

# Sequence-specific DNA Binding of the Phage Mu C Protein: Footprinting Analysis Reveals Altered DNA Conformation Upon Protein Binding

V. Ramesh and V. Nagaraja\*

Centre for Genetic  
Engineering, Indian Institute  
of Science, Bangalore  
560 012, India

The *mom* gene of bacteriophage Mu, which codes for a DNA modification function, is regulated in a complex manner at both transcriptional and translational levels. The phage-encoded C protein functions as an activator of *mom* transcription. The *mom* promoter has features of an activator-dependent weak promoter, and the C binding site is located upstream and overlapping the  $-35$  region and includes the palindromic sequence TTAT(N)<sub>6</sub>ATAA. The interactions of this activator protein at its binding site in  $P_{mom}$  has been investigated using four different chemical footprinting reagents. The protein footprint spans a region of 18 to 25 bp, depending on the nature of the chemical reagent used. Dimethylsulfate protection experiments revealed the base-specific interactions. The protected guanines are separated by 15 bp and are located beyond the interrupted palindromic sequence. A tripartite footprint was observed with hydroxyl radical, generated by Fe(II)-EDTA, which shows the binding of the protein to one face of the helix. The extent of protection conferred by the bound protein, however, is not uniform, suggesting that the interaction is asymmetric. The chemical nuclease 1,10-phenanthroline-copper, a minor groove specific ligand, shows hyper-reactivity upon protein binding in the top strand nucleotide triplet CAC, again confirming the protein-induced alterations in DNA conformation. Gel exclusion chromatography and chemical crosslinking experiment with the purified protein suggest that this mode of interaction is accomplished by a dimeric protein. This observation is supported by electrophoretic mobility shift assay using heterodimer of pure C protein and staphylococcal protein A-C fusion. The deletion analysis implicates a role for the carboxyl-terminal region of the protein in DNA binding.

© 1996 Academic Press Limited

\*Corresponding author

Keywords: Mu C protein; chemical footprinting; *mom* regulation

## Introduction

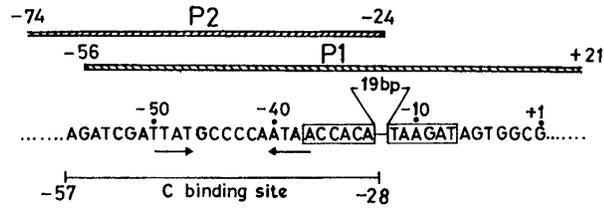
Regulation of gene expression is an intricately controlled process involving the DNA binding regulatory proteins and RNA polymerase (RNAP). Both positive and negative control circuits can operate at multiple steps to modulate the expression of a gene. The *mom* gene of bacteriophage Mu (Kahmann & Hattman, 1987) is a case that highlights this point very well. This gene, which

encodes a unique DNA modification function, is the last gene to be expressed in the lytic cycle of the phage. The *mom* specific modification of adenines in the sequence 5'...G/CAG/CNPy...3' to N<sup>6</sup>-acetamidoadenine serves an anti-restriction function for the phage, which has a broad host range. Though the biochemical pathway that leads to the DNA modification is not understood, it is known that *mom* is the only phage-encoded function involved in the process. Since this DNA modification not only affects restriction but also other phenomena like replication and transcription, it is essential that it occurs very late in the phage life cycle. Accordingly, the *mom* gene, which maps to the right extreme in the phage genome, is the last gene to be transcribed and is subjected to a complex and elaborate control mechanism both at the level

Abbreviations used: bp, base-pair(s); CRP, *E. coli* cyclic AMP receptor protein; DMS, dimethyl sulfate; ENU, ethylnitrosourea; HTH, helix-turn-helix; (OP)<sub>2</sub>Cu, 1,10-phenanthroline-copper; RNAP, RNA polymerase; PCR, polymerase chain reaction; IPTG, isopropyl- $\beta$ -D-thiogalactoside.

of transcription and translation to delay the *mom* expression until the final phase of Mu development. Transcription of the *mom* gene requires the methylation of adenines, by the *Escherichia coli* Dam methylase, in three GATC sites (located from positions -87 to -55) in the *mom* promoter region (Hattman, 1982). The methylation-dependent expression of this gene was the first example demonstrating the involvement of methylation in positive control. The negative control of *mom* regulation is mediated by a cellular protein OxyR. OxyR binds to a 43 bp region (-92 to -50) spanning the GATC sites, only when the adenines in these sequences are not methylated (Bolker & Kahmann, 1989). At the translation level, the phage Com protein encoded by the *mom* operon acts as a translational enhancer of *mom* mRNA. This zinc metallo-protein is an RNA binding protein specific for the translation initiation region of *mom* mRNA and stimulates translation by preventing/relieving an inhibitory secondary structure involving the initiation codon and the Shine-Dalgarno sequence (Hattman *et al.*, 1991; Wulczyn & Kahmann, 1991). Positive control of *mom* gene expression is brought about by the Mu C gene product, which is transcribed towards the middle phase of lytic cycle (Hattman *et al.*, 1985; Heisig & Kahmann, 1986; Margolin & Howe, 1986). The C protein is a site-specific DNA binding protein (Nagaraja *et al.*, 1988) and functions as an activator of late gene transcription from four late promoters  $P_{lys}$ ,  $P_I$ ,  $P_P$ , and  $P_{mom}$  (Bolker *et al.*, 1989; Margolin *et al.*, 1989). The transcription start sites from these promoters have been mapped. Sequence analysis of these promoters show poor similarity to the *E. coli* -35 region. These activation-dependent promoters, exhibiting characteristics of a weak promoter, are subjected to positive control by the transacting function C.

DNA binding studies with purified C protein (Ramesh *et al.*, 1994a,b) show that C binds to the *mom* promoter with an affinity comparable to that of  $\lambda$  cI and lac repressors at their respective binding sites. *In vitro* footprinting studies using MPE-Fe(II) located the C binding site in  $P_{mom}$  between -35 and -53, which includes a tetranucleotide inverted repeat sequence TTAT(N)<sub>6</sub>ATAA (Bolker *et al.*, 1989). Sequence alignment of the four phage late promoters revealed conservation of sequence upstream and overlapping the -35 region, which serves as the 3' half-site for C protein binding. The 5' half-site, however, is conserved to a lesser extent. Mutation analysis of the *lys* regulatory region showed that, in addition to the -10 region, a 19 bp region from -52 to -34 containing the C footprint is required for C-dependent promoter activity (Chiang & Howe, 1993). *In vivo* footprinting studies (Balke *et al.*, 1992) on  $P_{mom}$  revealed that, in the absence of C protein, *E. coli* RNAP binds to an upstream non-functional promoter P2. In the presence of C, polymerase binds to the downstream functional promoter P1 (which overlaps the C binding site) and is engaged in open complex



**Figure 1.** The *mom* regulatory region: the -10 and -35 positions are boxed. The dyad symmetric sequence within the C binding site and the overlapping promoters P2 and P1 are also indicated.

formation. This functional promoter shows the characteristics of an activator-dependent weak promoter: a poor -35 region (ACCACA) and a sub-optimal 19 bp spacer between the -10 and -35 sequences (Hawley & McClure, 1983; Raibaud & Schwartz, 1984; Harley & Reynolds, 1987). From this point of view, a likely mechanism for transcription activation, in addition to possible interactions with RNAP, is through structural alterations in DNA, mediated by the C protein, to facilitate RNAP recognition of the functional promoter P1 for productive open complex formation. To investigate this aspect, we have carried out footprinting analysis with purified C protein, using chemical footprinting agents. Our data indicate that, upon protein binding, the DNA in the recognition sequence undergoes a conformational change, which could have implications in the transcription activation process. Gel filtration chromatography, heterodimer DNA binding assay, protein crosslinking and deletion analysis further show that this mode of interaction is brought about by the binding of a dimeric protein involving the carboxyl-terminal region of the protein.

