

Identification of the Domains for DNA Binding and Transactivation Function of C Protein from Bacteriophage Mu

Bindu Diana Paul,¹ Aditi Kanhere,² Atanu Chakraborty,¹ Manju Bansal,² and Valakunja Nagaraja^{1,3,4,*}

¹Department of Microbiology and Cell Biology and ²Molecular Biophysics Unit, Indian Institute of Science, Bangalore-560012, and ³Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore 560064, India

ABSTRACT The C protein, a middle gene product of bacteriophage Mu, is the determinant of the transition from middle to late gene expression. C activates transcription from four late gene promoters, P_{lys} , P_I , P_P , and P_{mom} by binding to a site overlapping their -35 elements. Site-specific, high-affinity binding of C to its recognition sequence results in both axial and torsional distortion of DNA at P_{mom} , which appears to play a role in recruitment of RNA polymerase to the promoter for *mom* gene transactivation. To identify the regions of C protein important for its function, deletion and site-directed mutagenesis were carried out. We demonstrate here that a helix-turn-helix (HTH) motif located toward the carboxy terminal end of the protein is the DNA-binding domain and amino acid residues involved in transactivation overlap the HTH motif. Mutagenesis studies also aided in the identification of the region important for dimerization. Structure-based sequence alignment and molecular modeling in conjunction with mutational analysis suggest that the HTH motif is part of a three-helix bundle, with remarkable similarity to paired (*prd*), a developmental regulatory protein from *Drosophila*. Additional key residues identified in the model to be crucial for C protein structure and DNA binding were shown to be important by mutagenesis. These results provide a structural framework for C function and insight into the mechanism of transactivation at the *mom* promoter. *Proteins* 2003;52:272–282. © 2003 Wiley-Liss, Inc.

Key words: phage mu; helix-turn-helix; site-directed mutagenesis; paired protein; molecular modeling

INTRODUCTION

Bacteriophage Mu C protein is a specific transcription activator of late gene expression.^{1,2} Late gene expression in bacteriophage Mu begins with the expression of the *lys* gene, required for host cell lysis and culminates with the expression of *mom*, which modifies the phage DNA.³ In addition, the genes encoding head and tail proteins are turned on during the last phase of the lytic cycle. The *mom* gene product is the last to be expressed and is detrimental to the cell.⁴ As a consequence, it is regulated by a series of complex mechanisms during transcription and transla-

tion.² Two overlapping promoters, designated $P1$ (P_{mom}) and $P2$, have been identified upstream of the *mom* mRNA start site.⁵ In the absence of C protein, RNA polymerase is bound to the promoter $P2$ to transcribe leftward, in the opposite direction of P_{mom} .⁶ However in presence of C, which binds to its site, overlapping the -35 element of P_{mom} , rightward transcription is activated. The polymerase recruitment to P_{mom} in presence of C has been demonstrated by *in vitro* and *in vivo* footprinting experiments.⁵

The *mom* promoter is a typical transactivation dependent promoter⁷ comprising a weak -35 element and a suboptimal spacing of 19 bp between the -10 and the -35 elements. C protein has been shown to bind to its site at P_{mom} as a dimer of 33 kDa⁸ with high affinity.⁹ Mg^{2+} ions are essential for the binding process. Circular dichroism studies have revealed that Mg^{2+} ions trigger a conformational change in the protein, converting it to a structure with a higher α -helical content.¹⁰ C-mediated axial distortion of the *mom* regulatory region has been detected by dimethyl sulfate (DMS) probing and electrophoretic mobility shift assays (EMSA) using circularly permuted fragments.^{11,12} Based on methylation protection and interference analysis, the minimal C-binding site at P_{mom} was identified as 5'GATTATgcccacaATAACC 3'^{11,13} [Fig.1(A)]. Probing with the minor groove-specific ligand, $(OP)_2Cu$, revealed localized distortion and widening of the minor groove at the 3' end of the C-binding site, suggesting unwinding of DNA.¹¹ Protein-mediated unwinding was confirmed by coupled topoisomerase assays.¹³

Abbreviations: DMS, dimethyl sulfate; EMSA, electrophoretic mobility shift assay; bp, base pair; aa, amino acid; $(OP)_2Cu$, bis (1,10-phenanthroline) copper; LB, Luria Bertani medium; IPTG, isopropyl- β -D-thiogalactopyranoside; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; HTH, helix-turn-helix; *prd*, paired protein.

Grant sponsor: Department of Science and Technology postdoctoral fellowship to B.D.P. Grant sponsor: University Grants Commission and Council of Scientific and Industrial Research, Government of India senior research fellowships to A.K. and A.C. Grant sponsor: Department of Science and Technology, Government of India to V.N.

*Correspondence to: Valakunja Nagaraja, Department of Microbiology and Cell Biology, Indian Institute of Science, Bangalore-560012, India. E-mail: vnrj@mcbl.iisc.ernet.in

Received 22 April 2003; Accepted 8 January 2003

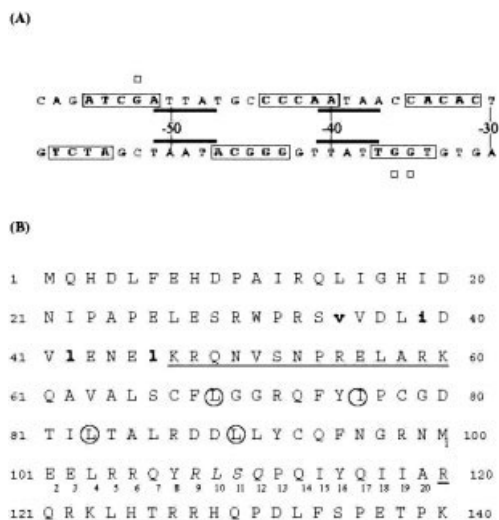


Fig. 1. (A) C recognition sequence. The dyad symmetry is indicated by horizontal bars. Bases that are boxed are regions protected from hydroxyl radicals.¹¹ Squares represent protected guanines in DMS footprinting analysis. (B) Primary sequence of C protein. The two helices of the helix-turn-helix (HTH) motif are shown in bold, with the turn in italics. The residues comprising the HTH motif are numbered, and the basic stretches are underlined. The hydrophobic residues in the first heptad repeat are in small bold letters, whereas those comprising the second heptad repeat have been circled.

Thus, the C protein has certain remarkable properties: (1) high-affinity binding comparable to that of the lambda cI repressor; (2) Mg²⁺-triggered conformational change as a prerequisite for binding; (3) axial distortion of DNA, and (4) asymmetric unwinding at 3' end of the recognition sequence. In addition to the DNA-binding motif and a dimerization interface, the protein possibly contains an interacting surface for RNA polymerase for transactivation function. Attempts to crystallize the protein have been unsuccessful so far, and hence we have resorted to an alternate approach. Here we describe attempts to locate the domains of C involved in DNA binding, dimerization, and transactivation and to identify the important residues that contribute to these key functions. A combination of sequence alignment, secondary structure prediction, fold recognition programs, and mutational data were used to build a model of the C protein-DNA complex. This knowledge based molecular model and mutational analysis provide insights into the possible roles of amino acid residues of C protein in determining its structure as well as DNA recognition.

