

---

**Localization and DNA sequence analysis of the *C* gene of bacteriophage Mu, the positive regulator of Mu late transcription**

---

William Margolin and Martha M. Howe

---

Department of Bacteriology, University of Wisconsin, Madison, WI 53706, USA

---

Received 13 March 1986; Accepted 19 May 1986

---

**ABSTRACT**

The *C* gene of bacteriophage Mu, required for transcription of the phage late genes, was localized by construction and analysis of a series of deleted derivatives of pKN50, a plasmid containing a 9.4 kb Mu DNA fragment which complements Mu *C* amber mutant phages for growth. One such deleted derivative, pWM10, containing only 0.5 kb of Mu DNA, complements *C* amber phages and transactivates the *mom* gene, one of the Mu late genes dependent on *C* for activation. The DNA sequence of the 0.5 kb fragment predicts a single long open reading frame coding for a 140 amino acid protein. Sequence analysis of DNA containing a *C* amber mutation located the base change to the second codon of this reading frame. Generation of a frameshift mutation by filling in a *Bgl*III site spanning codon 114 of this reading frame resulted in the loss of *C* complementation and transactivation activity. These results indicate that this open reading frame encodes the Mu *C* gene product. Comparison of the predicted amino acid sequence of the *C* protein with those of other transcriptional regulatory proteins revealed some similarity to a region highly conserved among bacterial sigma factors.

**INTRODUCTION**

Lytic growth of bacteriophage Mu, a temperate phage of *Escherichia coli* K12 and other enteric bacteria (see (1) for a review), proceeds in an ordered sequence which correlates with transcription of different segments of its 38 kb double stranded DNA genome (2). Early in the lytic cycle, transcripts are derived predominantly from the leftmost 9-10 kb of the genome due to the activity of a rightward-oriented promoter, *Pe*, located approximately 1 kb from the left end of the Mu genome (3,4). Two early genes essential for Mu DNA replication, *A* and *B*, (2,5,6) are expressed from this major early rightward promoter. Also transcribed from *Pe* is *ner* (2), whose gene product represses transcription from *Pe* at the end of the early phase (6,7). The onset of replication is followed by transcription of a region which includes the *C* gene at approximately 10 kb (3) and lies between the regions expressed early and late during phage development.

Transcription of the "late" region of the phage DNA (2) requires the *C*

gene product, since no late transcription is observed in a C amber mutant (3,8). The late region includes lys, which is essential for cell lysis (9), the genes involved in phage morphogenesis and host range determination (10-12), and mom (13,14), a gene required for Mu DNA modification (15,16).

Our objective is to understand how C activates the transcription of mom and the other phage late genes. There are many different strategies which Mu might use to positively regulate transcription (for a review, see (17)). For example, C could be a phage-encoded sigma factor, a CAP-like activator protein, an antiterminator, or a modifier of the host RNA polymerase. In this paper, we begin our studies to define the mechanism by which C positively regulates Mu late transcription by localizing and sequencing the C gene and determining the predicted amino acid sequence of the C protein. We also present additional evidence supporting the conclusion that Dad, the phage factor required for mom expression, is the C gene product (18).

### MATERIALS AND METHODS

#### Bacterial Strains, Phages, and Plasmids

All strains used were derivatives of Escherichia coli K12. Strain MH4985 (WD5021; gal lac rpsL; 19) and its recA derivative MH2500 (20) were used for marker rescue and complementation assays, respectively, and as hosts for isolation of the pWM plasmids. The strain containing pKN50 was MH3337, a derivative of EB105 ( $F^-(\Delta lac)RV \Delta(ara-leu)498$ ; 21). Strain MH7295, an MH2500 derivative containing  $F'lacI^{Q1} lacZ::Tn5$  (22) (obtained from C. Gross), was the parent of strains used for assaying complementation of Mu C amber phages by  $\lambda Q$ . Strain MH7388 was used for preparation of pWM3 from a dam<sup>-</sup> background: it is a pWM3-containing derivative of NK5772 (dcm6 dam3 metB1 galK2 galT22 lacY1 tsx78 thil1 tonA31 mtl-1; 23). Amber mutant phages were titered on either MH4939 ( $Q1$ ;  $F^- supE thr leu tonA lac$ ) or MH4945 ( $QD5003$ ;  $F^+ supF mel pro$ ) (11). Strains MH429 (thyA malt::Mu cts62) and MH3565 (thyA malt::Mu cts62 Cam1966) (11,12) were used as sources of RNA for S1 mapping experiments.

Mu amber mutant phages were from the collection of O'Day et al. (24). Strain MH3105, a derivative of MH4939 containing Mu cts62 Cam2005, and strains MH2853, MH408, MH1667, and MH3107, derivatives of MH4945 containing Mu cts62 phages with amber mutations Cam1966, Cam4005, lysam1030, and lysam7297, respectively, were used to prepare high titer Mu lysates (20).  $\lambda C1857 Qam117$  (25) was obtained from G. Somasekhar.

The plasmid pKN50 contains a 9.4 kb fragment of Mu DNA from 5.1 to 14.5 kb

---

cloned into pBR322 (21). Plasmid pCCV(-) (18) is a pACYC184 derivative containing the 5' region of the Mu mom gene translationally fused to lacZ. Plasmid pLQ323 is a pBR322 derivative containing the  $\lambda$ Q gene expressed from the lac operon promoter (26; G. Somasekhar and W. Szybalski, personal communication).

#### Media

LB broth and plates, soft agar, LBM, and SM were previously described (19). TCMG plates were as specified by Schumm et al. (27). Ampicillin and tetracycline were used in LB broth and agar (LB-Amp and LB-Tet) at 40 and 20  $\mu$ g/ml, respectively. LB-Thy broth was LB broth supplemented with 50  $\mu$ g/ml thymine.

