABLE 1. Purification of Mu C protein

Fraction	Total vol (ml)	Total protein (mg)	Purity (%)[a]
Crude extract	14	147	5–10
Ammonium sulfate	0.5	2	20
Sephacryl S200	3.5	0.7	50
Heparin-Sepharose	0.6	0.14	90

[a] Purity was estimated from Coomassie-stained gels.

Mu C protein (fraction III) in a total volume of 20 μl containing 100 mM NaCl, 25 mM Tris hydrochloride (pH 7.9), 10 mM $MgCl_2$, 1.2 mM EDTA, 2 mM DTT, 10 μg of bovine serum albumin (BSA), and 100 ng of sonicated calf thymus DNA. After 20 min at 25°C, 5 μl of loading buffer, which is identical to the reaction buffer except that it contains 5% Ficoll, was added. The samples were loaded on an acrylamide gel (20 by 20 cm; 5% acrylamide, 0.16% bisacrylamide, 100 mM Tris-borate [pH 8.3], 1 mM EDTA). Electrophoresis was carried out at 100 V for 8 to 10 h at room temperature.

MPE · Fe(II) footprint analysis. About 100 ng of 5′-end ^{32}P-labeled restriction fragment was incubated with 50 to 400 ng of purified Mu C protein in a total volume of 20 μl containing 100 mM NaCl, 25 mM Tris hydrochloride (pH 7.9), 10 mM $MgCl_2$, 1.2 mM EDTA, 10 μg of BSA, and 100 ng of sonicated calf thymus DNA for 20 min at 25°C. Cleavage reagent methidiumpropyl-EDTA (MPE; kindly provided by P. Dervan) at a concentration of 1.2 mM was mixed with an equal volume of 1.2 mM $Fe(II)(NH_4)_2(SO_4)_2$ and immediately diluted fivefold with H_2O; 2.5 μl of this mixture was added, and the reaction was started by addition of 2.5 μl of DTT (10 mM), kept for 10 min at 30°C, and stopped by freezing in dry ice-ethanol. Samples were thawed by adding 50 μl of a solution containing 0.5 M sodium acetate and 20 μg of tRNA per ml. DNA was extracted once with phenol and once with phenol-chloroform-isoamyl alcohol (25:24:1), precipitated with ethanol, and redissolved in loading dye (95% formamide, 10 mM EDTA, 0.1% xylene cyanol, 0.1% bromphenol blue). After incubation for 1 min at 90°C, samples were loaded on 0.2-mm-thick 6% acrylamide sequencing gels.

RESULTS

Construction of a plasmid overproducing Mu C protein. After a series of unsuccessful attempts to overproduce Mu C protein by making promoter fusions to λ p_L or T7 φ10, we reasoned that these failures could indicate either inefficient translation of the *C* open reading frame (ORF) or instability of the protein. To address the first possibility, we decided to make use of the plasmid vector system of Schauder et al. (38), in which the highly efficient translation initiation region (TIR) of the *atpE* gene of *E. coli* is fused at the ATG to the coding region of interest. Plasmid pJLA502 was linearized with *Nco*I and the protruding 5′ ends were filled with Klenow fragment to regenerate the ATG, followed by cleavage with *Bam*HI. pMuPH6R, a plasmid containing *C*, was linearized with *Hpa*I 14 bp upstream of the ATG, digested with *Bal*31 to remove the untranslated portion preceding *C*, and cleaved with *Bam*HI. The population of fragments encompassing *C* were cloned into pJLA502 as prepared above and transformed into the Sup^0 host JM83. To identify clones expressing functional C protein, Ap^r transformants were transferred to a plate spread with soft agar containing about 2×10^9 Mu *C*am phage at 37°C. The production of C from the plasmid allows *C*am complementation, phage development, and cell lysis. Four clones which showed reduced growth in the presence of the *C*am phage were analyzed further. They all showed full complementation of the *C*am phage when the phage was spotted on a lawn of cells harboring the respective plasmids and severe reduction of growth when shifted to 42°C (data not shown). When analyzed for the presence of C protein in crude cell extracts after a temperature shift to 42°C, in all cases a new band appeared (≈4% of total cellular protein, not shown) which had the expected size of C in plasmids pJLC1,-3, and -16 but was larger by about 1 kDa in pJLC9. The amount of these peptides could be increased about threefold by using the *lon* mutant JR500 (Fig. 2A) and was highest in strains harboring pJLC16. To determine the exact points of fusion between the vector and the *C* gene, the junctions were sequenced after the respective fragments were subcloned. The results are summarized in Fig. 2B. Only plasmid pJLC1 had the intended junction sequence, where fusion occurred at the ATG. In pJLC3 and pJLC16, the ATG of pJLA502 was deleted and ORF *C*, including 8 or 7 bp, respectively, of the 5′ untranslated region, was fused to the TIR. In pJLC9, the