MATERIALS AND METHODS

Bacterial Strains and Plasmids

Escherichia coli DH10B was the host strain for cloning experiments. The expression strain *E. coli* BL26(DE3) was obtained from Novagen (Madison, WI). The plasmid pVR7⁹, which carries the C gene under the control of the T7 ϕ_{10} promoter was used both for purification of C protein and for site-directed mutagenesis. The plasmid pCCV(-) was provided by Stanley Hattman, University of Rochester, NY. It is a derivative of pACYC184 having the *mom-lacZ* reporter construct.¹

TABLE I. Summary of Mutagenesis Studies†

Mutants	DNA binding	Transactivation	Dimerization
Δ C13	+	+	+
Δ C14	+	+	+
A-C Δ N18 Δ C13	+	ND	+
A-C Δ N18 Δ C27	-	ND	ND
Δ N43 Δ C48	ND	ND	+
V35R	ND	ND	+
L38R	ND	ND	+
L42R	ND	ND	+
L46R	ND	ND	+
L42RL46R	ND	ND	+
L69M	+	+	+
L83R	+	+	+
L90S	+	+	+
L83RL90S	ND	ND	+/-
K47D	+	+	+
N50D	+	+	+
I76L	+	+	+
G79A	-	-	ND
R87D	-	-	ND
F95A	+	+/-	+
G97A	-	-	ND
L103Q	-	-	+
R105D	+	+/-	+
S110A	-	-	ND
Q111A	+	+/-	+
I114N	-	-	+
R120D	-	-	ND
R122A	+	+	+
K123Q	+	+	+
H125A	+	+/-	+
R127L	+	+	+
R127 stop	+	+	+

†ND, not determined; +/-, weak DNA binding/transactivation/dimerization.

Mutagenesis and Preparation of Mutant Proteins

Site-directed mutagenesis was carried out using the GeneEditor™ *in vitro* mutagenesis kit (Promega, Madison, WI) using plasmid pVR7 as a template. Mutants were also generated by PCR based error-prone mutagenesis.¹⁴ The minimal heptad repeat (M44-Y92) was amplified using pVR7 as template. The PCR product was cloned into *Nco*I and blunted *Bam*HI sites of pET11d (Novagen) vector to obtain Δ N43 Δ C48. The deletion mutants, Δ N18 Δ C13 and Δ N18 Δ C27 (Table I), were generated as described earlier.⁸ The sequence of the mutants was confirmed by Sanger's dideoxy sequencing. The mutant proteins were expressed in *E. coli* BL26(DE3). Cells were grown in 20 ml LB containing 100 μ g/ml ampicillin up to an OD of 0.6 at A₆₀₀ and induced with 0.3 mM IPTG for 3 h at 37°C. In the case of L103Q and I114N, cells were induced at 16°C for 6 h to obtain the protein in a soluble form. After harvesting, cells were suspended in 1 ml of buffer containing 20 mM Tris-HCl, pH 8.0, 5 mM MgCl₂, 1 mM EDTA, 7 mM 2-mercaptoethanol, 5% glycerol, 150 mM NaCl, and 1 mM PMSF and disrupted by sonication. The extract was centrifuged at 20,000 g for 30 min at 4°C (S20 fraction). The

samples were subjected to SDS-PAGE and visualized by Coomassie Brilliant Blue staining. In cases where the mutant proteins were present mostly in the S20 pellet fraction, the band corresponding to the overexpressed protein was excised and eluted from a 15% SDS polyacrylamide gel using a BioRad Electroelutor. The eluted proteins were then subjected to acetone precipitation to remove SDS, denatured in buffer containing 6 M urea, and then renatured by stepwise dialysis against buffer containing 4, 2, and 1 M urea, respectively. The proteins were then dialyzed against buffer (10 mM HEPES, 5 mM MgCl₂, 1 mM EDTA, 150 mM NaCl, and 5% glycerol) lacking urea.

Preparation of DNA

Five picomoles of the top strand of a 25 mer C-binding site with the sequence 5' AGATCGATTATGCCCAATA-ACCAC 3' was labeled using 40 μ Ci of [γ -³²P]ATP (6000 Ci/mmol, New England Nuclear, Boston, MA) and mixed with equimolar amounts of the complementary strand and incubated for 5 min at 65°C. The mixture was allowed to cool slowly to room temperature following which the DNA was purified by passing through Sephadex G50.

Electrophoretic Mobility Shift Assays

The assays were carried out as described previously^{10,11} in 30- μ l reactions in binding buffer (20 mM Tris-HCl, pH 8.0; 5 mM MgCl₂, 1 mM EDTA, 7 mM 2-mercaptoethanol, 5% glycerol, and 150 mM NaCl). DNA (10,000 cpm) was incubated with 1 μ g of crude extracts harboring the mutant or wild type C protein, on ice for 15 min. The samples were loaded on a 10% native polyacrylamide gel and electrophoresed in 0.5 \times TBE (45 mM Tris-borate, 1 mM EDTA, pH 8.0) buffer at 150 V for 1 h at 4°C and subjected to autoradiography.

Dimerization Assay

The ability of the C protein and the mutant proteins to form dimers was assessed by chemical crosslinking using glutaraldehyde (Merck) as described earlier.¹¹ All the proteins under study were denatured in 6 M urea and sequentially renatured as described in the earlier section. Purified C or its mutants were dialyzed against crosslinking buffer (10 mM HEPES, pH 7.5, 5 mM MgCl₂, 1 mM EDTA, 150 mM NaCl, 7 mM 2-mercaptoethanol, 5% glycerol). C or mutant proteins (1 μ g) were incubated with different concentrations of glutaraldehyde (0.0025% and 0.005%) at room temperature for 30 min. The reactions were stopped by the addition of SDS loading dye, subjected to 15% SDS-PAGE, and visualized by silver staining. In order to assess the role of C residues in dimerization, wild type C protein (5 μ M) was denatured in 6 M guanidium HCl at 37°C for 30 min. One aliquot of this denatured protein was sequentially dialyzed against binding buffer (20 mM Tris-HCl, pH 8.0, 5 mM MgCl₂, 1 mM EDTA, 7 mM 2-mercaptoethanol, 5% glycerol, and 150 mM NaCl) as described in urea renaturation step. The remainder of denatured protein was treated with 1 mM *N*-ethylmaleimide at 37°C for 30 min and then sequentially dialyzed in

the same way as the untreated part. All the dialyzed fractions were assayed for DNA-binding activity by EMSA.