#### Plasmid DNA

Plasmid minipreps were made according to Birnboim and Doly (28). Large-scale plasmid DNA preparations for DNA sequence analysis and construction of deletions were made as previously described (29) except that LB medium was used instead of M9. Plasmid transformations were performed using the RbCl-dimethylsulfoxide method of Kushner (30). Restriction endonucleases and T4 DNA ligase were obtained from New England Biolabs and were used according to the manufacturers' suggested conditions. Agarose gel electrophoresis was performed essentially as described by Maniatis et al. (31).

#### Construction of Deleted Plasmid Derivatives

In general, plasmid DNAs were cut with restriction enzyme(s) having sites at the desired deletion endpoints, ligated, and then incubated with a restriction enzyme expected to cut only within the deleted segment. Ligations and fill-in reactions using Klenow fragment of DNA polymerase I were done as described by Maniatis et al. (31). The mixture was then transformed into MH2500, selecting for ampicillin resistant ( $\text{Amp}^R$ ) transformants on LB-Amp plates. The resulting plasmid-containing cells were screened for their "Lys" phenotype directly on the transformation plates; gummy colonies presumed to be Lys<sup>+</sup> were discarded when deletions removing lys were desired. The cells were then tested in complementation assays (see below) and their plasmid DNAs were restricted and analyzed by gel electrophoresis. Plasmid pWM1 was made by cutting pKN50 with NruI, and cutting with BamHI after ligation; the NruI site was restored by ligation. To make pWM2, pKN50 was cut with BglII and SalI, both sites were filled in by treatment with DNA polymerase Klenow fragment and dNTP's (restoring the BglII site), and the DNA was cut with BamHI after ligation. Plasmid pWM3 was made by cutting pKN50 with BglII followed by filling in and ligating as above, a procedure expected to create

a ClaI site; digestion of pWM3 DNA prepared from the dam<sup>-</sup> strain MH7388 confirmed the presence of the ClaI site. To make pWM5, pKN50 was digested with HpaI and EcoRI, the sites were filled in (destroying both sites), and after ligation the DNA was cut with ClaI. To generate pWM6, pKN50 was cut with PvuI and SalI, treated with Klenow polymerase to fill in the SalI site and to create a blunt end at the PvuI site, ligated (resulting in restoration of the SalI site), and then cut with BamHI. Plasmid pWM8 was made by using the same procedures as for pWM6, but with pWM5 as the starting DNA. Plasmid pWM10 was made from pWM8 by deleting DNA between the XmnI site in the Mu fragment and the XmnI site at position 2030 of pBR322, and pWM11 was made from pWM8 by deleting DNA between the BglII site and the pBR322 SalI site using the same procedures as for generation of pWM2. Plasmid pWM12, containing the Cam1966 mutation, was made by isolating phage DNA (32) from particles produced by heat induction of MH2853, cleaving the DNA with EcoRI and BamHI, ligating the fragments into pBR322 DNA cut with EcoRI and BamHI, transforming into MH4945 selecting for Amp<sup>R</sup>, and screening transformants for tetracycline sensitivity (Tet<sup>S</sup>). Plasmid DNAs from several Amp<sup>R</sup> Tet<sup>S</sup> transformants were restricted and analyzed by gel electrophoresis; a transformant whose restriction pattern was identical to that of pKN50 was saved. The presence of the Cam1966 allele was verified by the ability of pWM12 to complement Cam2005 in MH4945 (Cam2005, unlike Cam1966, is not suppressed by SuIII) and by complementation of lys ambers but not C ambers in MH2500 (Su<sup>-</sup>).

#### Complementation and Marker Rescue Assays

Phage lysates were prepared by heat induction of the appropriate lysogen and titered as described (20). Each plasmid to be tested was transformed into MH2500 (recA) for complementation tests and MH4985 (rec<sup>+</sup>) for marker rescue assays. The plasmid-containing cells were grown overnight in LB-Amp, diluted into LBM, grown to a cell density of  $\sim 10^9$  cells/ml, and plated in soft agar on LB plates. Approximately 30 to 50  $\mu$ l of each amber mutant phage lysate was spotted on the cell lawns at  $\sim 10^9$ ,  $10^7$ ,  $10^5$ , and  $10^3$  plaque forming units (PFU)/ml for complementation, and  $\sim 10^9$ ,  $10^8$ ,  $10^7$ , and  $10^6$  PFU/ml for marker rescue, and the plates were incubated at 37°C overnight. The efficiency of plating (EOP) was calculated as the number of plaques observed on the plasmid-containing cell lawn divided by the number observed on an Su<sup>+</sup> control strain. Strains containing pKN50 and pBR322 served as positive and negative controls, respectively.

#### Lambda Q - Mu C Complementation

Derivatives of MH7295 containing pKN50, pWM2, and pLQ323 were grown to  $10^9$

cells/ml in LBM, incubated at 42°C for 20 min, and plated on LB agar with and without 1mM IPTG (isopropyl  $\beta$ -D-thiogalactopyranoside, Sigma Chemical Co.). Mu cts62 Cam1966, Mu cts62 Cam2005, Mu cts62, and  $\lambda$ Qam117 were spotted as described for the complementation protocol above, and scored for growth after overnight incubation at 37°C. The EOPs in the presence of IPTG were less than  $10^{-7}$  for both C amber phages on strains with pLQ323 and less than  $10^{-4}$  for the  $\lambda$ Q amber on strains with pWM2 and pKN50; whereas, the EOPs were approximately 1 for C ambers on pKN50-containing cells and for the  $\lambda$ Q amber on the strain with pLQ323. The EOPs in the absence of IPTG were similar, except that the EOP was 0.1 for the  $\lambda$ Q amber on the strain with pLQ323. Complementation between two phages was assayed by adsorption of one phage at  $\sim 10^9$  PFU/ml to MH2500 (grown in LBM to  $\sim 10^9$  cells/ml), plating in soft agar on TCMG plates, and spotting the second phage on the cell lawns at  $\sim 10^4$  to  $10^5$  and  $\sim 10^6$  to  $10^7$  PFU/ml as described above. Plaques were scored after overnight growth at 37°C. No plaques were observed when  $\lambda$ I,  $\lambda$ R, and  $\lambda$ Q amber mutant phages were coinfecting pairwise with Mu C and P amber mutant phages at the highest phage concentration tested, whereas control coinfections ( $\lambda$ I+ $\lambda$ R,  $\lambda$ I+ $\lambda$ Q,  $\lambda$ Q+ $\lambda$ R, MuC+MuP) all yielded at least 200 plaques per spot.