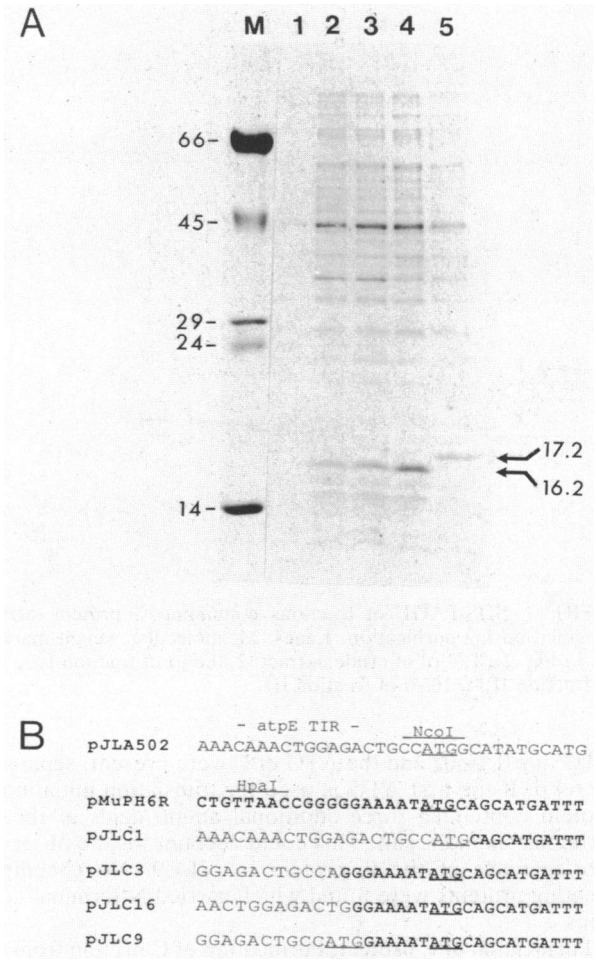


FIG. 2. Overproduction of C protein. (A) SDS-PAGE of crude cell extracts from strains overproducing Mu C protein. Lane M, Molecular weight markers: lysozyme, 14,000; trypsinogen, 24,000; carbonic anhydrase, 29,000; ovalbumin, 46,000; BSA, 66,000. Extracts were prepared from JR500 carrying the indicated plasmids: lane 1, pJLC16 (28°C); lane 2, pJLC1 (42°C); lane 3, pJLC3 (42°C); lane 4, pJLC16 (42°C); lane 5, pJLC9 (42°C); 28°C indicates that the sample was taken prior to heat induction, and 42°C indicates that samples were taken 3 h after the shift to 42°C. Bands were visualized after staining with Coomassie brilliant blue according to Weber and Osborn (45). The overproduced peptides are indicated by arrows. (B) Partial nucleotide sequence of TIR-*C* junctions. The pJLA502 sequence (lightface letters) was kindly provided by J. McCarthy; the pMuPH6R sequence (boldface letters) containing the start of gene *C* was taken from Heisig and Kahmann (11). ATG codons are underlined. Only restriction sites important for obtaining fusions are shown. The TIR region is not shown in its entirety.

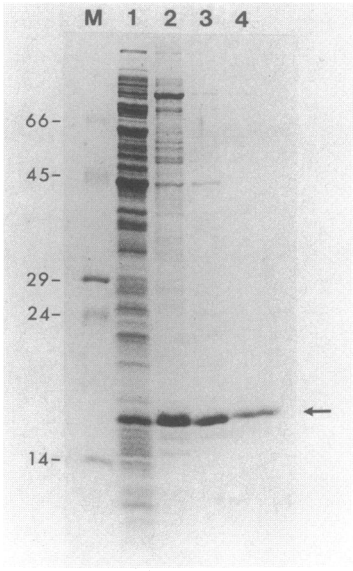


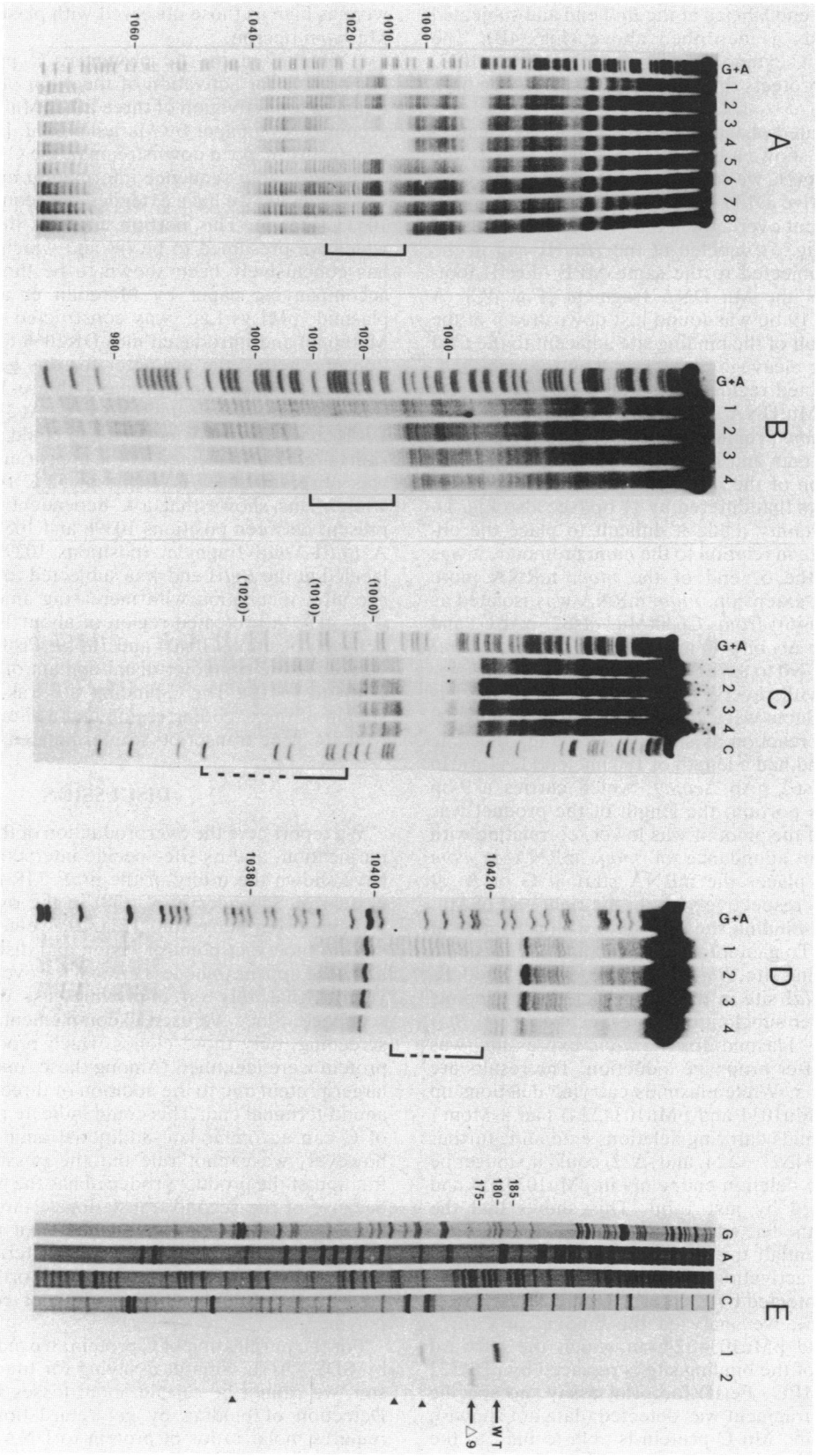
FIG. 3. SDS-PAGE of fractions containing C protein (arrow) obtained during purification. Lanes: M, molecular weight markers (as in Fig. 2); 1, 3 μl of crude extract; 2, 0.6 μl of fraction I; 3, 4 μl of fraction II; 4, 10 μl of fraction III.