## Results

The C binding site in the regulatory region of  $P_{mom}$ , identified by using the chemical nuclease MPE-Fe(II) (Bolker *et al.*, 1989) and DNase I footprinting (Gindlesperger & Hattman, 1994), is shown in Figure 1. While the chemical method using an end-labeled fragment yielded an 18 bp protected region, enzymatic footprinting with DNase I performed on supercoiled DNA resulted in a 28 bp protection. These studies thus located the C binding site upstream and overlapping the -35-like sequence of  $P_{mom}$ . The difference in sizes of the protected regions, however, could be either due to the different topological substrates used (linear versus supercoiled DNA) or due to the nature of the footprinting probe (MPE-Fe(II) versus DNase I). Our results of DNase I footprinting (not shown), performed on a linear 260 bp end-labeled fragment containing the C binding site, yielded an identical pattern of protection to that obtained by Gindlesperger & Hattman (1994), suggesting that the topological status of DNA (supercoiled or linear) does not affect the span of the protected region. To

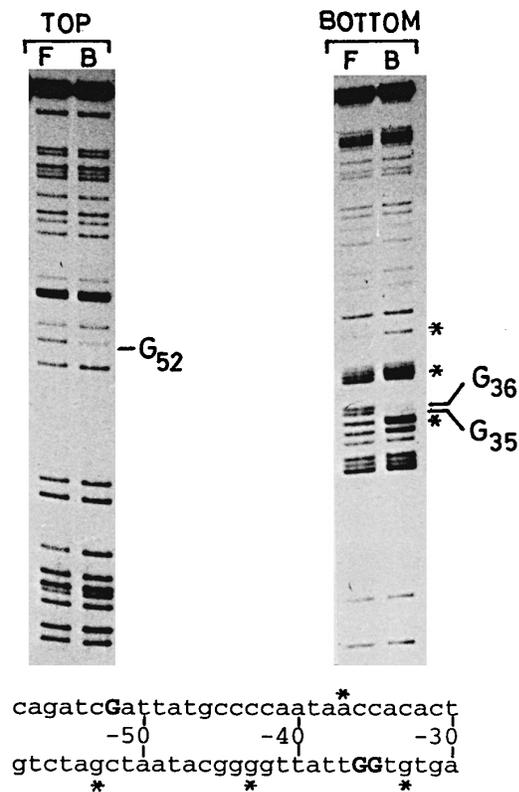
analyze further, the nucleotide contacts in this region, chemical modification protection and interference studies were carried out with dimethyl sulfate and ethylnitrosourea.

### Methylation protection analysis

The chemical dimethyl sulfate (DMS) methylates the N7 position of guanines and the N3 position of adenines, which are located in the major and minor grooves of DNA, respectively (Seeman *et al.*, 1976). Comparative analysis of the methylation of DNA in the presence and absence of protein yields data on the proximity of these functional groups to the protein in the complex. The proximity is usually interpreted as a contact with DNA at these sites, since there is a good agreement with the contacts obtained by co-crystal structure. The binding site of C in  $P_{mom}$  includes the palindromic sequence TTAT(N)<sub>6</sub>ATAA with the spacer being G + C-rich, and thus is amenable for probing with DMS. A preformed protein:DNA complex in the presence of saturating amounts of C protein was reacted with DMS, and the pattern of methylation was compared with that of naked DNA. The DNA was cleaved with piperidine to display the contacts with N7 group of guanines in the top and bottom strands (Figure 2). The pattern of DNA methylation by DMS under identical conditions in the absence of protein served as the control. Positions -52 G in the top strand and -36 G, -35 G in the bottom strand were protected against methylation to various extents, indicating the proximity of these functional groups to the protein. Since the N7 of guanine is a major groove determinant, the protection of these groups would mean a major groove interaction by the protein at these positions. The location of these protected guanines shows that the binding site extends beyond the interrupted palindromic sequence. On the other hand, -53 G, -43 G and -33 G in the bottom strand show enhanced methylation upon protein binding. Other residues (like -46 G in the top strand) also show enhanced reactivity but to a lesser extent and were not consistent. The cleavage conditions of the methylated bases were altered (A > G) to detect the N3 adenines protected against methylation. The pattern (not shown) did not reveal any significant protection of adenines in either of the strands. However, binding of the protein enhanced the methylation of -37 A on the top strand.

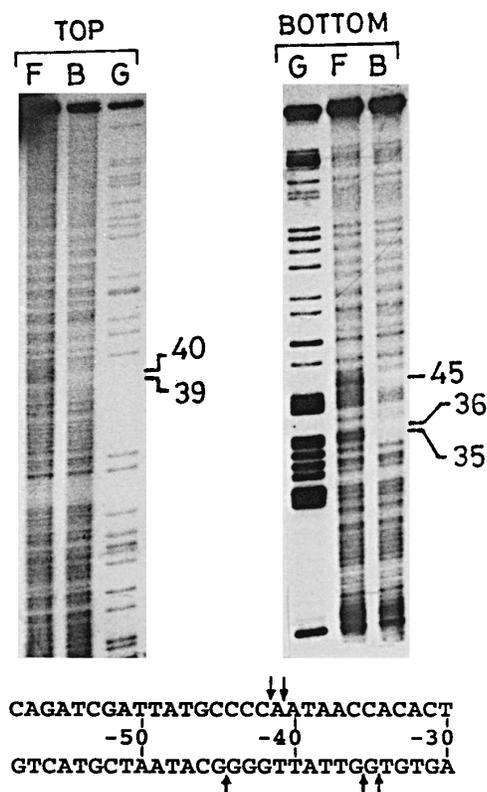
### Ethylation interference analysis

To identify the phosphate groups important in the C protein:DNA interactions, ethylation interference assays were carried out. The reagent ethylnitrosourea (ENU) preferentially ethylates phosphates in the backbone of DNA (Sun & Singer, 1975), and converts them into phosphotriesters. Strand scission of the ethylated DNA is achieved by



**Figure 2.** Guanine contacts of C protein obtained by DMS methylation protection. TOP and BOTTOM refer to the location of label in the corresponding strands of DNA. F and B indicate the methylation pattern in the absence and presence of C protein (500 ng). The protected guanines with the nucleotide positions are shown, and asterisks indicate the hyper-methylated guanines. The data are summarized in the nucleotide sequence below, in which the protected guanines are in bold uppercase, and the asterisks indicate the hyper-methylated purines.

alkali treatment in the subsequent step. Since the modification reaction does not support DNA binding (50% ethanol; 50°C; 60 minutes), only interference analysis can be carried out with this reagent. The 260 bp end-labeled DNA was partially ethylated, complexed with the protein, and the bound and unbound fractions (B and F, respectively) were resolved on a native gel by an electrophoretic mobility shift assay. The DNA was eluted from these regions, cleaved with alkali, and the footprinting pattern obtained is shown in Figure 3. At least two ethylated phosphate groups in the top strand at positions -40, -39 and three in the bottom strand at positions -45, -36, -35 interfered with protein binding. The interfering phosphates are 5' to the base positions indicated, and were determined by comparing with the 'G' sequencing ladders analyzed on adjacent lanes (Siebenlist & Gilbert, 1980). These positions map to the same or neighboring residues that were identified as being in contact by, methylation and other footprinting methods.



**Figure 3.** Phosphate contacts of C protein analyzed by its binding to partially ethylated DNA. The same labeling notation is used with "G" indicating the guanine specific sequencing ladder. The modified phosphates which interfere with binding are marked, and their locations are shown by vertical arrows in the nucleotide sequence below.