#### Mom Transactivation Assay

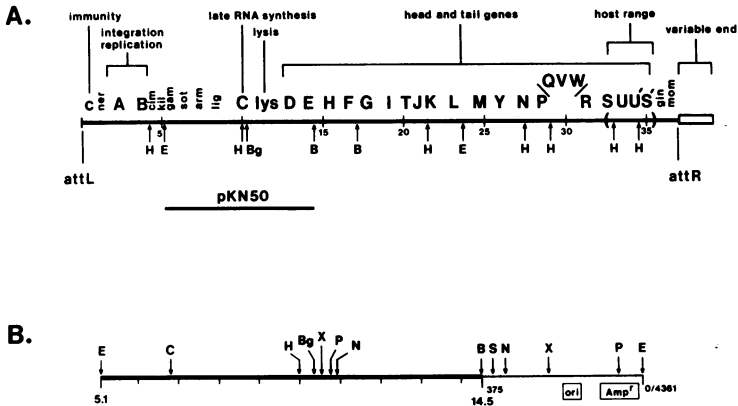
Competent MH4985 cells containing either pKN50 or one of the pWM plasmids were transformed with pCCV(-), and plated on LB-Amp-Tet plates supplemented with 20  $\mu$ g/ml XG (5-bromo-4-chloro-3-indolyl  $\beta$ -D-galactopyranoside, Sigma Chemical Co.). After 24-48 hours, the transformants were scored as being either blue or white. Quantitative assays for  $\beta$ -galactosidase activity on cultures of purified Amp<sup>R</sup> Tet<sup>R</sup> strains were performed according to Miller (33); portions of the same cultures were tested and found to have the expected C amber complementation and marker rescue properties.

#### DNA Sequencing

DNA sequences were determined using the method of Maxam and Gilbert (34,35). Cleaved plasmid DNA was labeled at the 5' end with [ $\gamma$ -<sup>32</sup>P]-ATP using T4 polynucleotide kinase (P-L Biochemicals) after treatment with calf intestinal alkaline phosphatase (Boehringer-Mannheim), or labeled at the 3' end with [ $\alpha$ -<sup>32</sup>P]-GTP using Klenow DNA polymerase. After cleavage with a second restriction enzyme and electrophoresis on 5% polyacrylamide gels, the labeled fragments to be sequenced were isolated from the gel slices by electroelution into dialysis bags (31).

#### S1 Mapping

As sources of in vivo Mu mRNA, cultures of MH429 and MH3565 (Mu C<sup>+</sup> and C



**Figure 1** (A). Genetic and physical map of bacteriophage Mu DNA. The bold horizontal line represents Mu DNA; the box at the right end and the vertical line at the left end denote covalently attached variable host DNA sequences. The parentheses near the right end represent the invertible G segment. Upper case letters above the Mu DNA (and *lys* and *lig*) indicate genes essential for Mu growth; lower case letters (except *lys* and *lig*) denote nonessential genes and functions. The symbols attL and attR denote the left and right phage attachment sites, respectively. The functions of some of the genes are indicated. Upward arrows designate the locations of the restriction cleavage sites for the enzymes *Hpa*I (H), *Eco*RI (E), *Bgl*II (Bg), and *Bam*HI (B). Numbers represent distances from attL in kilobases. The line below the map indicates the *Eco*RI-*Bam*HI fragment of Mu DNA carried in pKN50.

(B). Restriction sites in pKN50 used for making deletions. The bold and thin lines represent Mu DNA and pBR322 DNA, respectively, in the linear form of pKN50 generated by cleavage at the unique *Eco*RI site. The Mu DNA coordinates (in kb) of the cloned DNA fragment are given below the map as are the standard coordinates (in bp) for pBR322. Tick marks represent 1 kb intervals within the Mu DNA. Above the map are indicated the restriction sites used to cleave pKN50 to generate the deleted plasmids. All restriction sites shown are unique in pKN50 except for *Xmn*I, for which the only sites shown are those used for making pWM10. There are four additional *Xmn*I sites within the Mu DNA portion (S.F. Stoddard, M.T. Soltis and M.M. Howe, unpublished observations) and one additional site within the vector portion (31). Abbreviations for restriction endonuclease sites are as follows: B: *Bam*HI; Bg: *Bgl*II; C: *Cla*I; E: *Eco*RI; H: *Hpa*I; N: *Nru*I; P: *Pvu*I; S: *Sal*I; X: *Xmn*I. "Amp<sup>R</sup>" and "ori" denote the  $\beta$ -lactamase gene and the replication origin of pBR322.

amber mutant lysogens, respectively) were grown at 30°C in LB-Thy to a density of  $\sim 4 \times 10^8$  cells/ml, then shifted to 42°C to heat induce the prophage. The cells were harvested 35 min after induction by pouring onto ice. Cell lysis and extraction of the RNA using hot phenol was done as described by Salser et al. (36). S1 nuclease mapping of the *C* gene transcript was performed essentially as described by Burton et al. (37) using a hybridization temperature of 49°C.

### Computer Programs

Sequence alignment and homology searches were done using the GAP program of the University of Wisconsin Genetics Computer Group (UWGCG) (38), with gap weight = 5 and length weight = 0.3. Analysis of the open reading frames within the Mu DNA sequences and the predicted proteins was performed with the DNA/Protein Sequence Analysis System of International Biotechnologies, Inc.