ATG of pJLA502 and the ATG of *C* were present, separated by 6 bp. If the first ATG is used for translation initiation, a protein containing three additional amino acids at the N-terminus should result. This could account for the observed size increase of the C protein in pJLC9. No *C*-complementing proteins were found which carried N-terminal deletions.

Purification of C protein. Purification of C protein from the overproducing strain JR500(pJLC16) involved disruption of the cell paste in a French pressure cell, ammonium sulfate fractionation, chromatography on Sephacryl S200, and chromatography on heparin-Sepharose. The details are described in Materials and Methods. Fractions containing C protein (16.2 kDa) were identified by SDS-PAGE. The enrichment achieved in the individual steps is shown in Fig. 3. The last purification step yielded C protein which was about 90% pure. Fractions from heparin-Sepharose were also tested in gel retardation assays with a fragment containing the *mom* promoter (see below). Retention of the fragment correlated with the amount of C protein present in the respective fractions, which illustrated that specific DNA-binding activity resided in the 16.2-kDa protein (not shown).

C is a specific DNA-binding protein. To analyze whether C protein is able to interact with DNA directly or whether it does so only in conjunction with RNA polymerase, as sigma factors do, gel retardation assays were performed. Complexes were formed only with fragments containing the *mom* promoter and extending past the upstream *Cla*I site at position 1024 (data not shown). This could indicate that the binding site for C overlaps the *Cla*I site or is located so close to the fragment end that the binding was affected. To obtain these complexes, a 50:1 molar ratio of protein to DNA was required. This ratio did not change when a different C preparation was used. We therefore addressed the question of specificity of C protein binding by footprint analyses with the cleavage reagent MPE · Fe(II) (12). A *Hae*III-*Bsm*I fragment extending from the *Bsm*I site at position 838 to a *Hae*III site in the vector in pMuAS1C1 was 5′-end labeled at the *Hae*III end. After incubation with increasing amounts of purified C protein, the mixture was treated with MPE · Fe(II) (Fig. 4A). A single protected region was observed upstream of *com* between positions 1006 and 1024. In parallel, the same DNA fragment isolated from a *dam* mutant host was used (Fig. 4A) to see whether Dam methylation would affect the footprint. Although slightly better protection in Dam-modified DNA was seen (Fig. 4A), this effect was not always observed. This leads us to conclude that neither the size of the protected region nor the amount of protein needed to obtain protection differs significantly in Dam^+ and Dam^- DNA. Dam^- DNA was also used to determine the location of the C-binding site on the lower strand. A *Bcl*I-*Hae*III fragment (extending from the *Bcl*I site

FIG. 4. MPE-Fe(II) footprints of Mu C protein on C-dependent promoters and mapping of the *mom* promoter by primer extension. MPE-Fe(II) reactions were performed as described in Materials and Methods. Footprints are indicated by brackets; broken lines indicate a region of only partial protection by C protein. (A) C-binding to the Mu *mom* promoter (upper strand). A fragment extending from a *Hae*III site in pBR322 across position 1107 to the *Bsm*I site (position 838; see Fig. 1) was 5′-end ^{32}P-labeled at the *Hae*III end. Lanes 1–4, Fragment isolated from pMuAS1C1 (Dam^+); lanes 5–8, same fragment from pMuAS1C1 (Dam^-). Lanes 1 and 5 contained no added C protein; lanes 2 and 6 contained 50 ng of C protein; lanes 3 and 7 contained 100 ng of C protein; lanes 4 and 8 contained 200 ng of C protein. G+A is a Maxam-Gilbert sequencing reaction performed on the same fragment. Part of region I (positions 1068 to 1040) shows reduced sensitivity to MPE-Fe(II) cleavage. (B) C binding to the *mom* promoter (lower strand). A fragment extending from the *Bcl*I site (position 889; see Fig. 1) to the *Hae*III site in pBR322 was isolated from pMuAS1C1 (Dam^-) and ^{32}P labeled at the *Bcl*I 5′ end. G+A is a Maxam-Gilbert sequencing reaction performed on the same fragment. Lane 1, no C protein was added; lane 2 contained 100 ng; lane 3 contained 200 ng; and lane 4 contained 400 ng of C protein. (C) C binding to the D108 *mod* promoter. A fragment extending from a *Hae*III site in pBR322 across the *Cla*I site (see Fig. 5) to the *Bsm*I site located at a position corresponding to position 838 in Mu was 5′-end labeled with ^{32}P at the *Hae*III end. The fragment originated from D108-175. Lanes G+A and C are Maxam-Gilbert sequencing reactions performed on the same fragment. Lane 1, no C protein was added; lane 2 contained 100 ng, lane 3 contained 200 ng, and lane 4 contained 400 ng of C protein. Numbers in parentheses on the left indicate the corresponding positions in Mu DNA. (D) C binding to the *lys* promoter. A *Bgl*II-*Xmn*I fragment extending from positions 10299 to 10461 (34) was 5′-end ^{32}P labeled at the *Bgl*II end and incubated without protein (lane 1); with 100 ng of C protein (lane 2); with 200 ng of C protein (lane 3); or with 400 ng of C protein (lane 4). G+A indicates Maxam-Gilbert sequencing reactions performed on the same fragment. The region between positions 10370 and 10388 which contains the transcription terminator shows reduced sensitivity to MPE-Fe(II) cleavage. (E) Primer extension analysis of the *mom* mRNA. G, A, T, and C are DNA sequence reactions by the dideoxy chain termination procedure performed on plasmid DNA. The plasmid was a derivative of plasmid pTSV11 (47) carrying a 124-bp *mom* insert cloned into the plasmid's *Eco*RI and *Sma*I sites. Fragment size is indicated to the left of lane G (in nucleotides). The *Bsm*I fragment (positions 789 to 838) was used as a primer for extension on mRNA. Lane 1, extension products obtained with RNA isolated from C600(Mu *c*ts62 pAp1); lane 2, extension products obtained with RNA isolated from C600(Mu *c*ts62 pAp Δ*com*9). The major products are indicated on the right. WT indicates the major extension product from the wild type, and Δ9 indicates the extension product from *mom* mRNA carrying a 9-bp deletion in *com*. Arrowheads point to minor products of greater mobility which may be the result either of pausing by the reverse transcriptase or of extension on mRNA degradation intermediates. Details of the procedure used are given in Materials and Methods.