Transactivation Assays

The plasmids bearing either the wild type or mutant C gene (under the control of the T7 ϕ ₁₀ promoter) were transformed into *E. coli* BL26 (DE3) harboring the *mom-lacZ* fusion construct pCCV(-). Individual colonies were inoculated into LB medium, cell-free extracts were prepared and β -galactosidase activity was assayed according to Miller.¹⁵ The activity was measured in the uninduced cells itself because of the leaky expression by T7 RNA polymerase.

Modeling the Interaction of C Protein with Its Site

The secondary structure prediction was carried out using the PHD algorithm,¹⁶ which identified six helices in the protein, including a three-helix bundle encompassing the helix-turn-helix (HTH) motif.

To pinpoint more precisely the orientation of this motif vis-à-vis the DNA, different "Fold Recognition" programs such as 123D,¹⁷ TOPITS,¹⁸ MBA-FOLD,¹⁹ and 3D-PSSM²⁰⁻²² were used. When the entire C protein sequence was analyzed for comparison, no significantly matching sequence/structure was picked up. Therefore a modular approach was adopted using shorter fragments of the C protein. The best matches were obtained by 3D-PSSM²⁰⁻²² program for the carboxy terminus fragment spanning three helices connected by short loops. 123D could not pick up any significant hits, whereas TOPITS picked up functionally unrelated proteins. Among the top 20 hits 3D-PSSM picked up many DNA-binding proteins that are involved in transcriptional regulation. The hits given by MBA-fold program also had quite a few DNA-binding proteins. Homeodomain proteins were most significant in both cases. SCOP²³ classifies most of them into proteins containing a three-helix bundle motif. Finally, sequence alignment for this fragment was done using various programs such as GAP,²⁴ CLUSTAL-W,²⁵ LALIGN,²⁶ and ALION.²⁷ The prd paired protein²⁸ (alignment score 37), homeodomain²⁹ (alignment score 26), and NarL³⁰ (alignment score 17) were found to have significant similarity, with the prd paired homeodomain being the most similar. The similarity between 48 amino acid fragment from the C-terminal domain of the C protein and the N-terminal region of paired prd domain is 42% with 18.2% amino acids being identical.

The model of C protein was based on the structure of prd paired protein complexed with DNA (PDB code, 1pdn). The structure of the 10 mer duplex DNA (-53C to -44C), which is derived from the binding site of C protein [Fig. 1(A)], was generated using an in-house package NUC-GEN.³¹ This algorithm uses local step geometrical parameters for different dinucleotide sequences, which have been obtained from the crystal structures of DNA complexed with proteins. The model structure built was energy minimized using the AMBER package.³²

RESULTS

Features in the Primary Sequence of the C Protein

Analysis of the primary sequence of the C protein, as depicted in Figure 1(B), indicated the presence of a few putative motifs, which may have important roles in its function. An HTH motif present toward the carboxy terminal half of the protein (M100-A119) was detected using the weight matrix method of Brennan and Matthews³³ and Dodd and Egan.³⁴ A short heptad repeat with 4-3-4 pattern of hydrophobic residues, V35, I39, L42, and L46 as well as a 22 mer heptad repeat between L69 and L90 were located toward the N-terminal and middle region of the protein, respectively. In addition, two basic amino acid stretches, K47-K60 and R120-H129, were also identified. These putative motifs were considered to be likely candidates for the different properties *viz.* DNA binding, oligomerization, and transactivation of the protein.

The DNA Binding Activity of C Protein Is Mediated by the HTH Motif

As proposed for a 20-amino acid motif adopting an HTH conformation,³³ the sequence M100 to A119 in C protein has hydrophobic residues at positions 4 and 15 (*viz.* L103 and I114). In addition, amino acids 3 to 8 and 15 to 20 do not have any strong helix breaking residues [Fig. 1(B)]. The putative motif in C protein differs from a typical HTH in two less conserved positions, in not having an A or G at position 5 and G at position 9. It is important to note here that HTH motifs of several other proteins also do not obey all the stereochemical rules.³³ For example, the TrpR protein has a K at position 5.³³ The AraC protein has a C at position 9, and the bacteriophage D108 Ner and LysR have H at this position.³⁴ Some other variations at position 5 and 9 of HTH motif are presented in discussion section. The C protein has R at both these positions [Fig. 1(B)]. To test the functional significance of the putative HTH and to determine if this motif is involved in DNA binding, site-directed mutagenesis of the conserved hydrophobic residues at the fourth and fifteenth positions was carried out. The ability of the C protein mutants to bind DNA was assayed by electrophoretic mobility shift assays using a 25 mer end-labeled DNA having the C-binding site (see Materials and Methods). The wild-type protein forms a characteristic single complex with its cognate site [Fig. 2(A), lane 2]. In contrast, the mutants, L103Q and I114N, were found to be severely compromised in their DNA-binding activity [Fig. 2(A), lanes 3 and 4, respectively]. Although L103Q and I114N were defective in DNA binding [Fig. 2(A)], they formed dimers [Fig. 3(G,H); Table I]. A deletion mutant, Δ C27, which lacks the last 27 amino acids of the protein also does not bind DNA.¹¹ This deletion removes the last 6 amino acid residues from the recognition helix of the putative HTH motif. On the other hand, a deletion lacking the last 13 amino acids of the protein, having intact motif, binds to DNA, indicating that residues beyond R127 are not important for DNA binding.^{8,11}

To address whether the basic amino acids found downstream of the HTH motif are involved in DNA binding, the residues R120, R122, K123, H125, and