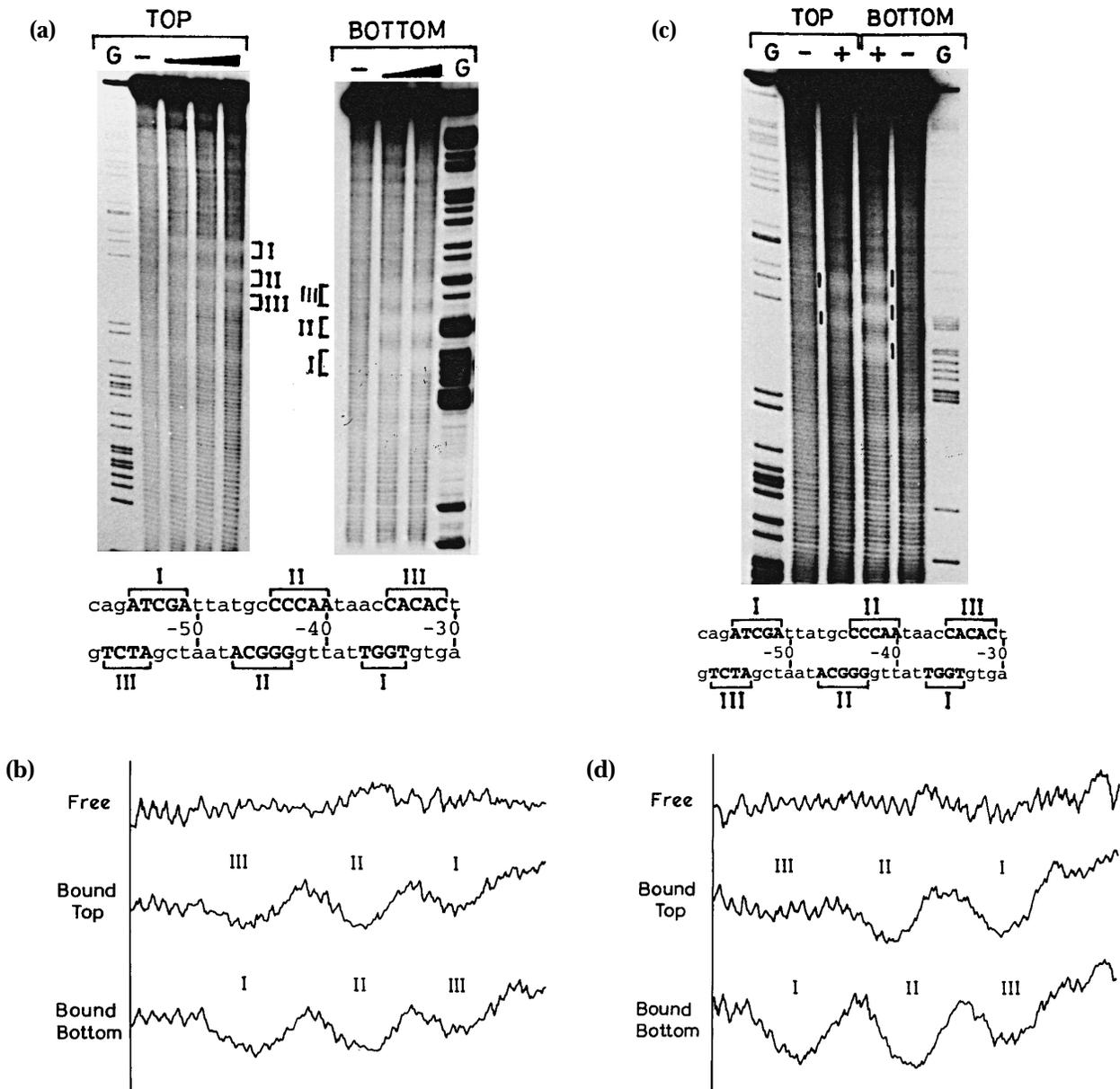
### Hydroxyl radical footprinting

Hydroxyl radicals generated by the Fenton reaction between Fe(II)-EDTA and hydrogen peroxide preferentially attack the C-4' of deoxyribose in the minor groove, leading to strand scission. Intimate contacts along the surface of the DNA are revealed by this powerful method (Tullius & Dombroski, 1986). Because of the small size and lack of sequence specificity of the hydroxyl radical, this method generates a high-resolution footprint. The hydroxyl radical footprinting was performed on C protein:DNA complex with increasing amounts of the protein. The reactants were individually tested for the integrity of the protein:DNA complex, and the reaction was performed under conditions of stable complex formation. Three regions of protection (I, II, III) were seen on each strand and the individual region covers four or five base-pairs (Figure 4(a)). The densitometric scan of the data is presented in Figure 4(b). The boundaries of the protected regions were accurately mapped with reference to the G sequencing ladders by scanning the autoradiograms and by comparing the cleavage pattern discerned from longer exposure of the gels. The protected regions in each strand are offset by 2 to

3 bp in the 3' direction, reflecting the cleavage chemistry of the hydroxyl radical that is centered in the minor groove of DNA. The centers of individual protected regions are separated by about 10 bp, indicating that C protein binds to one face of the helix. Also, the guanines protected against methylation and ethylated phosphates, which interfere with binding, map in the region of hydroxyl radical footprint. The protection pattern also reveals asymmetric interaction: the region III in the top strand spanning residues -35 to -31 is protected to a lesser extent compared with the other regions on either strands. This is clear in Figure 4(c), where the footprinting conditions are better optimized and the samples analyzed on adjacent lanes. The quantitative differences in footprint intensity are shown in Figure 4(d). The lower refractivity of region III is not due to band compression or other artefacts, as this pattern is consistently reproducible and further supported by using another footprinting reagent (see next section). This weaker protection probably reflects either less proximal contacts or alterations in DNA conformation to allow for the faster rate of free radical-mediated DNA backbone cleavage.

### (OP)<sub>2</sub>Cu footprinting

The results with Fe(II)-EDTA protection experiments indicate differential extents of protection, and region III in the top strand is more accessible to free radicals compared with other protected regions. To further probe this asymmetric mode of interaction, we used the chemical nuclease (OP)<sub>2</sub>Cu. This ligand binds to the minor groove of DNA and is sensitive to local conformational changes in DNA induced by the binding of the protein (Spassky & Sigman, 1985). The preformed C protein:DNA complex was subjected to cleavage with the reagent, and the pattern was compared with the naked DNA. Figure 5 shows the protection pattern on both the strands of DNA. The boundaries were mapped accurately by longer exposure of the gels and comparison with the sequencing ladders. The bottom strand protection spanned residues -58 to -34. The top strand protected region was from nucleotide position -57. The regions -33, -32 and -31 (C, A, C) showed enhanced reactivity on the top strand in the presence of protein, while the naked DNA is almost not cleaved in this region under the conditions employed. The enhanced cleavage by this groove specific ligand could probably indicate localized alterations (widening) of minor groove dimensions at, or near, the trinucleotide CAC. Such alternating purine-pyrimidine motifs are prone to DNA deformation (Travers, 1992). It is also possible that the hyper-reactivity could be simply due to the binding of the reagent to hydrophobic portions of the protein. However, the fact that hydroxyl radical generated by Fe(II)-EDTA, which is unlikely to interact in a similar manner, also shows increased



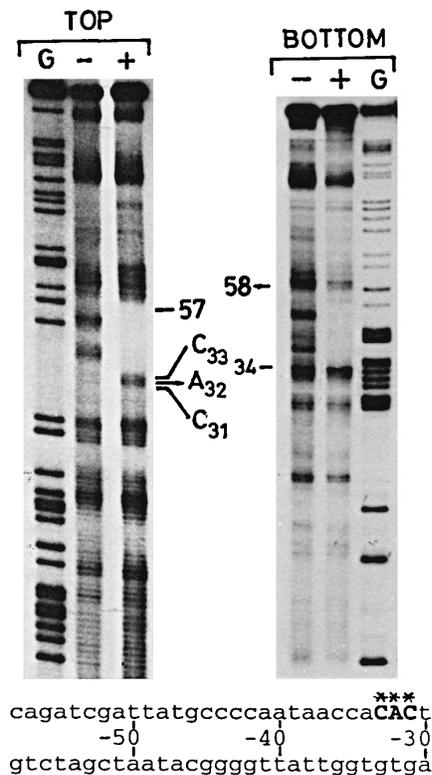
**Figure 4.** Hydroxyl radical footprinting of C protein. (a) The control (–) is the cleavage pattern in the absence of protein and G is the guanine ladder. Increasing amounts of C protein were used (100 ng, 200 ng, 500 ng for top strand, and 100 ng, 500 ng for the bottom strand). I, II and III are the three protected regions in each strand and are mapped in the sequence below. (b) Corresponding densitometric traces. The representative lanes were scanned from bottom to top of the autoradiogram. (c) The 2 to 3 bp offset in protection on each strand and the weaker protection in region III of top strand is better seen in this Figure. (–) and (+) refer to the cleavage pattern in the absence and presence of protein (200 ng), the other notations being same. (d) Densitometric scan of (c).

reactivity in the same region indicates that the sugar phosphate backbone structure in DNA is altered.