A
1060-
1040-
1020-
1010-
1000-
G+A
1 2 3 4 5 6 7 8
B
980-
1000-
1010-
1020-
1040-
G+A
1 2 3 4
C
(1020)-
(1010)-
(1000)-
G+A
1 2 3 4 C
D
10380-
10400-
10420-
G+A
1 2 3 4
E
185-
180-
175-
G A T C
1 2
WT
Δ9

at position 889 to a *Hae*III site in the vector portion of pMuAS1C1) was 5′-end labeled at the *Bcl*I end and subjected to footprint analysis as described above (Fig. 4B). The region protected by C extended from positions 1006 to 1023. This shows that C protects the same region of DNA on both strands.

Since the Mu-related phage D108 contains a modification function (*mod*) that shows a high degree of homology to the Mu *mom* function (42), we also analyzed the binding of C protein to the putative D108 *mod* promoter. To this end, a *Hae*III-*Bsm*I fragment overlapping the *Cla*I site was isolated from pD108-175 (Fig. 5), labeled at the *Hae*III end in the vector part, and subjected to the same MPE · Fe(II) footprinting regimen as the Mu DNA fragment (Fig. 4C). A protected region of 19 bp was found just downstream of the *Cla*I site. The left half of the binding site adjacent to the *Cla*I site showed weaker cleavage protection than the right half. Within the C protected region were 2 bp that differed from the corresponding Mu DNA sequence (Fig. 6).

Mapping of the *mom* promoter by primer extension. Previous studies by Hattman and Ives (9) and Plasterk et al. (33) aimed at localization of the *mom* promoter by S1 nuclease mapping gave results that differed by 11 bp (see also Fig. 1). Because this uncertainty made it difficult to place the observed C-binding site in relation to the *mom* promoter, it was desirable to map the 5′ end of the *mom* mRNA more precisely by primer extension. *mom* mRNA was isolated as described before (46) from C600(Mu *c*ts62 pAp1) and C600(Mu *c*ts62 pAp Δ*com*9) 40 min after induction. A *Bsm*I fragment (positions 790 to 837, Fig. 1) was used as the primer for the extension with reverse transcriptase (see Materials and Methods). Products were analyzed on a sequencing gel next to a sequence reaction as size markers (Fig. 4E). The most prominent band had a length of 180 nucleotides (nt). In the case of Mu *c*ts62 pAp Δ*com*9, which carries a 9-bp deletion in the *com* portion, the length of the product was reduced by 9 nt and the amount was lower, correlating with the observed lower abundance of *mom* mRNA in *com* mutants (46). This places the mRNA start at G or A, at position 971 or 970, respectively, from the right end of Mu.

Presence of the C-binding site correlates with activation of the *mom* promoter. To gain information about the functional limits of the C-binding site, *Bal*31 deletions were introduced starting from the *Cla*I site in pMu1034. Deletion endpoints were sequenced after subcloning the respective *Eco*RI-*Bcl*I fragments in M13. Plasmid-driven *mom* expression was tested in DK1066 after prophage induction. The results are summarized in Fig. 5. While plasmids carrying deletions up to position 1022 (pMu1034 and pMu1034Δ14) had a Mom^+ phenotype, all plasmids carrying deletions extending further to the right (pMu1034Δ27, -Δ24, and -Δ22) could no longer be activated by C. The deletion endpoints in pMu1034Δ14 and pMu1034Δ27 differed by just 1 bp. This shows that the outermost 2 bp on the left end of the region protected by C protein are nonessential; the A at position 1022, however, must be present for activation of the *mom* promoter in vivo. Since the region protected by C showed some dyad symmetry (see Discussion), we analyzed the binding of C to a fragment of plasmid pMu1034Δ24, in which the leftward symmetry element of the binding site is replaced by pBR322 sequences. In an MPE · Fe(II) footprint assay, no specific binding of C to this fragment was detected (data not shown).

The finding that the Mu C protein is able to bind to the D108 *mod* promoter led us to ask whether C is able to activate this promoter in vivo. To this end, plasmid pD108-175 was introduced into DK1066 and assayed for *mod* expression. As can be seen in Fig. 5, *mod* expression levels were as high as those observed with plasmids containing the Mu *mom* operon.