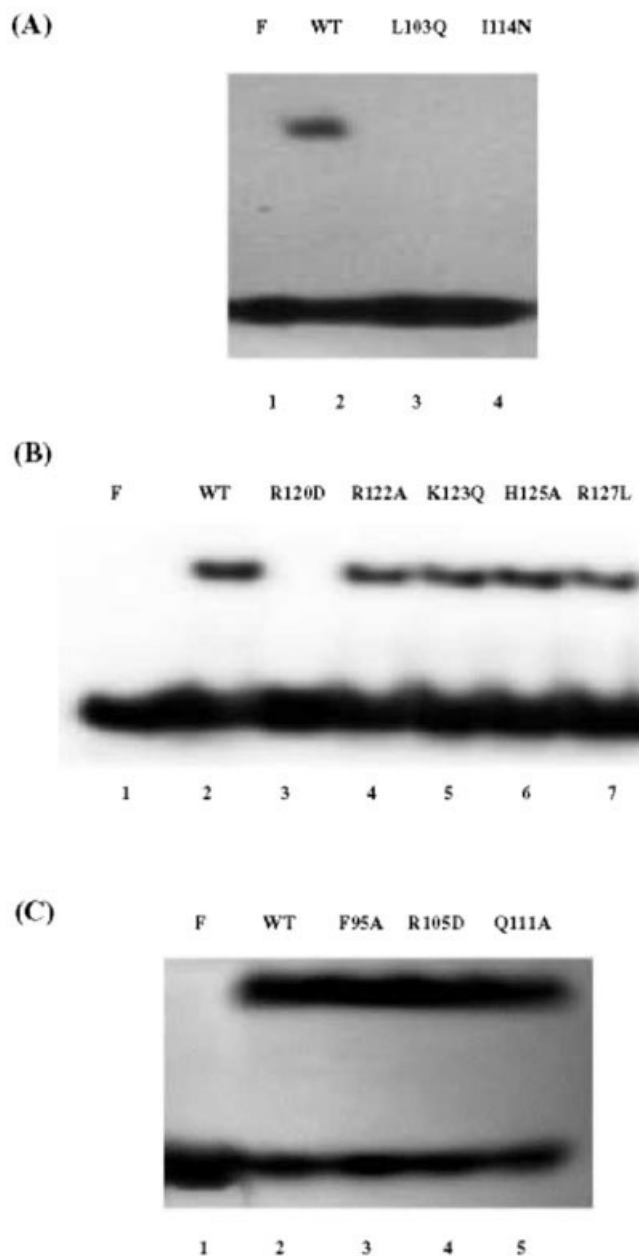


Fig. 2. Assessment of DNA binding. A 25-bp duplex oligonucleotide having the C binding site was incubated with the wild type and mutant proteins and electrophoresed on a 10% native polyacrylamide gel as described in Materials and Methods. (A) DNA-binding activities of the HTH mutants. (B) DNA-binding activity of the C terminal basic stretch mutants. F, free DNA; WT, wild type C protein. (C) DNA-binding activity of F95A, R105D, and Q111A mutants.

R127 were mutated (see Table I). All the mutants with the exception of R120D retain their ability to bind DNA. An evaluation of their DNA-binding properties revealed that these mutants bind DNA with an affinity comparable to that of the wild type C protein [Fig. 2(B), Table I]. Residues in the basic stretch preceding the heptad repeat [Fig. 1(B)] were also mutated. The two mutants, N50D and K47D, were proficient in DNA binding as well as transactivation (Table I). In an independent study,

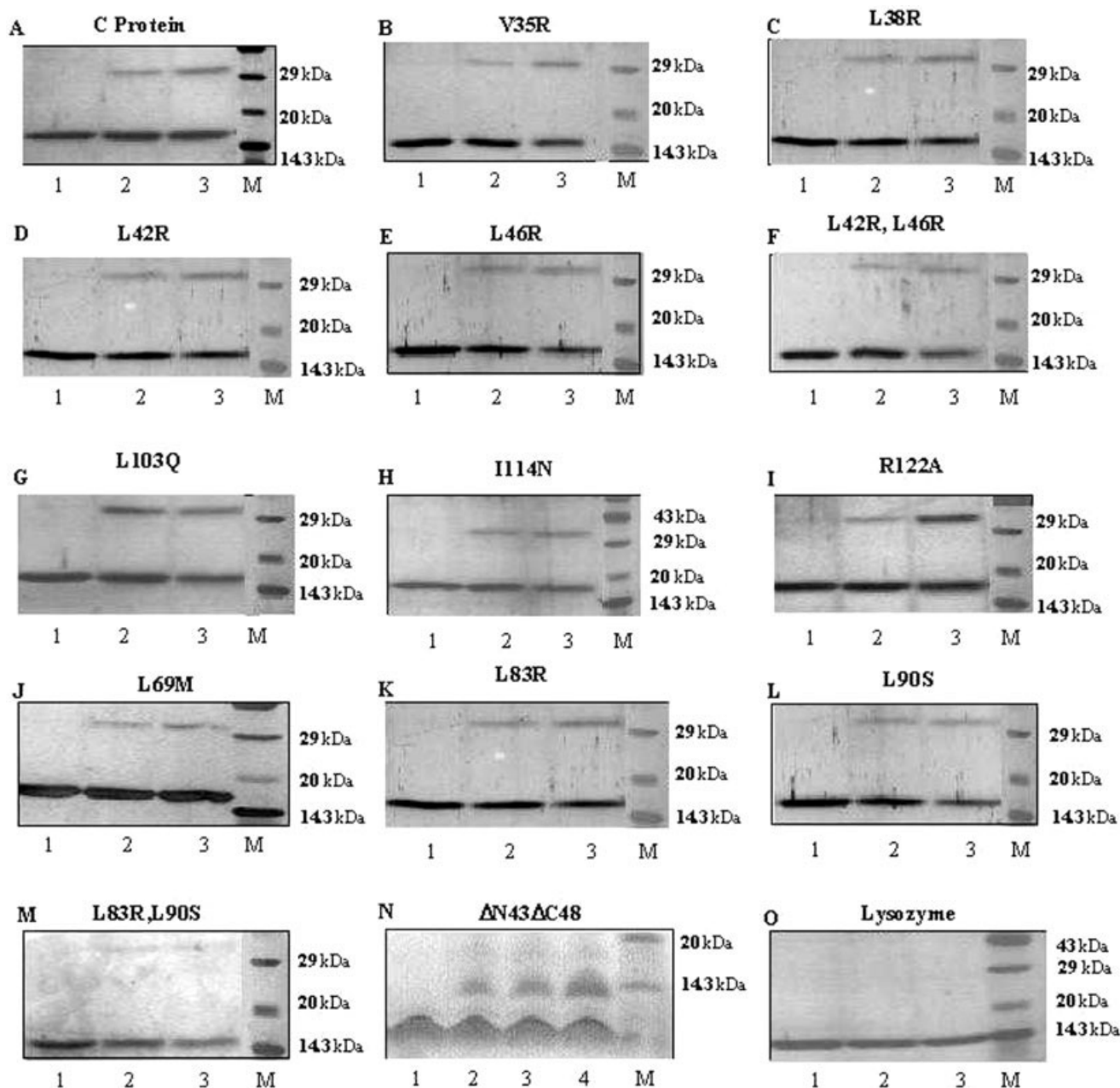


Fig. 3. Glutaraldehyde crosslinking of C protein and different mutants. Proteins were incubated with different indicated concentrations of glutaraldehyde for 30 min at room temperature, subjected to glycine SDS-PAGE, and visualized by silver staining (A–M and O). 1, Uncrosslinked; 2 and 3, treated with 0.0025% and 0.005% glutaraldehyde, respectively. In case of $\Delta N43\Delta C48$, after crosslinking, samples were subjected to 15% Tricine SDS-PAGE⁴⁰ and visualized by silver staining (N). 1, Uncrosslinked; 2–4, treated with 0.0025%, 0.003%, and 0.005% glutaraldehyde, respectively. The molecular weight markers indicated are lysozyme, 14.3 kDa; soybean trypsin inhibitor, 20 kDa; carbonic anhydrase, 29 kDa; and ovalbumin, 43 kDa.

replacement of each of the four K residues present in the protein did not affect either DNA binding or transactivation.² Mutations made elsewhere in the protein did not abolish DNA binding, indicating that the region comprising the HTH motif is the only DNA-binding region in C protein.