#### Deletion analysis

A search for putative DNA binding motifs in the 140 amino acid residues C open reading frame showed the presence of a helix–turn–helix (HTH) motif, spanning residues 100 to 119, in the carboxyl-terminal region of the protein (Bolker

*et al.*, 1989). Deletions in the C gene were created, using unique restriction enzyme sites located towards the 3' end of the gene. Two expression plasmids were constructed, pETAC13 and pETAC27, which encode for the proteins truncated at the carboxyl-terminal ends by 13 ( $\Delta$ C13) and 27 amino acid residues ( $\Delta$ C27), respectively. The expression of these proteins was confirmed by SDS-PAGE analysis of the S20 extracts prepared from the respective clones. The mobility of the overexpressed proteins present in the S20



**Figure 5.**  $(OP)_2Cu$  footprinting of C protein. The cleavage pattern of both the strands were analyzed in the absence (-) and presence (+) of 500 ng of C protein with their respective guanine specific sequencing ladders (G). The nucleotide positions mark the boundaries and the hyper-reactive nucleotides in the top strand are marked. Shown below is the nucleotide sequence in which the hyper-reactive nucleotides are in bold uppercase and marked with asterisks.

supernatant (Figure 6(a)) corresponds to their expected sizes, confirming their presence and integrity. These extracts were assayed for site-specific DNA binding activity using a specific 25 bp double-stranded oligonucleotide containing the C binding site in the presence of a large excess of non-specific DNA (poly(dI:dC)). Figure 6(b) shows the result of an electrophoretic mobility shift assay (EMSA) performed with these extracts. Extract overexpressing the C protein, when incubated with the labeled oligonucleotide, showed the formation of a single complex. The extract overexpressing the 13 amino acid deleted C protein ( $\Delta C13$ ) also showed the ability to form a single complex with a slightly faster mobility. The enhanced mobility of this complex indicates the specificity of the protein bound to labeled oligonucleotide used in the assay: only C or truncated versions of C protein, which retain the DNA binding activity, bind specifically to this oligonucleotide. In contrast, the extracts expressing  $\Delta C27$  failed to show any detectable DNA binding activity under identical conditions of binding. The 27 amino acid carboxyl-terminal deletion (at amino acid position 113) maps into the

putative HTH motif predicted by Bolker *et al.* (1989).

### Crosslinking analysis

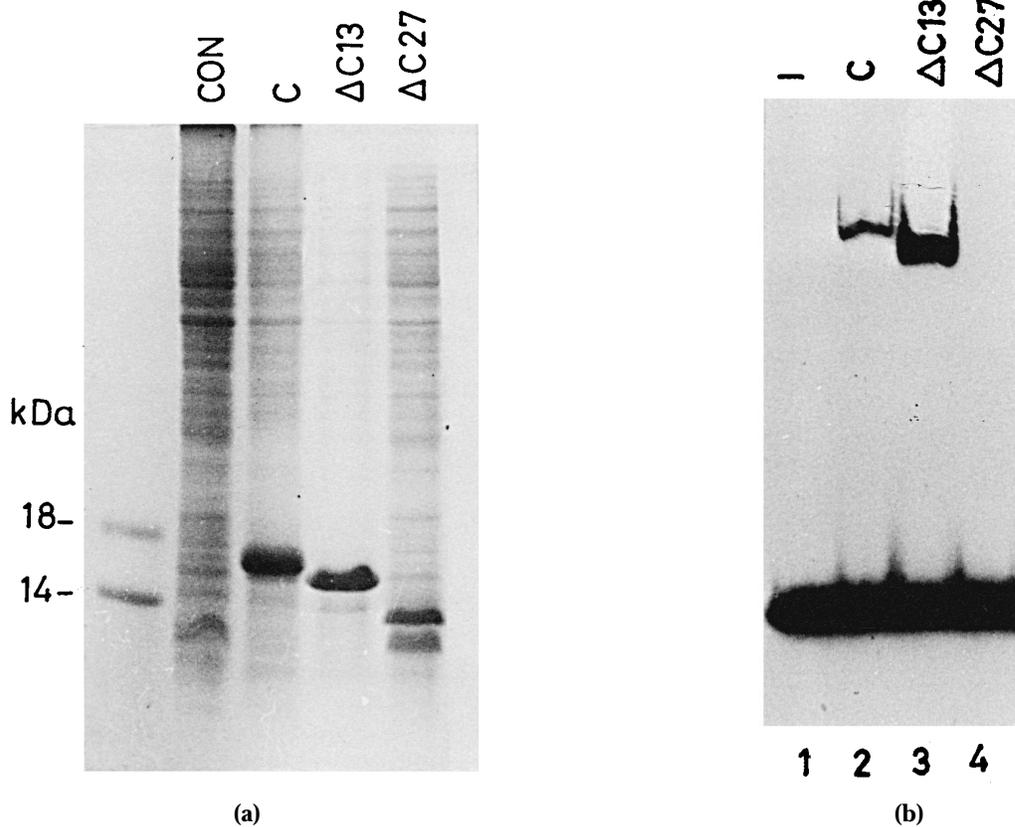
Protein crosslinking with bifunctional reagents, like glutaraldehyde and dimethylsuberimidate, has been used to investigate the subunit structure of oligomeric proteins. The subunits of oligomeric proteins are preferentially linked by a bifunctional crosslinking reagent in not too diluted protein solutions. In order to address the oligomeric state of C protein, chemical crosslinking was performed with glutaraldehyde. The purified protein was incubated with 0.025% and 0.05% glutaraldehyde, and the mixtures were subjected to SDS-PAGE and silver staining. The subunit molecular mass of C protein is 16.5 kDa and migrates as a single species corresponding to this size. Under the conditions of incubation, C protein crosslinked into dimers in both the concentrations of glutaraldehyde (Figure 7). The proportion of crosslinked dimers marginally increases in presence of DNA. However, no other higher crosslinked products were detected in the presence of DNA, suggesting that the protein also possibly binds as a dimer.

### Molecular exclusion chromatography

The presence of C protein in dimeric form in solution was further substantiated by subjecting the purified protein for gel filtration analysis (Figure 8). The two symmetric protein peaks, obtained at 11.5 and 17 minutes of elution, correspond to molecular masses of 32,000 and 16,000, respectively, in the calibration curve. SDS-PAGE analysis of both the samples showed the presence of 16.5 kDa protein band corresponding to the monomeric size of the C protein. These results confirm the presence of dimeric form of the protein under native conditions.

### Mixed oligomerization assay

The two different sets of experiments described above show that C protein is dimeric in free form and when bound to DNA. These experiments, however, do not rule out completely the tetrameric or other oligomeric form of the protein when bound to DNA. A mixed oligomerization assay is used to address this question. Since the dimers formed by full length and truncated C protein do not show appreciable difference in the size of the complex in electrophoretic mobility shift assay, the fusion Protein A-C was used. Protein A of *Staphylococcus aureus* does not bind DNA (Figure 9, lane 3) and is present as a monomer in solution (A. De, V.R. & V.N., unpublished results). Apart from serving as a flexible arm to increase the size of the protein, Protein A functions as an "affinity handle" (Nilsson & Abrahmsen, 1985). The A-C fusion protein formed a slow-moving complex with the DNA