C binding to the *lys* promoter. C protein is not only responsible for activation of the *mom* promoter but is also required for activation of three other Mu late promoters (see accompanying paper by Margolin et al. [26]). The first gene genetically mapped downstream of the C coding region is *lys* (15). Since only sequence information up to position 10461 was available, we have extended the sequence up to position 10533 (Fig. 6). This portion contains the start of an ORF which we presumed to be *lys* and which by marker rescue has conclusively been shown to be the Mu *lys* gene (see accompanying paper by Margolin et al. [26]). A fusion plasmid, pMLys-Lac, was constructed (see Materials and Methods) and introduced into DK1066 to address the question of whether the promoter for this gene is C regulated. Cells were plated on 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) at 32 and 37°C. At 37°C the prophage in DK1066 was partially induced and could provide C. Expression of *lacZ*, as indicated by the appearance of blue colonies, was observed at 37°C, while at 32°C the colonies stayed white. This shows that a C-dependent promoter must be present between positions 10396 and 10532 on the Mu map. A *Bgl*II-*Xmn*I fragment (positions 10299 to 10461) 5′-end labeled at the *Bgl*II end was subjected to MPE-Fe(II) cleavage after incubation with increasing amounts of C protein (Fig. 4D). A protected region of about 16 bp was observed between positions 10404 and 10419. Protection was slightly weaker in the promoter-distal domain of this site (positions 10404 to 10412). The C-binding site was located just downstream of the *C*-coding region and a site (10378 to 10404) at which *C* gene transcripts are terminated (11, 25).

DISCUSSION

We report here the overproduction of the Mu C protein, its purification, and its site-specific interaction with DNA. We have shown the utility of the *atpE* TIR-derived expression system of Schauder et al. (38) in the overproduction of C protein. From our results (Fig. 2), we conclude that in this system the exact point of fusion and distance between TIR and gene appear to be less critical for overexpression than in vectors that rely on improving just the Shine-Dalgarno sequence. Since we used *C* complementation for the initial screening, only those clones which produced functional C protein were identified. Among these, one clone produced a larger protein due to the addition of three amino acids to the amino-terminal end. This could indicate that the N-terminus of C can accommodate additional amino acids. Formally, however, we cannot rule out the possibility that a small fraction of the product produced has the genuine N-terminus because of the second translational start and that it is this fraction which is active. The fact that none of the active proteins contained deletions in the N-terminus may indicate that this portion of the protein is important for its activity, e.g., correct folding, dimerization, or contacts with RNA polymerase.

For the purification of C protein, we monitored the protein by SDS-PAGE without assaying for function. For this reason, we cannot be certain about losses in specific activity. Detection of binding by gel retardation and footprinting requires molar ratios of protein to DNA of 50:1, indicating that only a small fraction of the purified C protein is able to bind to DNA. In future, a C-dependent in vitro transcription assay will be needed to firmly establish this ratio and to

		Phage titer on C600	C600(P1Cm)	EOP
	1020 1010 ClaI			
pMuAS1C1	TACAGATCGATTATGCCCCAATAACCACACT	6×10^{9}	2×10^{8}	3.3×10^{-2}
pD108-175	TTATCATCGATTATGTCTCAATAACCACACT	1×10^{10}	1×10^{10}	1
pMu1034	TTATCATCGATTATGCCCCAATAACCACACT	7×10^{9}	1×10^{10}	1.4
pMu1034Δ14	ATGTTTGACATTATGCCCCAATAACCACACT	4×10^{9}	5×10^{9}	1.3
pMu1034Δ27	GTTTGACAGCTTATGCCCCAATAACCACACT	1×10^{10}	3×10^{5}	3×10^{-5}
pMu1034Δ24	TCTCATGTTTGACAGCCCCAATAACCACACT	6×10^{9}	3×10^{3}	5×10^{-7}
pMu1034Δ22	TCTCATGTTTGACAGCTTATCATAACCACACT	1×10^{10}	2×10^{3}	2×10^{-7}

FIG. 5. *mom* expression in plasmids carrying deletions extending into the C-binding site. The partial nucleotide sequences shown overlap the deletion endpoints. Lightface letters indicate pBR322, and boldface letters mark Mu and D108 sequences (in pD108-175). Mu sequences are taken from Kahmann (17); D108 sequences are derived from Szatmari et al. (42). The thick underline indicates the region that is protected by C protein. *mom* expression was assayed as described in Materials and Methods after transforming plasmids in DK1066. Values represent averages of at least two independent experiments. EOP, Efficiency of plating.

optimize the purification scheme. Nevertheless, we ascribe the observed footprints to C protein, based on the expected behavior of C in activation of the three promoters analyzed and on the deletion analysis of the *mom* promoter. Specific DNA binding of C protein was also inferred by the group of S. Hattman (reported in reference 31).

The site-specific binding of C protein to the *mom* promoter of Mu, the *mod* promoter of D108, and the *lys* promoter of Mu as assayed by gel retardation and footprinting suggests that the C protein is not a sigma factor but rather an accessory factor (35) that binds to DNA specifically and independently of RNA polymerase. In vivo activity of these promoters in the absence of C protein is barely detectable, and for the *mom* promoter we have shown that expression is stimulated at least 200-fold when C is provided in *trans* (M. Bölker, unpublished). The mapping of the 5′ end of the *mom* mRNA by primer extension to position 970 or 971 places the −10 region within the sequence TAAGAT (Fig. 6), which resembles the consensus sequence for *E. coli* promoters (TATAAT) (36). A corresponding consensus −35 region (TTGACA) was not detectable. However, it is the region upstream of position −35 that is protected by C protein. If we now take the C-binding site of *mom* for alignment with the binding site in the *lys* promoter, there is a 10 of 16 bp match (or 13 of 16 bp if one allows for pyrimidines at three positions) within the C-protected regions (Fig. 6). The putative −10 region of the *lys* promoter (TAGACT) has the same distance to the C-protected region as in *mom* and again resembles the *E. coli* consensus. Between the −35 and −10 regions, the two promoters share little homology except for a stretch of six T's between −12 and −17. Since the other C-dependent promoters do not share this hexamer motif (see accompanying paper by Margolin et al. [26]), we cannot assign a functional role to this sequence. The deletion analysis of the *mom* promoter (Fig. 5) clearly shows that nucleotides at position −53 and −52 are not important for function; however, deletion of the A in the adjacent ATTAT motif renders the *mom* promoter inactive. It may not be a coincidence that the same motif is conserved in the *lys* promoter, but here the sequence is located on the lower strand in the promoter-proximal part of the C-binding site (Fig. 6). In the *mom* and *mod* promoters, this ATTAT motif is conserved and furthermore is present twice and forms part of an imperfect palindrome (Fig. 6). The two halves are