### RESULTS AND DISCUSSION

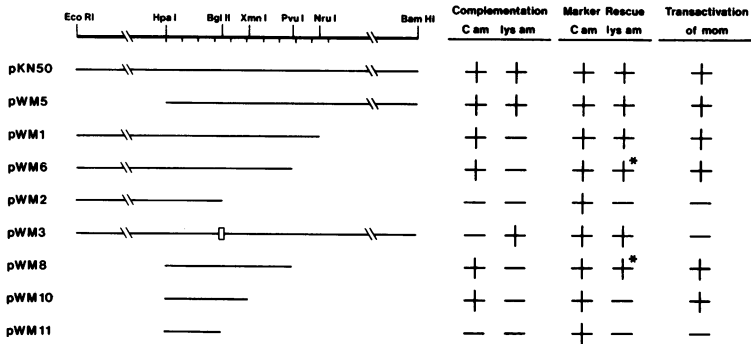
Previous analysis of  $\lambda$ pMu transducing phages carrying segments of the Mu genome had located the C gene to a region approximately 10 kb from the left end of Mu DNA, near a HpaI site (Figure 1A) (27,39). To locate C more precisely, we made a series of deletions in a plasmid expressing the gene, and assayed the deleted plasmids for their ability to complement or allow marker rescue of C amber mutant phages.

#### Expression of C and lys from pKN50

The starting plasmid was pKN50, which contains a 9.4 kb Mu DNA fragment from the EcoRI site at 5.1 kb to the BamHI site at 14.5 kb inserted between the EcoRI and BamHI sites of pBR322 (Figure 1A and B). Plasmid pKN50 expresses C, as evidenced by its ability to complement three different C amber mutant phages for growth in spot test assays (plating efficiencies of 0.3-1.0); these results confirm earlier observations of complementation of Cam4005 by pKN50 (40). Similar assays with two lys amber mutants indicated that pKN50 also expresses lys, a late gene to the right of C which requires C for its expression. Although these complementation assays were carried out in the presence of other phage functions, an additional property of the plasmid-containing cells suggested that some lys expression could occur in the absence of an infecting phage. Cells containing pKN50 flocculate in liquid cultures and form gummy colonies on plates, presumably due to lysis of some cells in the population. As discussed below, this "Lys" phenotype was not seen in Lys<sup>-</sup> plasmids, and could, in fact, be used to detect transformants containing deletions spanning the lys gene.

#### Precise Location of the C gene on pKN50

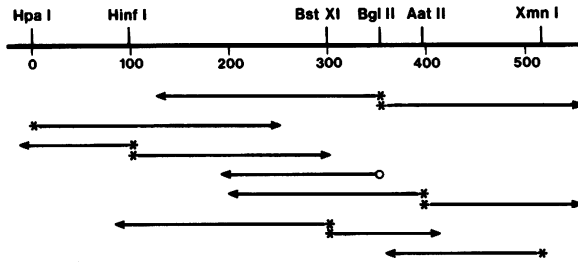
Deletions in pKN50 were first made to the left of the HpaI site and to the right of the nearby NruI, PvuI, and BglII sites to generate pWM5, pWM1, pWM6, and pWM2, respectively (Figure 2). As shown in Figure 2, all four plasmids allowed rescue of the three C amber mutants, and all except pWM2 could complement the C amber mutant phages for growth. This result indicated that the C gene was entirely within the 0.8 kb HpaI-PvuI fragment common to pWM5,



**Figure 2. Properties of the pKN50 deletion derivatives.** Horizontal lines below the restriction map denote the DNA present in plasmid pKN50 and its pWM derivatives identified at the left of the figure. Double slashes denote that several kb on either side of the *C* gene region has not been included in the figure so that this region could be enlarged and drawn to scale (tick marks at 0.1 kb intervals). The restriction sites shown represent unique sites in the Mu DNA fragment, except for *Xmn*I for which there are 4 additional sites (S.F. Stoddard, M.T. Soltis, and M.M. Howe, unpublished results). The box in pWM3 represents the four bp insertion obtained by filling in the *Bgl*II site of pKN50. The right portion of the figure indicates the results of complementation, marker rescue, and *mom* transactivation assays (see Materials and Methods). For all complementation results, a "+" denotes an plating efficiency (relative to a *Su*<sup>+</sup> host) of greater than 0.1, while a "-" denotes an efficiency of less than  $10^{-3}$ . For marker rescue, a "+" represents a plating efficiency of  $10^{-3}$  to  $10^{-5}$ , while a "-" denotes an efficiency of less than  $10^{-7}$ . All three *C* amber mutant phages behaved similarly in both types of assays. In addition to the data shown, complementation of *Cam*1966 by plasmids was analyzed in burst size experiments performed as previously described (20). Cells containing plasmids pWM5, pWM6, pWM8, and pKN50 yielded bursts of 114 PFU/cell or higher, while those containing pBR322 and pWM3 yielded burst sizes of 0.10 and 0.05 PFU/cell, respectively. The two *lys* amber phages, *lys*am1030 and *lys*am7297, behaved similarly in the complementation experiments, but in the marker rescue assays *lys*am1030 could rescue from pWM6 and pWM8 while *lys*am7297 could not (indicated by +\* in the figure). The last column lists the results of the *mom* transactivation assays measuring  $\beta$ -galactosidase production from the *mom-lacZ* fusion plasmid pCCV(-) (see Materials and Methods); a "+" indicates  $\beta$ -galactosidase activity of at least 47 units (33), while a "-" indicates activity of less than 0.59 units. Up to three-fold differences in enzyme activity were observed for replicate cultures of the same strain or different transformants containing the same plasmid. For pKN50 and pWM5, a few transformants gave considerably lower levels of  $\beta$ -galactosidase activity, presumably a consequence of lysis due to *lys* expression by these plasmids. Transformants of MH4985 (*Su*<sup>-</sup>) containing pCCV(-) and the *Cam*1966 plasmid pWM12 produced 0.83 units of  $\beta$ -galactosidase.

pWM1, and pWM6. Complementation and marker rescue results with *lys* mutants suggested that the *lys* gene spans both the *Pvu*I and *Nru*I sites, since pWM6 and pWM1 allowed rescue of some *lys* amber mutants, but both plasmids failed to complement the *lys* amber mutants for growth (Figure 2).