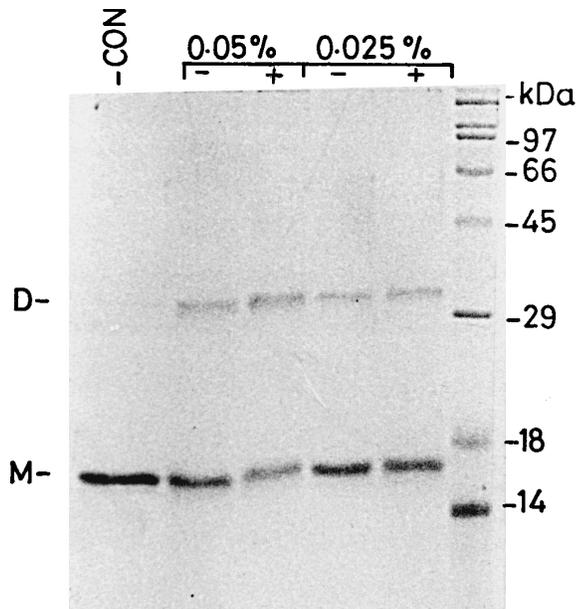
**Figure 6.** Expression and DNA binding analysis of C protein and its derivatives. (a) SDS-PAGE analysis of the S20 extracts expressing the full-length (C) and the carboxyl-terminal deleted C proteins ( $\Delta$ C13,  $\Delta$ C27). The molecular mass markers and their sizes in kDa are indicated. CON indicates the extract prepared from cells harboring the vector pET20b. (b) Electrophoretic mobility shift assay of the S20 extracts mentioned above. (-) indicates the free DNA in lane 1. One microgram samples of the indicated extracts were used in lanes 2 to 4. The intense shift in lane 3 is probably due to more  $\Delta$ C13 in the extracts.

fragment having a C binding site (Figure 9, lane 4 and 5). Incubation of pure A-C and C together in DNA binding assays resulted in the appearance of a unique intermediate complex, indicating the formation and binding to DNA of a heterodimer between the two proteins (Figure 9).

## Discussion

Using different footprinting reagents, the contact points for the phage Mu DNA binding protein C, at its cognate binding site located upstream and overlapping the  $-35$  sequence in  $P_{mom}$ , have been determined. The protein footprint covers a region of 18 to 25 bp depending on the footprinting agent used. Figure 10(a) summarizes all the determined contact points, and a helical representation on B-DNA is depicted in Figure 10(b). This extended region of footprint is due to the binding of a dimer of C protein of subunit molecular mass of 16.5 kDa. In order to make such extended contacts, it is likely that the protein distorts the DNA, and the results of footprinting experiments support this hypothesis. Upon protein binding, the guanines at positions  $-33$ ,  $-43$  and  $-53$  in the bottom strand and adenine at  $-37$  in the top strand are hyper-methylated. The functional groups on the purines for

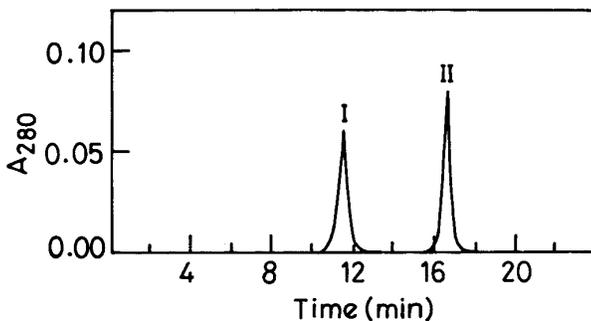
methylation are located in the grooves of DNA, and their enhanced reactivity could be a result of a more exposed site caused by protein-induced structural alterations. The protected guanines on the top and bottom strands are separated by 15 bp, and hence are out of phase by about half-a-turn of the helix. Two possible explanations can account for this kind of a contact. In the first case, the protein could contact these major grooves by binding to opposite faces of the helix. Alternatively, the DNA between the two regions could undergo a significant conformational change when bound to protein, so that the contacted grooves lie on the same side of the helix. The results of hydroxyl radical footprinting indicate that the protein binds to only one face of the helix, thus ruling out the first possibility. Other proteins like the *E. coli* cyclic AMP receptor protein (CRP) bind DNA as a dimer, with a HTH motif, and the co-crystal structure shows a large conformational change in DNA resulting in overall  $\sim 90^\circ$  bend (Schultz *et al.*, 1991). The phage P22 proteins Arc and Mnt also have extended contacts, but they bind to DNA as tetramers with  $\beta$ -sheet motif (Brown *et al.*, 1990). The contact pattern of the C protein, however, is more like a protein with a HTH motif (Pabo & Sauer, 1992; Suzuki & Yagi, 1994). Deletion analysis results also support a role



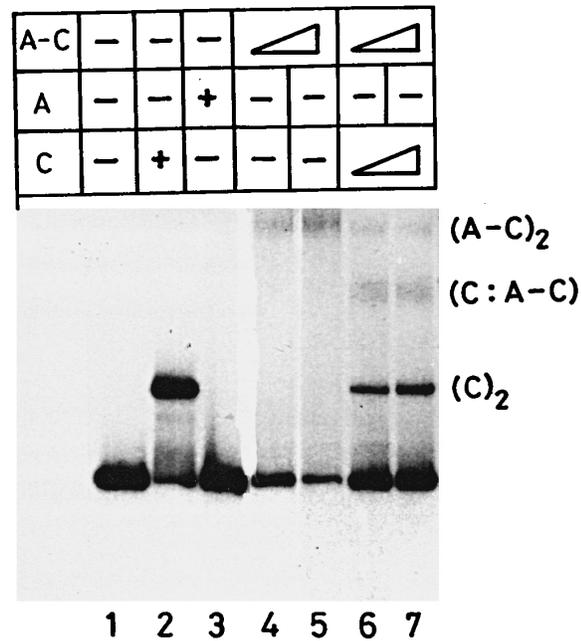
**Figure 7.** Analysis of crosslinked products. The amounts of glutaraldehyde used are indicated (%) and CON lane contains the untreated protein sample. (-) and (+) refer to the absence and presence of DNA, and D and M represent monomers and dimers. The sizes of molecular mass markers in kDa are indicated.

for such a putative motif located towards the carboxyl-terminal end of the protein.

The C binding site includes a TGTG motif located in its 3' half-site. This alternating purine-pyrimidine stretch, prone to helical deformations, is also present in the CRP recognition sequence (for a review see Kolb *et al.*, 1993). CRP introduces two  $\sim 45^\circ$  kinks in the two symmetrically located TGTG motifs to produce a net  $\sim 90^\circ$  bending in its binding site. The CRP-induced bends also result in widening of minor grooves. The role of this kink-prone motif, which is present only in the 3'



**Figure 8.** Size exclusion chromatography profile of C protein. An 80  $\mu$ g sample of C protein was loaded and analyzed as described in Materials and Methods. Standard marker proteins of known molecular mass were run separately to obtain a calibration curve. The proteins used were bovine serum albumin (66 kDa), ovalbumin (43 kDa), chymotrypsin (25 kDa) and lysozyme (13.7 kDa).



**Figure 9.** Heterodimer DNA binding assay. The end-labeled 220 bp DNA fragment (140 pmol, 10,000 cpm) was incubated with C, Protein A and Protein A-C and then analyzed on native polyacrylamide gel. Lane 1, free DNA; lane 2, 10 ng C protein; lane 3, 200 ng Protein A; lanes 4 and 5, 30 and 60 ng of Protein A-C; lanes 6 and 7, mixture of 10 ng C protein with 30 and 60 ng Protein A-C fusion, respectively.

half-site, in C protein's interaction is highlighted by the studies with chemical nucleases. It is clear from the DMS footprinting data that the protein is in close proximity to this motif, since the protected guanine ( $-35$  G) is a part of it. In spite of the bound protein, hydroxyl radicals react more only with the top strand in region III. Fe(II)-EDTA, because of its net negative charge, cannot approach DNA and thus is not sensitive to localized conformational changes induced by the protein. In contrast,  $(OP)_2Cu$  is a positively charged hydrophobic ligand that binds to the minor groove formed by three nucleotide pairs, causing strand scission *in situ* (Sigman *et al.*, 1991). Because of these properties it is more sensitive to conformational changes in DNA. Proteins like *EcoRI* bind exclusively in the major groove and do not affect the accessibility of the minor groove, hence do not show protection to  $(OP)_2Cu$  reactivity (McClarín *et al.*, 1986). On the other hand, this reagent has been used to detect the single-stranded regions formed by the *E. coli* RNAP during open complex formation, which are cut efficiently only on the template strands (Kuwabara & Sigman, 1987). Strained intermediates and polymerase-induced changes at the start of transcription have also been detected using this reagent (Thederahn *et al.*, 1990). Binding of C protein enhances the reactivity at the nucleotide triplet CAC in the top strand overlapping the  $-35$  region. Since the reagent is known to intercalate into underwound DNA and binds to a minor groove

formed by 3 bp, this increased reactivity in the bound complex could mean alterations in torsional flexibility of DNA. Compared to the 17 ( $\pm$ 1) bp spacing seen in large number of the promoters,  $P_{mom}$  has 19 bp separating the -10 and -35 regions (Figure 1). The 19 bp spacer means an additional twist angle of at least  $34^\circ$ , and the C protein-induced torsional change in DNA can compensate adequately for this difference. The reagent  $KMnO_4$  was used to detect single-stranded regions caused by this conformational change (Sasse-Dwight & Gralla, 1989). Absence of reactivity to the reagent in this region indicates that the binding of C protein does not lead to strand separation. The location of a high affinity binding site for the protein overlapping the -35 element could suggest that RNAP approaches the DNA from the face not contacted by the C protein. The increased reactivity to chemical nucleases reflects the accessibility of this region, possibly to the binding of RNAP.