The interaction of C protein with its site on the *mom* promoter had earlier been studied using hydroxyl radical footprinting, which indicated that the protein bound to one face of the DNA double helix interacting at the successive

major grooves.¹¹ Three patches of protection were observed on each strand of the C-binding site, which were offset by 2–3 bp in the 3' direction. Such a pattern is characteristic of DNA binding mediated by an HTH motif.³⁵ Taken together, the mutational analysis and the footprinting data confirmed a role for the HTH motif in DNA binding. To get a better understanding of the additional residues of C protein that may be involved in DNA recognition, a detailed molecular modeling study was undertaken (see below).

Dimerization Interface

Glutaraldehyde crosslinking and molecular exclusion chromatography experiments carried out earlier showed that C exists as a dimer in solution.⁸ That the dimeric protein binds DNA was established by using gene fusion technology to construct staphylococcal protein A-C fusion and using the heterodimer of C and A-C fusion in mobility shift experiments.^{8,11} These results clearly ruled out DNA binding by monomeric C protein. Two regions in the C protein sequence depicted in Figure 1(B) are likely candidates for the dimerization interface. The first one (L32-D49), located toward the amino terminal of the protein, is a putative coiled coil motif, according to COILS algorithm.³⁶ A coiled coil is a heptad repeat of residues designated a–g, where the residues at positions a and d are hydrophobic, contributing to dimerization. The hydrophobic residues, V35, L38, L42, and L46, were mutated to nonconserved R to check for disruption of dimerization. In addition, a double mutant L42RL46R was also generated. The ability of the wild-type C and the mutant proteins to form dimers was assessed by chemical crosslinking assays. Under identical experimental conditions, C protein and the mutants formed dimers [Fig. 3(A–F)]. These results essentially rule out the possibility of a role for the putative coiled-coil motif in dimerization. *N*-Ethylmaleimide modification of C67, C78, and C93 residues (Materials and Methods) did not abolish either DNA binding or dimerization property of the protein (not shown). Parallely, the second heptad repeat [L69–L90, Fig. 1(B)], which may have a role in dimerization, was also subjected to mutational analysis. The hydrophobic residues L69, I76, L83, and L90 were changed to M, L, R, S, respectively. These mutants, L69M, I76L, L83R, and L90S behaved essentially like the wild-type protein both in their DNA binding as well as transactivation function, indicating that their dimerization status was unaltered (Table I). Glutaraldehyde crosslinking studies also showed that the dimerization is not affected in these mutants [Fig. 3(J–L)]. However, L83RL90S, a double mutant in the heptad repeat, exhibited compromised dimerization [Fig. 3(M)]. These results are reminiscent of the data obtained with several other dimeric proteins subjected to such mutational analysis. It has been demonstrated previously in case of GCN4 and fos-jun heterodimers that disruption of more than one residue of the zipper region is required to abolish dimerization.^{37,38} Further, mutation of two residues was necessary to demonstrate the function of MetR, a prokaryotic regulatory protein.³⁹ The results presented in Figure 3(M), which indicate the involvement of second heptad repeat of C protein in dimerization, were further substantiated by the dimerization ability of a protein fragment containing minimal heptad repeat. The deletant protein Δ N43 Δ C48 (see Materials and Methods) comprising of only 49 residues formed dimers assigning a role for the second heptad repeat region in dimerization [Fig. 3(N)].

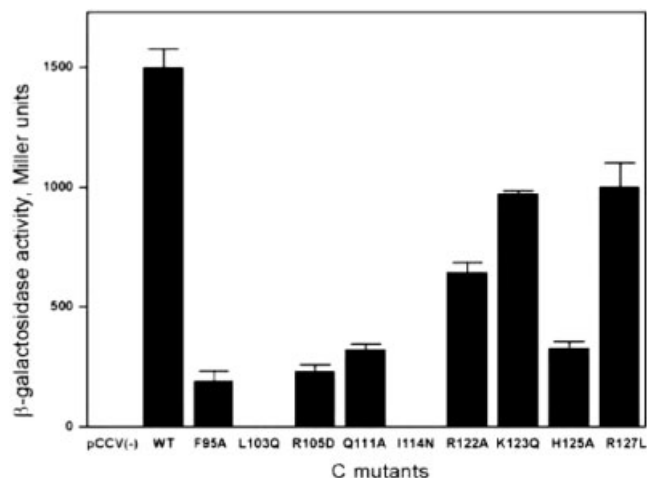


Fig. 4. Transactivation potential of C protein mutants. The values reported in Miller units are the mean of three independent experiments.

Transactivation Region of C Overlaps the HTH Motif

The transactivation function of C protein at P_{mom} was assayed using a reporter construct (see Materials and Methods). The HTH mutants, L103Q and I114N, were transactivation defective (Fig. 4), very likely because of their inability to bind DNA efficiently [Fig. 2(A)]. In contrast, F95A, which binds DNA [Fig. 2(C), lane 3], does not stimulate transcription from the *mom-lacZ* fusion construct (Fig. 4). Thus, it is a representative of positive control (pc) class of mutants.⁴¹ Mutants R105D and Q111A, which are within the HTH motif [Fig. 1(B)] bind DNA efficiently [Fig. 2(C), lanes 4 and 5] but exhibit reduced transactivation levels (Fig. 4). The carboxy terminal basic stretch mutants bound DNA with an affinity comparable to that of the wild type, yet transactivation of the *mom* promoter was reduced to varying degrees, indicating that these residues may also play a role in positive control (Fig. 4). Mutations in the N-terminal basic stretch did not affect DNA binding or transactivation. Among the transactivation defective mutants, F95A is located just before the HTH motif and R105D and Q111A map to the HTH motif, whereas mutants R122A, K123Q, H125A, and R127L are located immediately after the HTH motif [Fig. 1(B)]. Taken together, our results show that pc mutants are located within HTH and in its vicinity implying that transactivation region of C protein overlaps the DNA interaction region.