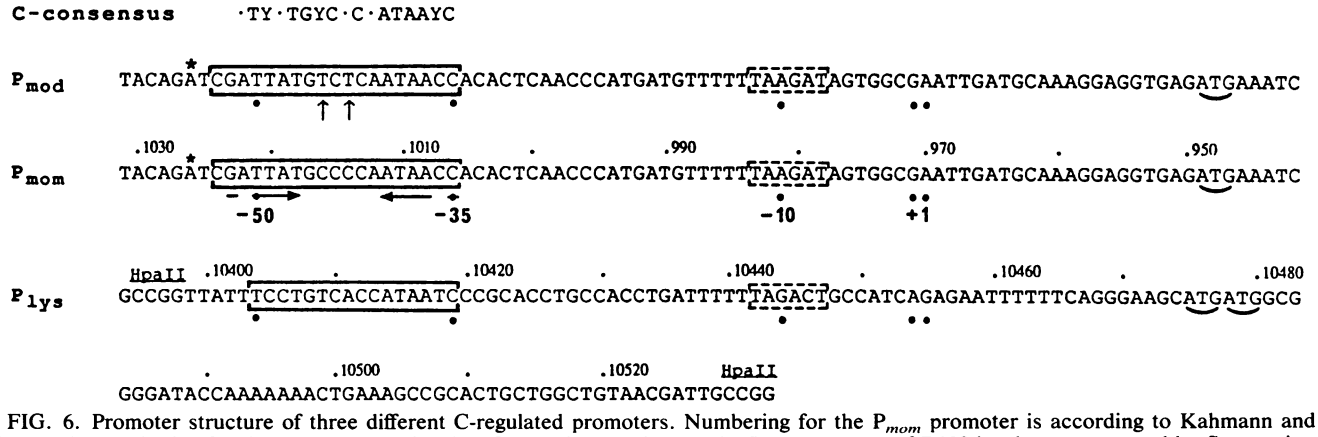


FIG. 6. Promoter structure of three different C-regulated promoters. Numbering for the P_{mom} promoter is according to Kahmann and Kamp (19); numbering for the P_{lys} promoter is taken from Priess et al. (34); the P_{mod} promoter of D108 has been sequenced by Szatmari et al. (42). C-binding sites are boxed (thick line), putative −10 regions are boxed (broken line), and translation starts are indicated by an arch. The +1 mRNA start for *mom* as determined by primer extension is indicated by dots; other numbers refer to this mRNA start. Symmetry elements are indicated by horizontal arrows. Vertical arrows denote differences between the P_{mod} and P_{mom} promoters.

separated by a 6-bp pyrimidine-rich spacer. It is tempting to speculate that the palindromic nature of the site may indicate that C protein binds to such sites as a dimer. Future experiments with sedimentation analysis should settle the question of whether active C protein exists as a monomer or as a multimeric form in solution. In the *lys* promoter, the promoter-distal site does not match the ATTAT motif very well, and this may explain the weaker protection of this region in the footprint. Since we analyzed binding of C protein to only three promoters, including the nearly homologous promoters of *mom* and *mod*, the derived consensus C-binding sequence (Fig. 6) must remain preliminary.

It is interesting that the two different approaches that were used by us and Margolin et al. (26) to define a consensus sequence for activation by Mu C protein have led to the same alignment of sequences but differences in the derived C consensus. The ATAA motif which we consider a right-half site is close to the left border of the consensus defined by Margolin et al. (26). Their consensus instead extends further to the right, into a region which is not protected by C protein. Conceivably, conserved nucleotides in this region are necessary for specific RNA polymerase contacts. The fact that Margolin et al. (26) have not included the region upstream of −42 in their consensus may indicate that binding of C protein in this region is not necessary in all of the promoters or is weaker in this region, as can be seen by the reduced protection of the promoter-distal part of the C-binding site of the *lys* and *mod* promoters. At least in the case of the *mom* promoter, sequences up to position −51 are necessary for activation by C protein, as shown by deletion analysis. Since for the other promoters, the functional limits of the activation sites have not yet been determined, it is presently not evident why there is such a low degree of homology in the promoter-distal portions of the sites.

In other systems where binding of positive activators has been analyzed, promoters also lack consensus −35 sequences if they are completely dependent on activation. The binding sites for the activators are located either within the −35 or just upstream of this region for the λ *c*I, *araC*, and *malT* proteins (14, 23, 35). For λ *c*II protein, the binding sites flank the −35 region on either side (13). Only in the case of λ *c*I protein has it been demonstrated that there are protein-protein contacts between the positive activator and RNA polymerase (8, 14). The strict conservation of distances between C-binding sites and −10 regions in the various promoters analyzed could indicate a similar activation mechanism. In this respect it should also be interesting to determine whether C protein increases the affinity of RNA polymerase to the *mom* promoter or affects the rate of open complex formation. For all the described positive activators, a helix-turn-helix motif is characteristic (32). We therefore analyzed the primary amino acid sequence of C protein for the presence of such a motif. Using the method of Dodd and Egan (6), we detected a stretch of 20 amino acids near the carboxyl-terminus of C which shows considerable homology to λ Cro-like DNA-binding domains (Fig. 7), despite the fact that in two of the three most highly conserved amino acids C does not match the consensus. For the C protein, a score of 1,776 was estimated by this merely statistical method. All proteins found in the protein sequence data base of the Protein Identification Resource (PIR) scoring more than 1,700 (16 proteins of 2,560) were judged to be Cro-like (6).