The pattern of interaction seen here has a striking similarity with that of regulatory protein MerR, which regulates gene expression by binding to the 19 bp spacer region of the *mer* promoter and allosterically modulates the expression of a mercuric ion resistance operon in response to mercury by controlling the DNA conformation (Frantz & O'Halloran, 1990). The metalloprotein introduces localized unwinding of DNA at the center of its binding site, resulting in a symmetric DNA distortion that reorients the suboptimally phased promoter elements (Ansari *et al.*, 1992, 1995). The asymmetric interaction observed in the case of C protein would be necessary because of its binding upstream and overlapping of the -35 element, unlike MerR. The C protein-induced conformational change in DNA could also help to displace the RNAP from its upstream site P2 to facilitate its binding to the functional downstream site P1 and

initiate transcription. The extensive protein-DNA contacts would help to offset the free energy requirement for this conformational change. In addition to this phenomenon, the C protein could also make direct contacts with RNAP and influence transcription activation.

The positive control by Dam methylation, and repression by OxyR in the absence of methylation, are intriguing features of *mom* expression. Further, the binding sites for the two regulatory proteins, the repressor OxyR and activator C, overlap partially, suggesting mutually exclusive occupancy of the sites. While the OxyR binding site spans all the three GATC sequences, the C binding site includes only the promoter-proximal GATC sequence. OxyR fails to bind to a methylated *mom* promoter, whereas the binding of C is not affected by the methylation status (Bolker & Kahmann, 1989). Thus, the balance and interplay amongst the proteins Dam, OxyR and C at their target site, is a critical factor that decides the amounts and temporal fate of *mom* gene product.

## Materials and Methods

### Construction of plasmids

Unique restriction sites *Aat*II and *Bgl*II at the 3' end of the C gene cloned in the expression plasmid pVN61 (Ramesh *et al.*, 1994a) and the *Xba*I site in this plasmid were manipulated in such a way that the coding sequences of the C gene and the TAG sequence within the *Xba*I site are brought in-frame. The protocols involved in DNA cloning were followed as described by Sambrook *et al.* (1989). Briefly, the plasmid pVN61 was digested with *Aat*II + *Xba*I, and the ends were repaired by treating the DNA with phage T4 DNA polymerase in the presence of all the four dNTPs. The large fragment containing the C gene was purified and the ends were self-ligated to generate p $\Delta$ C13. To obtain p $\Delta$ C27, the plasmid pVN61



**Figure 10.** Summary of footprinting data. (a) The nucleotide sequence showing the hydroxyl radical protected positions are in uppercase and bracketed. Hyper-reactive residues (determined from DMS and  $(OP)_2Cu$  reactions) are marked with asterisks and the protected guanines are indicated by filled circles. The vertical arrows indicate the position of phosphate interference. (b) Helical representation of C-binding sites on B-DNA. The guanine contacts (x) and the phosphate contacts (●) are indicated.

was first digested with *Bgl*III and the overhangs were removed with  $S_1$  nuclease treatment. The DNA was digested with *Xba*I and the resulting 5' overhangs were filled by treating with Klenow fragment of *E. coli* DNA polymerase I in the presence of all four dNTPs. The large fragment obtained after these manipulations was purified and circularized by ligation to generate p $\Delta$ C27. To enhance the level of expression of truncated C protein from these plasmids, the respective truncated genes were subcloned into the phage T7 promoter-based expression vector pET20b following a polymerase chain reaction (PCR)-based strategy (Ramesh *et al.*, 1994a). For this purpose, the sequences were amplified by PCR using primers 5'-GCTAGC-CATGGACTTATTCGAACACGATCC-3', forward (having the *Nco*I site and start codon), and 5'-CAGGCTGA-AAATCTTCTC-3', reverse (annealing downstream of the *Hind*III site in the expression vector pKK223-3, Pharmacia). This mode of amplification resulted in the removal of a potential secondary structure at the translation initiation region of C mRNA and also in the loss of the second and third codons. The amplified PCR product was digested with *Nco*I + *Hind*III and cloned at the respective sites of the vector pET20b to generate pET $\Delta$ C13 and pET $\Delta$ C27. These plasmids code for 13 and 27 amino acid carboxyl-terminal truncated proteins  $\Delta$ C13 and  $\Delta$ C27, respectively. The plasmid pVR7 (Ramesh *et al.*, 1994a) served as the expression construct for the C protein (also lacking the second and third amino acid residues). The binding of this recombinant C protein to DNA in electrophoretic mobility shift assays, and in DNase I footprinting reactions, was identical to the wild-type C protein. The plasmid pRIT2T (Pharmacia) served as a source for Protein A. The plasmid pVN111 is a fusion between Protein A and C genes generated by cloning 0.8 kb *Bam*HI-*Pst*I fragment containing the C gene from plasmid pVN8 (Balke *et al.*, 1992) into vector pRIT2T.

### Expression and assay of deletion mutants

The expression plasmids pVR7, pET $\Delta$ C13 and pET $\Delta$ C27 were used to transform *E. coli* BL21(DE3)/pLysS, and the transformants were selected on chloramphenicol (25  $\mu$ g/ml) and ampicillin (100  $\mu$ g/ml) containing LB agar plates. The colonies were grown in LB (3.0 ml) containing the two drugs and induced with 0.3 mM IPTG at  $A_{600}$  nm for three hours. The cells were harvested, suspended in 200  $\mu$ l of ice-cold buffer 1 (10 mM Tris-HCl (pH 8.0), 5 mM  $MgCl_2$ , 50 mM NaCl, 1 mM dithiothreitol, 5% glycerol) and lysed by sonication. The extracts were clarified at 20,000  $g$ , 4°C (S20) and the supernatant was analyzed on SDS-15% polyacrylamide gels (Laemmli, 1970). The DNA binding activity of the overexpressed proteins was monitored in the S20 extracts, using an end-labeled 25 bp double-stranded oligonucleotide (5'-AGATCGATTATGCC-CATAACCAC-3') having the sequence corresponding to the C binding site in  $P_{mom}$ , by electrophoretic mobility shift assay. The assay was performed in a 20  $\mu$ l volume by incubating the extracts (1  $\mu$ g) with poly(dI:dC) (300 ng) in buffer 1 for ten minutes on ice followed by the addition of labeled oligonucleotide. The incubation was continued for ten minutes and the samples were analyzed on a 4% polyacrylamide gel (30:0.8 acrylamide:bis-acrylamide) equilibrated and electrophoresed in 0.5 TBE (45 mM Tris-borate, 1 mM EDTA) for two hours/100 V at 4°C.

### Labeling of the DNA fragment used in footprinting

A 260 bp *Eco*RI + *Hind*III fragment, containing the C binding site, from plasmid pUW4 (Ramesh *et al.*, 1994a) was end labeled at the *Hind*III site and used in footprinting reactions. Labeling of the top strand was achieved by digesting the plasmid with *Hind*III and rendering the ends blunt using a Klenow fragment of *E. coli* DNA polymerase I in presence of [ $\alpha$ - $^{32}$ P]dCTP (3000 Ci/mmol) and other dNTPs. To label the bottom strand, the *Hind*III-digested plasmid was dephosphorylated with calf intestinal alkaline phosphatase and then rephosphorylated with [ $\gamma$ - $^{32}$ P]ATP (5000 Ci/mmol) using T4 polynucleotide kinase, and digested with *Eco*RI. The labeled fragment was resolved on a 4% polyacrylamide gel, located by autoradiography. The corresponding region of the gel was excised and the DNA eluted by crush and soak method of Maxam & Gilbert (1980).