Modeling the Interaction of C Protein with Its Site on the *mom* Promoter

The secondary structure prediction of C protein based on the PHD algorithm¹⁶ is shown in Figure 5. The protein is predominantly α -helical, containing six putative helices (α 1– α 6). The three α -helices at the C-terminus (α 4– α 6) are connected by two short loops, thus forming a putative three-helix bundle.⁴² In this, helices 5 and 6 represent HTH motif confirmed by the mutational analysis described above. To further characterize the interaction of the C

PHD results

```

.....1.....2.....3.....4.....5.....6
AA      MQHDLFEHDPAIRQLIGHIDNIPAPELESRWPRSVVDLIDVLENEBKRONVSNPRELARK
SUB_sec LLL.LLLL.HHHHHHHHH.LLLLLLLLLL.HHHHHHHHHHHHHHH.LLLL.HHHH
Rel_sec 986455884999999986359997767677278999999999998747997527899

.....7.....8.....9.....10.....11.....12
AA      QAVALS CFLGGRQFYI POGDTIL TALRDDLLYCQFN GRNMBELRRQYRLSQPQIYQIIAR
SUB_sec HHHHHHH.LLL.EEELLL.HHHHH.HHHHH.LLLLHHHHHHH...HHHHHHHHH
Rel_sec 999999991898187559826999994599999937898899998623227999999999

.....13.....14
AA      QRKLHTRRHQFDLFSPEPK
SUB_sec HHHHHHHH.LLLLLLL
Rel_sec 99999898624368999999

```

Fig. 5. Secondary structure prediction of C protein by the PHD algorithm.¹⁶ The letters H, E, and L represent α -helix, extended β -strand and loop, respectively. The expected accuracy of prediction is >82%. Rel_sec indicates the reliability index for secondary structure prediction SUB_sec. A dot (.) indicates that no reliable prediction is made for the corresponding residue.

A)

```

C PROTEIN      sp|P06022  80 DTIL TALRDD LLYCQFN GRNMBELRRQYRL SQPQIYQIIA RQRKLHTRRH 129
prd PAIRED     sp|P06601  16 RPLPNNIRLK IVEMAADGIR PCVISRQLRV SHGCVSKILN RYQETGSIRP  65
NarL           sp|P10957 153 VNQLTPRERD ILKLIAGLP NKMIAARRLDI TESTVKVHVK HMLKKMKLKS 202
Prd HOMEODOMAIN sp|P06601 222 ASQLELERA FERTQYDIYT REELAQRNL TEARIQVWFS NRRARLRKQ- 281

```

(B)

```

      80      90      100      110      120
77  PCGDTIL TALRDDLLYCQFN GRNMBELRRQYRLSQPQIYQIIARQRKL 124
    |  :  :  |  :  .  |  :  ||  |  .  :  :  |  |
13  INGRPLPNNIRLKIVEMAADGIRPCVISRQLRVSHGCVSKILNRYQET  60

```

Fig. 6. Similarities between the C protein, NarL, and prd sequences. (A) The multiple sequence alignment between C protein, NarL, paired domain and homeodomain of prd protein using Clustal W.²⁵ The regions shown here (along with their SwissProt identification code) correspond to a three-helix bundle consisting of a typical HTH motif an additional α -helix preceding it. The nonpolar residues involved in the formation of the hydrophobic pocket and the glycine in the loop between the first two helices are shown in bold. In the alignment for homeodomain, P238 is shown as an insertion. (B) The pairwise alignment between 48-amino acid fragment from the C-terminal domain of the C protein and the N-terminal region of paired domain of prd. The GAP²⁴ algorithm was used for the alignment. The identical residues are connected by a vertical bar, whereas conservative substitutions are indicated by dots. The two proteins contain similar residues at many critical positions, which are indicated in bold. Overall there is 42% similarity and 18.2% identity. The secondary structure observed in the crystal structure of the prd domain bound to DNA is shown below the sequence.²⁸

protein with DNA, we adopted a modular approach (see Materials and Methods). The use of different fold recognition programs along with careful manual analysis indicates that the best match occurs between the fragment at the carboxy terminus consisting of the three helices connected by short loops and three helix bundle observed in the crystal structures of the NarL protein from *E. coli*,³⁰ paired domain,²⁸ and homeodomain²⁹ of prd protein from *Drosophila* [Fig. 6(A)]. The similarity is especially striking between a 48-residue fragment of C protein and the DNA binding N-terminal fragment of the prd paired domain

[Fig. 6(B)] from *Drosophila* prd (42% similarity and 18.2% identity), whose crystal structure along with its cognate DNA has been solved at 2.5 Å resolution (PDB code, 1pdn).²⁸ Several residues are highly conserved in the paired domains of various proteins as well as within the NarL family of proteins and are believed to be responsible for similarity of the three-helix fold in the DNA-binding domain of these proteins. The three hydrophobic residues I26/I163, I39/I176, and V 50/V 187 in the prd and NarL proteins [Fig. 6(A)] apparently form a hydrophobic cluster, which fixes the relative orientation of all three helices. The

loop between the two helices constituting the HTH motif could be held in place by V45/I182, whereas a glycine in the loop connecting the first two helices in both the proteins (G33/G170) takes up a glycine-specific conformation in the Ramachandran map, with (ϕ, ψ) being approximately $(88^\circ, 10^\circ)$ to ensure a proper orientation of the first two helices.^{30,28} We find a striking similarity in the C protein sequence, which contains three nonpolar residues, L90, L103, and I114, at identical locations to the I/I/V triad in prd and NarL proteins, whereas L109 occurs in the turn constituting the HTH motif and a glycine (G97) occupies the loop position between the first two helices [Fig. 6(B)].

The structure of prd paired domain-DNA complex revealed that in addition to the HTH motif, a type II β -turn preceding the three-helix bundle was also involved in binding to the minor groove of DNA. This type II β -turn (I13-R16) includes G15 in a glycine-specific conformation with (ϕ, ψ) being $(100^\circ, -31^\circ)$ and whose carbonyl oxygen forms a hydrogen bond with the 2-amino group of a guanine in the DNA. C protein also has a glycine (G79) at an identical position. Hence we mapped the 48-residue C protein sequence, shown in Figure 6(B), to the prd paired domain crystal structure and superimposed a 10-bp region from the Mu *mom* DNA sequence [from -53C to -44C, out of the C-binding site shown in Fig. 1(A)] to the double-helical DNA of the prd-DNA complex. An energy minimized, stereochemically acceptable model showing a 48-amino acid fragment of C, spanning a region between P77 to L124 (which overlaps the I13 to T60 fragment of prd, comprising the β -turn and the three-helix bundle) docked with a DNA fragment from -53C to -44C is shown in Fig. 7(A). The amino acid residues of C can be superimposed on the prd protein-DNA complex with several critical hydrogen bonds and van der Waals contacts being retained. In particular, amino acids G15, R23, R41, S46, and R56, which interact with DNA, either directly or through water-mediated hydrogen bonds in prd, have corresponding residues G79, R87, R105, S110, and R120 in identical positions in C. These similarities are in addition to the conserved features described above [Fig. 6(A,B)], which are essential for the formation of the three-helix bundle. The hydrophobic pocket formed by L90, L103, and I114 from helices 4, 5, and 6, respectively, is depicted in Figure 7(B).