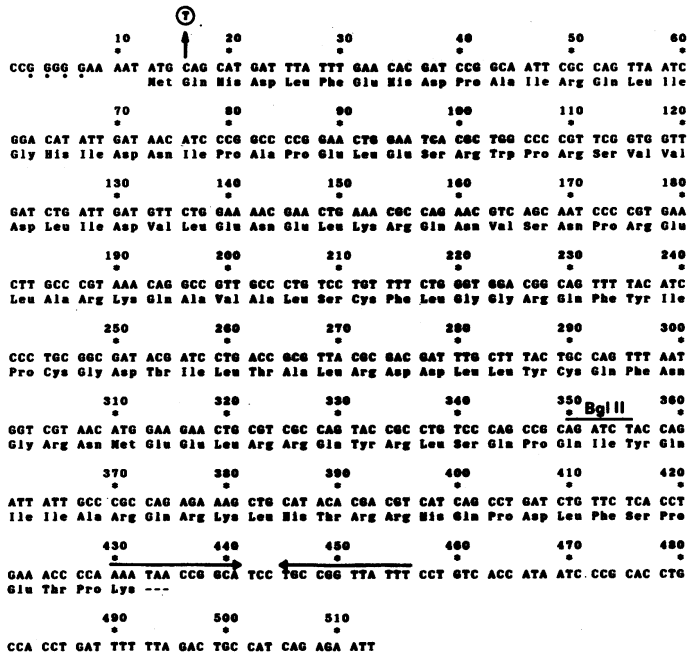
**Figure 3.** Sequencing strategy for pKN50 and pWM8. The thick line at the top denotes the Mu DNA in the region of interest with distances in base pairs from the HpaI site given below the line. The restriction sites used for sequencing are indicated above the line. The arrows represent the extent of DNA sequence determined from a given fragment using the Maxam-Gilbert technique (34,35). Asterisks and open circles denote positions of 5' and 3' end labels, respectively.

The inability of pWM2 to complement C amber mutants suggested that the BglII site was within the C gene. Further support for this possibility was provided by the properties of pWM3, a derivative of pKN50 containing a four base-pair insertion at the BglII site made by filling in the single stranded ends of BglII-cleaved DNA and religating to create a ClaI site. Figure 2 shows that pWM3 could not complement C mutants, but did complement lys mutants. Presumably, the four base-pair insertion caused a frameshift mutation in C. Complementation for lys could occur because the infecting lys amber phage would produce C product, which then would transactivate the lys gene on pWM3, resulting in lys expression. By this logic, pWM3 should not make Lys protein in the absence of the phage. As expected, it did not display the gummy colony phenotype we have attributed to lys expression. In contrast, pWM5, which contains intact C and lys genes, did display this "Lys" phenotype.

To confirm and further define the location of the C gene, three additional plasmids, containing the 0.8kb HpaI-PvuI fragment (pWM8), the 0.5 kb HpaI-XmnI fragment (pWM10), and the 0.35 kb HpaI-BglII fragment (pWM11) were constructed and characterized (Figure 2). Complementation of C ambers by pWM8 confirmed that the C gene is contained in the 0.8 kb fragment, and complementation by pWM10 narrowed its location to the 0.5 kb HpaI-XmnI fragment. The lack of complementation but positive marker rescue for all three C amber mutant phages by the 0.35 kb HpaI-BglII Mu DNA fragment in pWM11 was consistent with the results suggesting that C spans the BglII site.

#### DNA Sequence of the C gene

To locate the C reading frame and derive the amino acid sequence of the C



**Figure 4. Nucleotide sequence of the 0.5 kb *HpaI-XmnI* fragment and the predicted amino acid sequence of C.** The nucleotide sequence of the upper strand of the DNA is given from 5' to 3', and is numbered from the 5' end. The base at position 1 is the 3' cytosine in the *HpaI* recognition site; *XmnI* cleaves immediately 3' to the T at position 513. The postulated C reading frame begins at the first ATG in this fragment, reading left to right, and the predicted amino acid sequence is shown below the DNA sequence. Bases complementary to the 16S rRNA (Shine-Dalgarno sequence) are indicated by dots (bases 2-7). The *BglII* site, spanning bases 350-355 within the reading frame, is indicated. The inverted repeats at the 3' end of the coding region are indicated with inverted arrows. The TAA translation stop codon for C is marked by dashes. The position of the *Cam1966* mutation, a C to T transition in the second codon, is indicated by the circled T.

protein, the DNA sequence of the 0.5 kb *HpaI-XmnI* fragment was determined using plasmids pKN50 and pWN8. As indicated in the sequencing strategy depicted in Figure 3, the sequence was determined for both strands from all regions of the fragment, including those containing restriction sites used for end labeling.

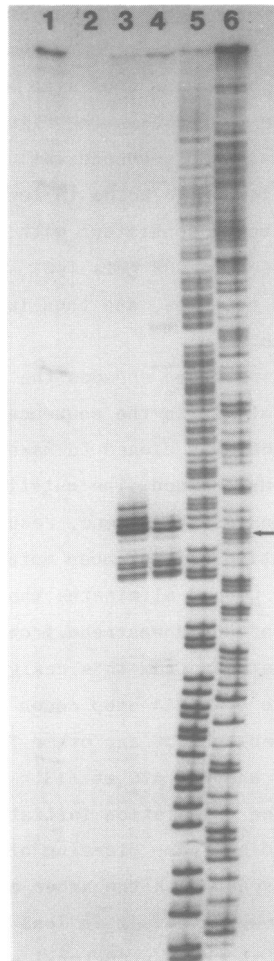
The nucleotide sequence of the 515 bp *HpaI-XmnI* fragment present in pWN10 and the predicted amino acid sequence of the C protein are presented in Figure 4. A search of all six possible frames revealed only one large open reading frame, entirely contained within the 515 bp fragment, which we attribute to C. All other open reading frames found within the fragment, starting

either with ATG or GTG, were less than 56 amino acids in length. The C reading frame starts with an ATG 15 bp downstream from the HpaI site (at +13 in the figure), spans the BglIII site, and ends with a TAA stop codon 437 bases downstream from the HpaI site (at +435 in the figure). The ATG is preceded by a relatively poor Shine-Dalgarno sequence (41), as shown in Figure 4. The deduced protein sequence is 140 amino acids in length with a calculated molecular weight of 16,500, a value consistent with the molecular weight of approximately 15,500 for C seen on SDS gels (42). The protein contains 25 basic residues and 18 acidic residues, and thus in the absence of modifications is likely to be basic.