### C protein used in footprinting reactions

C protein was purified from *E. coli* BL26(DE3) carrying plasmid pVR7 as described (Ramesh *et al.*, 1994a,b). The protein was stored at a concentration of 500  $\mu$ g/ml in buffer I and used in footprinting and crosslinking experiments.

### Methylation protection (Fairall *et al.*, 1986)

For the protection from methylation by DMS the C protein:DNA complex (500 ng protein; 1 pmol DNA,  $10^5$  cpm) was formed in 100  $\mu$ l volume in buffer 1 and equilibrated at 25°C for ten minutes. DMS was freshly diluted in binding buffer to a final concentration of 2%, and 10  $\mu$ l was added to the preformed complex. After two minutes of incubation the reaction was stopped by the addition of 25  $\mu$ l DMS stop solution (1.5 M sodium acetate, 1 M 2-mercaptoethanol), deproteinized by extraction with phenol and chloroform, and the DNA was precipitated with three volumes of ethanol and 5  $\mu$ g carrier tRNA. The DNA was reprecipitated, washed with 70% ethanol and suspended in 90  $\mu$ l sterile distilled water. The control samples were methylated under identical conditions in the absence of C protein. The strand cleavage at the methylated guanines were performed by the addition of 10  $\mu$ l piperidine to 90  $\mu$ l methylated DNA and heating at 90°C for 30 minutes followed by freeze-drying.

### Ethylation interference (Siebenlist & Gilbert, 1980)

DNA (10 pmol,  $10^6$  cpm) in 100  $\mu$ l of 50 mM sodium cacodylate (pH 8.0), was mixed with an equal volume of freshly prepared saturated solution of ethylnitrosourea in ethanol. The samples were incubated for 60 minutes at 50°C, followed by two precipitations with ethanol and ammonium acetate in presence of carrier tRNA. Binding reactions were carried out with the ethylated DNA in buffer I, analyzed on a 4% non-denaturing gel and the DNA from the free and bound fractions were eluted. The eluted DNA was dissolved in 30  $\mu$ l of 10 mM sodium phosphate (pH 7.4) 1 mM EDTA. A 5  $\mu$ l sample of 1 M NaOH was added and the mixture was incubated for 30 minutes at 90°C. The mixture was neutralized with 5  $\mu$ l of 1 M HCl and the DNA was recovered by precipitation.

### Hydroxyl radical footprinting (Tullius & Dombroski, 1986; Wendy *et al.*, 1991)

C protein:DNA complexes were formed in 170  $\mu$ l (1 pmol end-labeled DNA,  $10^5$  cpm, 100 to 500 ng protein) in a buffer containing 10 mM Tris-HCl (pH 8.0), 5 mM MgCl<sub>2</sub>, 50 mM NaCl, 1 mM dithiothreitol, 0.5% glycerol. The reaction was initiated by the addition of 8  $\mu$ l of 125 mM Fe(II)/250 mM EDTA (pH 8.0), 8  $\mu$ l 0.84% H<sub>2</sub>O<sub>2</sub> and 14  $\mu$ l 20 mM sodium ascorbate, incubated for two minutes at 25°C and terminated with 20  $\mu$ l of 100 mM thiourea. Following protein removal by extraction with phenol:chloroform, the DNA was precipitated with ethanol in the presence of carrier tRNA.

### 1,10-Phenanthroline-copper [(OP)<sub>2</sub>Cu] footprinting (Sigman *et al.*, 1985)

C protein:DNA complexes were formed in 180  $\mu$ l binding buffer as described above. Ten microlitres of 2 mM 1,10-phenanthroline/0.15 mM CuSO<sub>4</sub> and 10  $\mu$ l of 58 mM 3-mercaptopropionic acid were added and incubated for one minute at 25°C. Digests were quenched by adding 5  $\mu$ l of 100 mM 2,9-dimethyl 1,10-phenanthroline, deproteinized by extraction with phenol:chloroform, and the DNA was precipitated with ethanol in the presence of carrier tRNA.

### Analysis of cleavage products

The precipitated DNA was washed, dried, suspended in 5  $\mu$ l formamide dye mixture (90% formamide, 10 mM NaOH, 1 mM EDTA, 0.01% xylene cyanol, 0.01% bromophenol blue), heated to 90°C for two minutes and electrophoresed on a denaturing gel (5% polyacrylamide, 50% urea). The gels were dried, autoradiographed and scanned using LKB Ultrosan XL densitometer (Pharmacia).

### Glutaraldehyde crosslinking

Purified C protein was dialyzed against 20 mM sodium phosphate (pH 8.0) and used in crosslinking experiments. Glutaraldehyde (25%, v/v) was diluted in sterile water to a concentration of 0.5% just before use. Three micrograms of purified protein was incubated with the indicated amounts of glutaraldehyde at 25°C in a reaction volume of 20  $\mu$ l. The crosslinking in the presence of DNA (25 ng of the double-stranded oligonucleotide mentioned above) was performed under identical conditions except that the protein was preincubated with DNA for ten minutes at 25°C prior to the addition of glutaraldehyde. The reaction was terminated by the addition of SDS sample buffer, separated on SDS-15% polyacrylamide gels (Laemmli, 1970) and visualized by silver staining (Schoenle *et al.*, 1984).

### Gel exclusion chromatography

Purified C protein (80 to 100  $\mu$ g) was applied to FPLC Hiload™ 26/60 Superdex™ 75 gel filtration column (Pharmacia, 22 ml bed volume) preequilibrated with 20 mM Tris-HCl (pH 7.5), 150 mM KCl, 5 mM MgCl<sub>2</sub> and 7 mM 2-mercaptoethanol. The elution was carried out in the same buffer at 48 ml/h flow rate. After monitoring the A<sub>280</sub> profile, the fractions were analyzed by SDS-PAGE.

### Heterodimer DNA binding assay

A 220 bp EcoRI-BamHI fragment from plasmid pUW4 (Ramesh *et al.*, 1994a) was used for electrophoretic mobility shift assay. The Protein A and A-C fusion proteins were purified by immunoglobulin G affinity chromatography (Nilssen & Abrahmsen, 1985). Purified C protein and A-C fusion protein were incubated individually or as a mixture with end-labeled DNA fragment in buffer 1 for ten minutes at 4°C. The samples were electrophoresed on 4% polyacrylamide gels in 0.5 × TBE for six hours at 4°C. Gels were dried and subjected to autoradiography.

### Acknowledgements

We thank K. Madhusudan, Amitabha De and S. Unniraman for help in experiments, D. R. Radha for expert technical assistance and U. Varshney for discussions. V.R. is supported by the Jawaharlal Nehru Centre for Advanced Scientific Research. This work is funded by a grant from the Department of Science and Technology, Government of India.