A further point worthy of discussion is the nature of the repression of the *mom* promoter when the upstream GATC sites are unmethylated. We have shown that although the C-binding site is adjacent to GATC site III, C binding in vitro is not significantly affected by DNA methylation. This is in agreement with the proposal that region I is recognized by a cellular repressor when the GATC sites are un- or hemimethylated. This model is especially attractive since mutation of site III allows partial *mom* expression in a *dam* mutant background, while single mutations in either of the upstream GATC sites (I and II) have no effect (39). It will be interesting to test whether occupancy of region I by the repressor can be shown to inhibit C binding in vivo or in vitro.

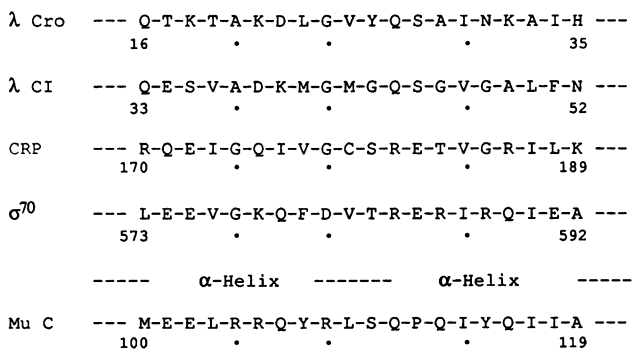


FIG. 7. Putative DNA-binding domain of Mu C protein. Alignment of a 20-amino-acid region of C protein with the helix-turn-helix motif characteristic for the DNA-binding domains of λ Cro, λ *c*I, CRP, and σ^{70}. The highly conserved positions of this motif are indicated by dots. Numbers refer to amino acid positions; sequence data are taken from Pabo and Sauer (32), Margolin and Howe (25), and Heisig and Kahmann (11).

ACKNOWLEDGMENTS

We thank J. McCarthy for plasmids, M. M. Howe and J. Rouviere-Yaniv for strains, and P. Dervan for MPE. The communication of results by W. Margolin, G. Rao, and M. M. Howe prior to publication is gratefully acknowledged. We thank C. Koch and A. Klippel for their advice on protein purification and footprinting techniques.

This work was supported by the Deutsche Forschungsgemeinschaft (Ka 411/4-1).

LITERATURE CITED

1. **Appleyard, R. K.** 1954. Segregation of lambda lysogenicity during bacterial recombination in *Escherichia coli* K12. Genetics **39**:429–439.
2. **Bolivar, F.** 1978. Construction and characterization of new cloning vehicles. III. Derivatives of plasmid pBR322 carrying unique *Eco*RI sites for selection of *Eco*RI generated recombinant DNA molecules. Gene **4**:121–136.
3. **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72**:248–254.
4. **Bukhari, A. I., and E. Ljungquist.** 1977. Bacteriophage Mu: methods for cultivation and use, p. 749–756. *In* A. I. Bukhari, E. Ljungquist, F. de Bruijn, and H. Khatoon (ed.), DNA insertion elements, plasmids and episomes. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
5. **Chow, L. T., R. Kahmann, and D. Kamp.** 1977. Electron microscopic characterization of DNAs of non-defective deletion mutants of bacteriophage Mu. J. Mol. Biol. **113**:591–609.
6. **Dodd, I. B., and J. B. Egan.** 1987. Systematic method for the detection of potential lambda Cro-like DNA-binding regions in proteins. J. Mol. Biol. **194**:557–564.
7. **Fried, M., and D. Crothers.** 1981. Equilibria and kinetics of *lac* repressor-operator interactions by polyacrylamide gel electrophoresis. Nucleic Acids Res. **9**:6505–6525.
8. **Guarente, L., J. S. Nye, A. Hochschild, and M. Ptashne.** 1982.