To confirm that this reading frame encodes the C protein and that translation initiates at the ATG at +13 in the sequence, the EcoRI-BamHI fragment from a C amber mutant (Cam1966) was cloned to make pWM12, and the sequence of its HpaI-BglIII fragment was determined. The mutation was found to be a C to T transition in the second codon of the C gene, resulting in a TAG termination codon (Figure 4). This position of the amber mutation is consistent with the postulated reading frame for C, and eliminates the possibility that translation of C might initiate farther downstream from the ATG at +13. Examination of the DNA sequence upstream from this designated ATG reveals the presence within the HpaI site of a TAA stop codon which is in the same reading frame as C, and the absence of any other TTG, GTG, or ATG start codons between the HpaI site and the ATG at +13. Thus, all information indicates that this ATG is the translation initiation codon for the C protein. The fact that the reading frame starting at +13 and ending at +435 is the only reading frame that spans both the amber mutation and the BglIII frameshift mutation, both of which result in loss of C activity, is strong evidence that it corresponds to the C structural gene.

The cloned Mu DNA upstream of the C gene in pWM8 and pWM10 is not long enough to accommodate a promoter. Presumably, weak transcription from within pBR322, such as transcription initiated at the P4 promoter and reading through the 104 RNA terminator (43), allows sufficient C expression to complement C ambers and to transactivate Mu late transcription.

The 3' end of the gene and downstream sequences contain two inverted repeats which might form a large stable hairpin structure (Figure 4) with a calculated free energy of approximately -14 kcal/mol (44). This structure resembles a rho-independent transcription terminator (45). Results from S1 mapping experiments suggest that it serves as the C gene transcription terminator (Figure 5). Late RNA from induced C<sup>+</sup> and Cam1966 Mu lysogens



**Figure 5. S1 mapping of the 3' end of the C gene transcript.** The DNA probe was made by labeling at the BglIII site, which is within the C gene reading frame and is approximately 100bp upstream of the stem-loop structure predicted to be the 3' end of the C transcript. BglIII- cleaved pWM6 DNA was 3' labeled with Klenow DNA polymerase and [ $\alpha$ - $^{32}$ P]-ATP and then cleaved with SalI, which generated a 0.45 kb probe extending far to the right of the stem-loop structure. Approximately 0.1 picomole of DNA,  $\sim$ 50 $\mu$ g of Mu late RNA (see Materials and Methods), 20 $\mu$ g of tRNA, and 5000 units of S1 nuclease (30 minute units, Boehringer-Mannheim Biochemicals) were used in the S1 protection experiments which were performed as described by Burton et al. (37). Samples were run on 6% polyacrylamide gels containing 7M urea, and the gel was autoradiographed using Kodak XAR-5 film. Lane 1: Probe + tRNA. Lane 2: Probe + tRNA + S1 nuclease. Lane 3: Probe + tRNA + S1 + Cam1966 RNA. Lane 4: Probe + tRNA + S1 + C<sup>+</sup> RNA. Lanes 5 and 6 are A+G and C+T sequencing ladders from a 5' end labeled probe used to generate size markers; the arrow denotes a fragment 100 bases long.

---

protected a DNA probe 3' end-labeled at the BglIII site; since approximately 100 bases separate the BglIII site and the 3' end of the stem-loop structure, the S1-protected hybrid was expected to be about 100 bases in length. Figure 5 shows that for both the C<sup>+</sup> and C<sup>-</sup> phages, there are two predominant 3' ends, one located within the stem (93-95 bases from the labeled end, corresponding to positions 448-450 in Figure 4) and the other within or immediately downstream from the T residues at the 3' end of the stem (positions 456-458 in Figure 4, 101-103 bases from the labeled end). The small amount of protection of the full-length probe indicates that there is little apparent transcription readthrough beyond this site in both wild type and C amber phages, suggesting that C-dependent transcription of lys is not simply due to readthrough from the upstream C gene.

#### Transactivation of the Mu mom Gene

Recent evidence has suggested that the Mu Dad transactivation function required for transcriptional activation of the Mu mom gene promoter is in fact C (18). In that work, pWM8, containing 0.8 kb of Mu DNA, was shown to transactivate  $\beta$ -galactosidase expression from pCCV(-), a compatible plasmid containing a mom-lacZ fusion. In this work we have used pCCV(-) to assay for mom transactivation by several other pWM plasmids. For all plasmids tested, the ability to transactivate mom correlated with the ability to complement Mu C amber mutants for growth (Figure 2), a result which supports the identification of C as Dad. In particular, the transactivation ability of pWM10, containing the shorter 515 bp HpaI-XmnI fragment, strengthens the conclusion that C is the only phage protein required to transactivate mom.

#### Lack of Complementation with Lambda Q

Previously it was reported (8) that a  $\lambda$ Q<sup>+</sup> phage was able to complement a Mu C amber mutant phage for growth, while a  $\lambda$ Q<sup>-</sup> phage did not. Because this observation suggested that  $\lambda$ Q could replace C in activating Mu late transcription, we attempted to complement Mu C and  $\lambda$  Q amber mutant phages with both plasmids and phages expressing the other function. Plasmid pLQ323, which expresses  $\lambda$ Q from P-lac upon induction with IPTG (G. Somasekhar and W. Szybalski, personal communication), did not detectably complement Mu Cam1966 phage for growth, although it complemented  $\lambda$ Qam117 (see Materials and Methods). Similarly,  $\lambda$ Qam117 mutant phages were not detectably complemented for growth by pKN50 or pWM8, two of the C expressing plasmids described above. In coinfections of Mu Cam1966 with  $\lambda$ cI857 R or I late amber mutant (Q<sup>+</sup>) phage, or a Mu cts62 Pam7345 (C<sup>+</sup>) with  $\lambda$ Qam117 phage, there was no detectable complementation (see Materials and Methods). Therefore, we have