Validation of the Model

The model depicted in Figure 7 assigns critical structure determining roles to G79 and G97. The G79 is predicted to occur in a type-II β turn preceding the three-helix bundle and can directly interact with the minor groove of DNA, whereas G97 is part of a turn between the first two helices and is essential for maintaining the geometry of the three-helix bundle. G79A and G97A mutants were generated and tested for their DNA-binding ability. As predicted by the model, both the mutants, G79A as well as G97A did not bind to DNA (Fig. 8). Mutants L103Q and I114N, which would also affect the formation of the three helix bundle, do not bind DNA [Fig. 2(A)]. However the mutant L90S, involving the hydrophobic residue from the first

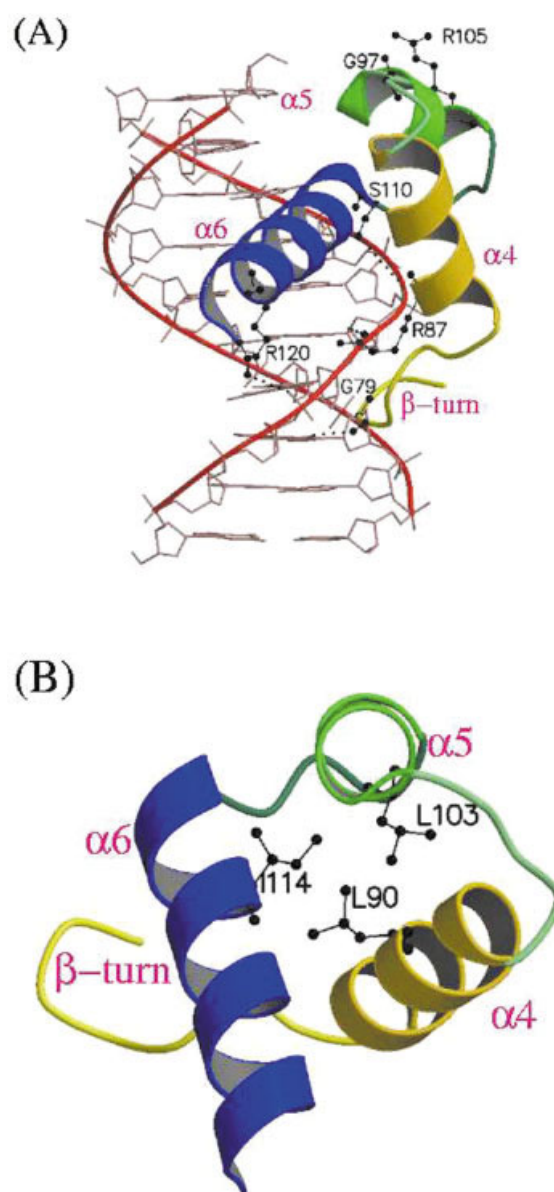


Fig. 7. Energy minimized model structure showing the interaction of C protein with its DNA-binding site. (A) The C-terminal region of C protein (P77 to L124), which is predicted to take up a β -turn conformation, followed by a three-helix bundle structure (according to the prediction shown in Fig. 5 and observed in prd protein) is shown docked with a 10 mer duplex DNA (-53C to -44C). The C protein has been modeled using the crystal structure of the paired domain²⁸ as a template. The protein structure is shown as a ribbon diagram, with only residues critical for the formation of the structure or those involved in hydrogen bonding to DNA being shown. The DNA is in red and the three helices numbered $\alpha 4$, $\alpha 5$ and $\alpha 6$ (as in Fig. 6) are colored yellow, green, and blue, respectively. The direct hydrogen bonds that can form between the protein and DNA in the model structure are shown by broken lines. (B) The DNA-binding fragment of C protein as viewed down the middle helix ($\alpha 5$) of the three-helix bundle. It shows the structurally important hydrophobic pocket formed by conserved nonpolar residues L90, L103, and I114 from the three helices.

helix of the three-helix bundle, binds DNA, probably because the small size of the mutated residue does not disrupt the overall structure. In addition to the above

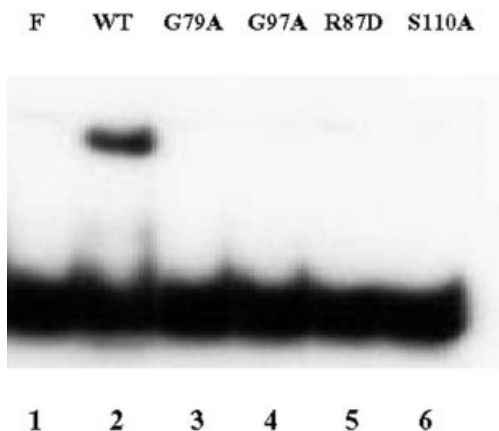


Fig. 8. DNA-binding activities of the mutants G79A, G97A, R87D, and S110A. A 25-bp duplex C binding site was incubated with the wild-type and mutant proteins and electrophoresed on a 10% native polyacrylamide gel as described in Materials and Methods.

residues, the model also identifies roles for R87, S110, and R120 in DNA recognition. These residues could be directly or indirectly involved in hydrogen bonding with DNA. R87 occurs in the first helix of the three-helix bundle, whereas S110 occupies the third position in the turn of the HTH motif. R120 is present near the C-terminus of the HTH motif. In addition, Q111 is present at a position occupied by H47 in *prd* protein involved in DNA recognition. However, Q111 is located slightly farther away in our model. The predictions for structural and/or functional role of the above residues were directly tested by constructing the mutants, R87D, R120D, S110A, and Q111A. The mutant proteins were expressed and assayed for their DNA-binding activity as described earlier. Among these mutants, only Q111A bound at C recognition sequence [Figs. 2(B) and 8 and Table I], validating the model presented in Figure 7.

DISCUSSION

In this article, we have dissected the functional domains of late gene transcription activator protein C of phage Mu. The results demonstrate that an extended HTH motif, consisting of a three-helix bundle preceded by a β -turn, is responsible for DNA binding with amino acids in close proximity and overlapping the motif contributing to the transactivation process.