### References

- Ansari, A. Z., Chael, M. L. & O'Halloran, T. V. (1992). Allosteric underwinding of DNA is a critical step in positive control of transcription by Hg-MerR. *Nature*, **355**, 87–89.
- Ansari, A. Z., Bradmer, J. E. & O'Halloran, T. V. (1995). DNA-bend modulation is a repressor-to-activator switching mechanism. *Nature*, **374**, 371–375.
- Balke, V., Nagaraja, V., Gindlesperger, T. & Hattman, S. (1992). Functionally distinct RNA polymerase binding sites in the phage Mu *mom* promoter region. *Nucl. Acids Res.* **20**, 2777–2784.
- Bolker, M. & Kahmann, R. (1989). The *Escherichia coli* regulatory protein OxyR discriminates between methylated and unmethylated states of the phage Mu *mom* promoter. *EMBO J.* **8**, 2403–2410.
- Bolker, M., Wulczyn, F. G. & Kahmann, R. (1989). Role of bacteriophage Mu C protein in activation of the *mom* gene promoter. *J. Bacteriol.* **171**, 2019–2027.
- Brown, B. M., Bowie, J. U. & Sauer, R. T. (1990). Arc repressor is tetrameric when bound to operator DNA. *Biochemistry*, **29**, 11189–11195.
- Chiang, L. W. & Howe, M. M. (1993). Mutational analysis of a C-dependent late promoter of bacteriophage Mu. *Genetics*, **135**, 619–629.
- Fairall, L., Rhodes, D. & Klug, A. (1986). Mapping of the sites of protection on a 5 S RNA gene by the *Xenopus* transcription factor IIIA. A model for the interaction. *J. Mol. Biol.* **192**, 577–591.
- Frantz, B. & O'Halloran, T. V. (1990). DNA distortion accompanies transcriptional activation by the metal-repressor gene-regulatory protein MerR. *Biochemistry*, **29**, 4747–4751.
- Gindlesperger, T. L. & Hattman, S. (1994). *In vitro* transcriptional activation of the phage Mu *mom* promoter by C protein. *J. Bacteriol.* **176**, 2885–2891.
- Harley, C. B. & Reynolds, R. P. (1987). Analysis of *E. coli* promoter sequences. *Nucl. Acids Res.* **15**, 2343–2361.
- Hattman, S. (1982). DNA methyltransferase-dependent transcription of the phage Mu *mom* gene. *Proc. Natl Acad. Sci. USA*, **79**, 5518–5521.

- Hattman, S., Ives, J., Margolin, W. & Howe, M. M. (1985). Regulation and expression of the bacteriophage Mu *mom* gene: mapping of the transactivation (Dad) function to the C region. *Gene*, **39**, 71–76.
- Hattman, S., Newman, L., Krishna Murthy, H.M. & Nagaraja, V. (1991). Com, the phage Mu *mom* translational activator, is a zinc-binding protein that binds specifically to its cognate mRNA. *Proc. Natl Acad. Sci. USA*, **88**, 10027–10031.
- Hawley, D. K. & McClure, W. R. (1983). Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucl. Acids Res.* **11**, 2237–2255.
- Heisig, P. & Kahmann, R. (1986). The sequence and *mom* transactivation function of the C gene of bacteriophage Mu. *Gene*, **43**, 59–67.
- Kahmann, R. & Hattman, S. (1987). Regulation and expression of the *mom* gene. In *Phage Mu* (Symonds, N., Toussaint, A., Van de Putte, P. & Howe, M. M., eds), pp. 93–109, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Kolb, A., Busbu, S., Buc, H., Garges, S. & Adhya, S. (1993). Transcriptional regulation by cAMP and its receptor protein. *Annu. Rev. Biochem.* **62**, 749–795.
- Kuwabara, M. D. & Sigman, D. (1987). Footprinting DNA–protein complexes *in situ* following gel retardation assays using 1,10-phenanthroline-copper ion: *Escherichia coli* RNA-polymerase-*lac* promoter complexes. *Biochemistry*, **26**, 7234–7238.
- Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680–685.
- Margolin, W. & Howe, M. M. (1986). Localization and DNA sequence analysis of the C gene of bacteriophage Mu, the positive regulator of Mu late transcription. *Nucl. Acids Res.* **14**, 4881–4897.
- Margolin, W., Rao, G. & Howe, M. M. (1989). Bacteriophage Mu late promoters: four late transcripts initiate near a conserved sequence. *J. Bacteriol.* **171**, 2003–2018.
- Maxam, A. M. & Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**, 494–560.
- McClarin, J. A., Frederick, C. A., Wang, B. C., Greene, P., Boyer, H. W., Grable, J. & Rosenberg, J. M. (1986). Structure of the DNA- *EcoRI* endonuclease recognition complex at 3 Å resolution. *Science*, **234**, 1526–1541.
- Nagaraja, V., Hecht, G. & Hattman, S. (1988). The phage Mu late transcription factor, C, is a site-specific DNA binding protein. *Biochem. Pharmacol.* **37**, 1809–1810.
- Nilssen, B. & Abrahmsen, L. (1985). Fusions to staphylococcal protein A. *Methods Enzymol.* **185**, 144–161.
- Pabo, C. O. & Sauer, R. T. (1992). Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.* **61**, 1053–1095.
- Raibaud, O. & Schwartz, M. (1984). Positive control of transcription initiation in bacteria. *Annu. Rev. Genet.* **18**, 173–206.
- Ramesh, V., De, A. & Nagaraja, V. (1994a). Engineering hyperexpression of bacteriophage Mu C protein by removal of secondary structure at the translation initiation region. *Protein Eng.* **7**, 1053–1057.
- Ramesh, V., De, A. & Nagaraja, V. (1994b). Overproduction and purification of C protein, the late gene transcription activator from phage Mu. *Protein Exp. Purif.* **5**, 379–384.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd edit., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Sasse-Dwight, S. & Gralla, J. D. (1989). KMnO<sub>4</sub> as a probe for *lac* promoter DNA melting and mechanisms *in vivo*. *J. Biol. Chem.* **264**, 8074–8081.
- Schoenle, E. J., Adams, L. D. & Sammons, D. W. (1984). Insulin-induced rapid decrease of a major protein in fat cell plasma membrane. *J. Biol. Chem.* **259**, 12112–12116.
- Schultz, S. C., Shields, G. C. & Steitz, T. A. (1991). Crystal structure of a CAP–DNA complex: the DNA is bent by 90°. *Science*, **253**, 1001–1007.
- Seeman, N. C., Rosenberg, J. M. & Rich, A. (1976). Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl Acad. Sci. USA*, **73**, 804–808.
- Siebenlist, U. & Gilbert, W. (1980). Contacts between *Escherichia coli* RNA polymerase and an early promoter of phage T7. *Proc. Natl Acad. Sci. USA*, **77**, 122–126.
- Sigman, D. S., Spassky, A., Rimsky, S. & Buc, H. (1985). Conformational analysis of *lac* promoters using the nuclease activity of 1,10-phenanthroline-copper ion. *Biopolymers*, **24**, 183–197.
- Sigman, D. S., Kuwabara, M. D., Chen, C.-H.B. & Bruice, T. W. (1991). Nuclease activity of 1,10-phenanthroline copper in study of protein–DNA interactions. *Methods Enzymol.* **208**, 414–433.
- Spassky, A. & Sigman, D. (1985). Nuclease activity of 1,10-phenanthroline-copper ion. Conformational analysis and footprinting of the *lac* operon. *Biochemistry*, **24**, 8050–8056.
- Sun, L. & Singer, B. (1975). The specificity of different classes of ethylating agents toward various sites of HeLa cell DNA *in vitro* and *in vivo*. *Biochemistry*, **14**, 1795–1802.
- Suzuki, M. & Yagi, N. (1994). DNA recognition code of transcription factors in the helix-turn-helix, probe helix, hormone receptor and zinc finger families. *Proc. Natl Acad. Sci. USA*, **91**, 12357–12361.
- Thederabad, T. B., Spassky, A., Kuwabara, M. D. & Sigman, D. S. (1990). Chemical nuclease activity of 5-phenyl-1,10-phenanthroline-copper ion detects intermediates in transcription initiation by *E. coli* RNA polymerase. *Biochem. Biophys. Res. Commun.* **168**, 756–762.
- Travers, A. (1992). DNA conformation and configuration in protein–DNA complexes. *Curr. Opin. Struct. Biol.* **2**, 71–77.
- Tullius, T. D. & Dombroski, B. A. (1986). Hydroxyl radical footprinting: high-resolution information about DNA–protein contacts and application to λ repressor and cro proteins. *Proc. Natl Acad. Sci. USA*, **83**, 5469–5473.
- Wendy, D. J., Hayes, J. J., Levin, J. R., Weidner, M. F., Dombroski, B. A. & Tullius, T. D. (1991). Hydroxyl radical footprinting. *Methods Enzymol.* **208**, 380–414.
- Wulczyn, F. G. & Kahmann, R. (1991). Translational stimulation: RNA sequence and structure requirements for binding of Com protein. *Cell*, **65**, 259–269.

Edited by R. F. Schleif

(Received 7 November 1995; received in revised form 8 April 1996; accepted 15 April 1996)