E. COLI SIGMA-70	L D L I q E G N I G L M K A V d K F E Y
E. COLI SIGMA-32	A D L I q E G N I G L M K A V R R F N P
BACILLUS SIGMA-29	E D L I s i G T I G L I K A V N T F N P
BACILLUS SIGMA-43	L D L I H E G N M G L M K A V E K F D Y
SALMONELLA NTR A	V E V D L D E V E A V L K R I q R F D P
MU C PROTEIN	v D L I . D V L E N E L K R q N V S N P
SP01 gp34	E D L M g T I D q I F L q L V S E Y N P
SP01 gp28	q D L E q E G L L K L Y D C W E K W C F
T4 gp55	q E M I A D G I E A S I K G L H N F D E

**Figure 6. Similarity between C and sigma factors.** Amino acids 36 to 54 of C are compared and aligned with the most highly conserved region of several sigma factors: amino acids 402-421 of sigma-70 (48), 76-95 of sigma-32 (49), 85-104 of sigma-29 (50), 161-180 of sigma-43 (47), 168-187 of ntrA (51), 44-63 of SP01 gp34 (52), 49-68 of SP01 gp28 (53), and 69-88 of T4 gp55 (54). A gap of one amino acid was introduced into the C sequence to achieve optimal alignment. Bacterial sigma factors are listed above the C sequence and phage sigma factors are given below it. The single letter amino acid code is used for ease of comparison. If there are at least four chemically similar amino acids at a given position among the eight sequences being compared with C, those residues are represented by large letters; otherwise, the residues are represented by small letters. Amino acids defined as chemically similar for this figure are K,R,H; D,E; Q,N; I,L,V,M; S,T; F,W,Y; A,G; C and P are not in a group (57; W. Fitch, personal communication).

been unable to obtain evidence suggesting that the  $\lambda$ Q and Mu C proteins can functionally replace each other in vivo.

Search for Proteins Containing Sequence Similarity to C

We have compared the amino acid sequence of C to those of other proteins which positively regulate transcription in an attempt to identify sequence similarities with these proteins that might provide clues to the role of the C protein. The UW Genetics Computer Group GAP program, which finds the best alignment between two sequences by inserting gaps, was used for this purpose. There were no significant similarities detected between Mu C and  $\lambda$  cII, N, and Q; several host and phage sigma factors (see Figure 6); N of phages P22 and 21; ogr of phage P2; and CAP and AraC of *E. coli*. However, visual comparison of the C protein sequence with the most highly conserved regions among known bacterial and phage sigma factors (46-54; M. Gribskov and R.R. Burgess, personal communication) revealed some similarity to C in the region from amino acids 36 to 54 (Figure 6). Seven of the 19 amino acid positions in this small region of C are chemically similar to positions 402-421 of the major *E. coli* sigma-70, while 11 of 19 positions are chemically similar to, and 7 of 19 are identical to positions 168-187 of the ntrA sigma factor involved in nitrogen regulation in *Salmonella typhimurium* (54,55; S. Kustu,

personal communication). Although there is no obvious sequence similarity between C and other regions in these sigma factors, C shares with these proteins a clustering of acidic amino acids near the N terminus and basic residues near the C terminus. Some similarity to the helix-turn-helix domain often present in DNA binding proteins (56) was also found in the N terminus, between residues 17 and 27.

The observation that a small region of C has some sequence similarity to a highly conserved and presumably functionally important region of sigma factors is intriguing, especially in view of the relatively weak sequence homology between the bacterial sigma factors and phage sigma factors. Clearly, this observation in no way provides evidence that C is a sigma factor; however, it does suggest an experimental direction for investigation into the mechanism of C activity.

#### Conclusions

In summary, we have located and sequenced the C gene of bacteriophage Mu and derived the amino acid sequence of the C protein. The positions of the Cam1966 mutation, the frameshift mutation at the BglIII site in pWM3, and the 3' end of the C gene transcript, the reported molecular weight of the protein, and all the complementation, marker rescue, and mom transactivation results are consistent with our postulated reading frame for C. We are now undertaking the biochemical and genetic experiments required to define the mechanism of action of the C protein in stimulating Mu late transcription and to identify the sites at which it acts.

#### ACKNOWLEDGEMENTS

We thank S. Hattman and W. Szybalski for the plasmids pCCV(-) and pIQ323, M. Gribskov for unpublished sequence homology data, J. Erickson and W. Fitch for help with the sequence homology searches, and S. Kustu and C. Gross for helpful discussions. This work was supported by the College of Agricultural and Life Sciences, University of Wisconsin-Madison, by PHS grant No. AI-12731 and NSF grants PCM79-04055 and PCM81-09769 to M.M.H.; W.M. was supported by an NSF predoctoral fellowship. M.M.H. holds a University of Wisconsin Vilas Research Professorship.

#### REFERENCES

1. Toussaint, A., and Résibois, A. (1983) in *Mobile Genetic Elements*, J.A. Shapiro, Ed., pp. 105-158, Academic Press, New York.
2. Wijffelman, C., Gassler, M., Stevens, W.F., and van de Putte, P. (1974) *Mol. Gen. Genet.* 131, 85-96.
3. Marrs, C.F. (1982) Ph.D. Thesis, University of Wisconsin-Madison.
4. Krause, H.M., Rothwell, M.R., and Higgins, N.P. (1983) *Nucl. Acids Res.* 16, 5483-5495.
5. Wijffelman, C.A., and Lotterman, B. (1977) *Mol. Gen. Genet.* 151, 169-174.