Mutant lambda phage repressor with a specific defect in its positive control function. Proc. Natl. Acad. Sci. USA **79:** 2236–2239.
9. **Hattman, S., and J. Ives.** 1984. S1 nuclease mapping of the phage Mu *mom* gene promoter: a model for the regulation of *mom* expression. Gene **29:**185–198.
10. **Hattman, S., M. Goradia, C. Monaghan, and A. I. Bukhari.** 1983. Regulation of the DNA-modification function of bacteriophage Mu. Cold Spring Harbor Symp. Quant. Biol. **47:**647–653.
11. **Heisig, P., and R. Kahmann.** 1986. The sequence and *mom*-transactivation function of the C gene of bacteriophage Mu. Gene **43:**59–67.
12. **Hertzberg, R. P., and P. B. Dervan.** 1984. Cleavage of DNA with methidiumpropyl-EDTA-iron(II): reaction conditions and product analyses. Biochemistry **23:**3934–3945.
13. **Ho, Y. S., D. L. Wulff, and M. Rosenberg.** 1983. Bacteriophage λ protein cII binds promoters on the opposite face of the DNA helix from RNA polymerase. Nature (London) **304:**703–708.
14. **Hochschild, A., N. Irwin, and M. Ptashne.** 1983. Repressor structure and the mechanism of positive control. Cell **32:** 319–325.
15. **Howe, M. M.** 1987. Late genes, particle morphogenesis, and DNA packaging, p. 63–74. *In* N. Symonds, A. Toussaint, P. van de Putte, and M. M. Howe (ed.), Phage Mu. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
16. **Hultmark, D., R. Klemenz, and W. J. Gehring.** 1986. Translational and transcriptional control elements in the untranslated leader of the heat-shock gene hsp22. Cell **44:**429–438.
17. **Kahmann, R.** 1983. Methylation regulates the expression of a DNA-modification function encoded by bacteriophage Mu. Cold Spring Harbor Symp. Quant. Biol. **47:**639–646.
18. **Kahmann, R., and S. Hattman.** 1987. Regulation and expression of the *mom* gene, p. 93–109. *In* N. Symonds, A. Toussaint, P. van de Putte, and M. M. Howe (ed.), Phage Mu. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
19. **Kahmann, R., and D. Kamp.** 1987. Sequence of the right end of Mu, p. 297–308. *In* N. Symonds, A. Toussaint, P. van de Putte, and M. M. Howe (ed.), Phage Mu. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
20. **Kahmann, R., A. Seiler, F. G. Wulczyn, and E. Pfaff.** 1985. The *mom* gene of bacteriophage Mu: a unique regulatory scheme to control a lethal function. Gene **39:**61–70.
21. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227:**680–685.
22. **Leach, D., and N. Symonds.** 1979. The isolation and characterization of a plaque-forming derivative of bacteriophage Mu carrying a fragment of Tn*3* conferring ampicillin resistance. Mol. Gen. Genet. **172:**179–184.
23. **Lee, N. L, W. O. Gielow, and R. G. Wallace.** 1981. Mechanism of *araC* autoregulation and the domains of two overlapping promoters, P_C and P_{BAD}, in the L-arabinose regulatory region of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **78:**752–756.
24. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
25. **Margolin, W., and M. M. Howe.** 1986. Localization and DNA sequence analysis of the *C* gene of bacteriophage Mu, the positive regulator. Nucleic Acids Res. **14:**4881–4897.
26. **Margolin, W., G. Rao, and M. M. Howe.** 1989. Bacteriophage Mu late promoters: four late transcripts initiate near a conserved sequence. J. Bacteriol. **171:**2003–2018.
27. **Marinus, M. G., and N. R. Morris.** 1975. Pleiotropic effects of a DNA adenine methylation mutation (*dam-3*) in *Escherichia coli* K12. Mutat. Res. **28:**15–26.
28. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65:**499–560.
29. **Mertens, G., A. Hoffmann, H. Blöcker, R. Frank, and R. Kahmann.** 1984. Gin-mediated site-specific recombination in bacteriophage Mu DNA: overproduction of the protein and inversion *in vitro*. EMBO J. **3:**2415–2421.
30. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
31. **Nagaraja, V., G. Hecht, and S. Hattman.** 1988. The phage Mu "late" gene transcription activator, C, is a site-specific DNA binding protein. Biochem. Pharmacol. **37:**1809–1810.
32. **Pabo, C. O., and R. T. Sauer.** 1984. Protein-DNA recognition. Annu. Rev. Biochem. **53:**293–321.
33. **Plasterk, R. H. A., M. Vollering, A. Brinkman, and P. van de Putte.** 1984. Analysis of the methylation-regulated Mu *mom* transcript. Cell **36:**189–196.
34. **Priess, H., B. Bräuer, C. Schmidt, and D. Kamp.** 1987. Sequence of the left end of Mu, p. 277–296. *In* N. Symonds, A. Toussaint, P. van de Putte, and M. M. Howe (ed.), Phage Mu. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y..
35. **Raibaud, O., and M. Schwartz.** 1984. Positive control of transcription initiation in bacteria. Annu. Rev. Genet. **18:**173–206.
36. **Reznikoff, W. S., and W. R. McClure.** 1986. *E. coli* promoters, p. 1–33. *In* W. S. Reznikoff and L. Gold (ed.), Maximizing gene expression. Butterworths, Boston.
37. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
38. **Schauder, B., H. Blöcker, R. Frank, and J. E. G. McCarthy.** 1987. Inducible expression vectors incorporating the *Escherichia coli atpE* translational initiation region. Gene **52:**279–283.
39. **Seiler, A., H. Blöcker, R. Frank, and R. Kahmann.** 1986. The *mom* gene of bacteriophage Mu: the mechanism of methylation-dependent expression. EMBO J. **5:**2719–2728.
40. **Silhavy, T. J., M. L. Berman, and L. W. Enquist.** 1984. Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
41. **Swinton, D., S. Hattman, P. F. Crain, C.-S. Cheng, D. L. Smith, and J. A. McCloskey.** 1983. Purification and characterization of the unusual deoxynucleotide α-*N*-(9-β-D-2′-deoxyribofuranosylpurin-6-yl)glycinamide, specified by the Mu modification function. Proc. Natl. Acad. Sci. USA **80:**7400–7404.
42. **Szatmari, G. B., M. Lapointe, and M. S. DuBow.** 1987. The right end of transposable bacteriophage D108 contains a 520 base pair protein-encoding sequence not present in bacteriophage Mu. Nucleic Acids Res. **15:**6691–6704.
43. **Toussaint, A.** 1976. The DNA modification function of temperate phage Mu-1. Virology **70:**17–27.
44. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19:**259–268.
45. **Weber, K., and M. Osborn.** 1969. The reliability of molecular weight determinations by dodecyl sulfate-polyacrylamide gel electrophoresis. J. Biol. Chem. **244:**4406–4412.
46. **Wulczyn, F. G., and R. Kahmann.** 1987. Post-transcriptional regulation of the bacteriophage Mu *mom* gene by the *com* gene product. Gene **51:**139–147.
47. **Wyckoff, E., L. Sampson, M. Hayden, R. Parr, W. M. Huang, and S. Casjens.** 1986. Plasmid vectors useful in the study of translation initiation signals. Gene **43:**281–286.