6. van de Putte, P., Giphart-Gassler, M., Goosen, N., Goosen, T., and van Leerdam, E. (1980) *Cold Spring Harbor Symp. Quant. Biol.* 45, 347-353.
7. van Leerdam, E., Karreman, C., and van de Putte, P. (1982) *Virology* 123, 19-28.
8. van Meeteren, R. (1980) Ph.D. Thesis, Univ. of Leiden, The Netherlands.
9. Faelen, M., and Toussaint, A. (1973) *Virology* 54, 117-124.
10. van de Putte, P., Cramer, S., and Giphart-Gassler, M. (1980) *Nature* 280, 218-222.
11. Grundy, F.J., and Howe, M.M. (1984) *Virology* 134, 296-317.
12. Grundy, F.J., and Howe, M.M. (1985) *Virology* 143, 485-504.
13. Toussaint, A., Desmet, L., and Faelen, M. (1980) *Mol. Gen. Genet.* 177 351-353.
14. Plasterk, R.H.A., Vrieling, H., and van de Putte, P. (1983) *Nature* 301, 344-347.
15. Toussaint, A. (1976) *Virology* 70, 17-27.
16. Hattman, S. (1980) *J. Virol.* 34, 277-279.
17. Raibaud, O., and Schwartz, M. (1984) *Ann. Rev. Genet.* 18, 173-206.
18. Hattman, S., Ives, J., Margolin, W., and Howe, M.M. (1985) *Gene* 39, 71-76.
19. Howe, M. (1973) *Virology* 54, 93-101.
20. Howe, M.M., O'Day, K.J., and Schultz, D.W. (1979) *Virology* 93, 303-319.
21. Schumann, W., Bade, E.G., Forgie, R.A., and Howe, M.M. (1980) *Virology* 104, 418-425.
22. Muller-Hill, B., Fanning, T., Geisler, N., Gho, D., Kania, J., Kathmann, P., Meissner, H., Schlotmann, M., Schmitz, A., Triesch, I., and Beyreuther, K. (1975) in *Protein-Ligand Interactions* (ed. H. Sund and G. Blauer), pp.211- 227, Walter de Gruyter, Berlin.
23. Simons, R.W., Hoopes, B.C., McClure, W.R., and Kleckner, N. (1983) *Cell* 34, 673-682.
24. O'Day, K., Schultz, D., Ericson, W., Rawluk, L., and Howe, M.M. (1979) *Virology* 93, 320-328.
25. Campbell, A. (1961) *Virology* 14, 22-32.
26. Somasekhar, G., and Szybalski, W. (1983) *Gene* 26, 291-294.
27. Schumm, J.W., Moore, D.D., Blattner, F.R., and Howe, M.M. (1980) *Virology* 105, 185-195.
28. Birnboim, H. C., and Doly, J. (1979) *Nucl. Acids Res.* 7, 1513-1523.
29. Marrs, C.F., and Howe, M.M. (1983) *Virology* 126, 563-575.
30. Kushner, S. R. (1978) in *Genetic Engineering*, H.B. Boyer and S. Nicosia Eds., pp. 17-23, Elsevier/North-Holland, Amsterdam.
31. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) *Molecular Cloning*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
32. Williams, B.G., Blattner, F.R., Jaskunas, S. R., and Nomura, M. (1977) *J. Biol. Chem.* 252, 7344-7354.
33. Miller, J. (1972) *Experiments in Molecular Genetics*, pp. 352-355, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
34. Maxam, A.M., and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* 74, 560-564.
35. Maxam, A.M., and Gilbert, W. (1980) *Methods Enzymol.* 65, 499-559.
36. Salser, W., Gesteland, R.F., and Bolle, A. (1967) *Nature* 215, 588-591.
37. Burton, Z.F., Gross, C.A., Watanabe, K.K., and Burgess, R.R. (1983) *Cell* 32, 335-349.



- 
38. Devereaux, J., Haeberli, P., and Smithies, O. (1984) *Nucl. Acids Res.* 12, 387-395.
  39. Magazin, M., Howe, M., and Allet, B. (1977) *Virology* 77, 677-688.
  40. Staufenbiel, M., and Schumann, W. (1983) *FEMS Microbiol. Lett.* 20, 369-374.
  41. Shine, J., and Dalgarno, L. (1975) *Nature* 254, 34-38.
  42. Giphart-Gassler, M., Wijffelman, C., and Reeve, J.M. (1981) *J. Mol. Biol.* 145, 139-163.
  43. Stuber, D., and Bujard, H. (1981) *Proc. Natl. Acad. Sci. USA* 78, 167-171.
  44. Tinoco, I., Uhlenbeck O. C., and Levine, M. (1971) *Nature* 230, 362-367.
  45. Rosenberg, M., and Court, D. (1979) *Ann. Rev. Genet.* 13, 319-353.
  46. Stragier, P., Parsot, C., and Bouvier, J. (1985) *FEBS Letters* 187, 11-15.
  47. Gitt, M.A., Wang, L.F., and Doi, R.H. (1985) *J. Biol. Chem.* 260, 7178-7185.
  48. Burton, Z.F., Burgess, R.R., Lin, J., Moore, D., Holder, S., and Gross, C.A. (1981) *Nucl. Acids Res.* 9, 2889-2903.
  49. Landick, R., Vaughn, V., Lau, E. T., Van Bogelen, R. A., Erickson, J. W., and Neidhart, F. C. (1984) *Cell* 38, 175-182.
  50. Stragier, P., Bouvier, J., Bonamy, C., and Szulmajster, J. (1984) *Nature* 312, 376-377.
  51. Merrick, M.J., and Gibbins, J.R. (1985) *Nucl. Acids Res.* 13, 7607-7620.
  52. Costanzo, M., and Pero, J. (1983) *Proc. Natl. Acad. Sci. USA* 80, 1236-1240.
  53. Costanzo, M., Brzustowicz, L., Hannett, N., and Pero, J. (1984) *J. Mol. Biol.* 180, 533-547.
  54. Gram, H., and Ruger, W. (1985) *EMBO J.* 4, 257-264.
  55. Hirschman, J., Wong, P.K., Sei, K., Keener, J., and Kustu, S. (1985) *Proc. Natl. Acad. Sci. USA* 82, 7525-7529.
  56. Pabo, C.O., and Sauer, R.T. (1984) *Ann. Rev. Biochem.* 53, 293-321.
  57. Grantham, R. (1974) *Science* 185, 862-865.