HTH is a well-characterized motif found in a large number of DNA-binding proteins involved in regulation of gene expression. Thus it is not surprising to find variations in the architecture of this motif to accommodate subtle differences in interaction of different proteins with their target sequences. The motif generally forms a part of a domain with varying number of helices. For instance, HTH in the λ -repressor and the 434 repressors is a part of a four-helix bundle, whereas in λ -Cro and catabolite activator protein (CAP), the motif is a part of a three-helix bundle closed off by a β -sheet.⁴³ Based on the way additional helices pack against the HTH motif, different types of HTH domains can be distinguished. Despite the diver-

sity among the HTH DNA-binding motifs, residues in certain positions (4 and 15) are highly conserved.^{33,34} These residues are hydrophobic amino acids and are buried in the interior of the motif. Although these most conserved residues are retained in C, changes are seen at other positions. Instead of A or G, which are generally preferred at position 5 and G at position 9, R is present at both positions in C protein [Fig. 1(B)]. A point to be noted is that HTH motifs in many proteins do not follow all the rules. For instance, K is found at the fifth position of the motif in Trp repressor,³³ whereas bacteriophage Mu Ner has a S residue in that position.³⁴ The σ^{70} of *E. coli* and σ^{54} of *Klebsiella pneumoniae* have D and E at the ninth position of their respective HTH motifs.^{44,45} The σ^{28} and σ^{30} of *Bacillus subtilis* have N at this position.⁴⁶ Further, the paired domain of eukaryotic developmental regulatory proteins, *viz.* the *prd* protein, do not contain G in the turn region of the HTH motif. Thus, sequence variations in the HTH motif are not uncommon. Given the diversity in sequences of binding sites, variations in the HTH motif appear to have evolved to suit individual regulatory requirements.

Mutations in amino acid residues within and in the vicinity of the HTH of C protein, when mutated, result in a reduced level of expression of the reporter gene as in the case of many other transactivators. For example, the residues present adjacent to the HTH motif in FIS are required for transactivation at the *proP2* promoter.⁴⁷ In the λ -repressor, an acidic patch on the surface of the HTH is involved in transcription activation.⁴⁸ The amino acid residues identified to participate in transactivation of *P_{mom}* could influence the process either indirectly or directly. The residues could indirectly stimulate transcription activation by influencing mode of interaction *viz.*, DNA distortion mediated by C protein.¹¹⁻¹³ Alternately, there could be a direct contact of these residues with one of the subunits of the *E. coli* RNA polymerase. It has been shown previously that neither the alpha-CTD nor sigma, which are the sites of interaction of most transactivators, contact C protein.⁴⁹ Although the β and β' subunits have been identified as other targets of activators,^{50,51} the interaction of C with either of these subunits, if any, is yet to be established.

Modeling of the DNA-binding region of the C protein using a combination of sequence and structural homology, revealed that the HTH motif is part of a three-helix bundle. This bundle shared regions of similarity with the *prd* protein from *Drosophila*²⁸ and the NarL protein from *E. coli*.³⁰ Apart from other common features, the C protein has three hydrophobic residues at positions identical to the nonpolar triad in the *prd* and NarL proteins. A 48-amino acid sequence of C could replace the DNA-binding region of the *prd* protein in the *prd* protein-DNA complex, with minimal alteration in the side chain conformations. The docking of the β -turn and three-helix bundle with DNA fragment, as shown in Figure 7(A), neatly explains several experimental observations. For example, the nesting of the β -turn in the minor groove could protect the GC-rich spacer region (-46G to -42C) from hydroxyl

radicals [Fig. 1(A)]. The characteristic three patches of protection in each strand of DNA observed in case of C is indicative of a dimeric protein binding to one face of the DNA with monomers contacting the DNA in successive major grooves and each other across the minor groove.³⁵ The N-terminal residues in the recognition helix ($\alpha 6$) can provide protection to -52G in the top strand, whereas the same residues in the other monomer could protect the -35G and -34G in the bottom strand. These residues were shown to be contacted by C protein by protection and interference analysis.^{11–13} In addition, mutation of L103 and I114 to nonpolar residues would destabilize the HTH motif and hence DNA binding. Mutation of the residues G97 and G79 would also disrupt the characteristic structural motif and affect DNA binding. Although R87D and R120D would lead to unfavorable interactions with the DNA backbone, the R105D as well as mutations beyond R120 will not affect DNA binding because these residues do not contact DNA. The results of mutational analysis support all these predictions from the model.

The striking similarity in the DNA-binding regions of the prd protein, NarL, and C protein is rather intriguing considering their diverse origin. Although the first protein is a eukaryotic regulator involved in developmental processes in *Drosophila*, the latter two transactivators are of prokaryotic origin carrying out different functions. The paired protein binds to DNA as a monomer, whereas C binds as a dimer. The NarL protein is also binds as a dimer.⁴⁸ Among the three proteins, C protein binds to DNA with high affinity⁹ and untwists DNA in order to facilitate RNAP recruitment to the promoter.¹³ A parallel to this mode of interaction is seen with the paired homeodomain interaction at its target site. Binding of prd homeodomain monomer to its site distorts DNA to recruit the second monomer.²⁹ Docking of the NarL protein onto the prd-DNA complex reveals that the full molecule of NarL, as seen in the crystal structure, cannot directly interact with DNA because the amino terminal domain of the protein would block the DNA-binding domain present at the carboxy terminal region, from contacting it.³⁰ A conformational change mediated by a Mg^{2+} -dependent phosphorylation is probably responsible for removing this block. The C protein also undergoes a mandatory Mg^{2+} -induced conformational change to bind DNA,¹⁰ which could involve a reorientation of the N- and C-terminal domains. In spite of their diverse cellular functions, these proteins share certain characteristics. The conserved HTH architectures observed in these proteins seem to reflect features conserved in their interaction with target DNA during the transactivation process.

CONCLUSIONS

We have identified the DNA-binding domain of the transcriptional activator protein C of phage Mu using secondary structure predictions, molecular modeling, and site-directed mutagenesis. Our studies indicate that a HTH motif toward the carboxy terminal end of the protein is the DNA-binding domain and that it is a part of a three-helix bundle. Structure-based homology modeling

using the prd domain of the paired protein of *Drosophila* as a template revealed a role for additional residues in DNA recognition. Mutagenesis of these residues further supported the model and provides a structural basis for understanding the recognition of *mom* promoter by C protein. In addition, residues involved in the transactivation function of C protein were also identified and these residues map within or close to the HTH motif.

ACKNOWLEDGMENTS

We thank S. Hattman for plasmid pCCV(-), D.R. Radha for technical assistance, and the reviewer for many suggestions. Bioinformatics and Phosphorimager facility at the Indian Institute of Science are acknowledged. B.D.P. was a recipient of the postdoctoral fellowship from the Department of Science and Technology, and A.K. and A.C. are recipients of the senior research fellowships from the University Grants Commission and Council of Scientific and Industrial Research, Government of India, respectively. The work was supported by a grant (to V.N) from the Department of Science and Technology, Government of India.

REFERENCES

- Hattman S, Ives J, Margolin W, Howe MM. Regulation and expression of the bacteriophage Mu *mom* gene: mapping of the transactivation (*dad*) function to the C region. *Gene* 1985;39:71–76.
- Hattman S. Unusual transcriptional and translational regulation of the bacteriophage Mu *mom* operon. *Pharmacol Ther* 1999;84:367–388.
- Toussaint A. The DNA modification function of temperate phage Mu-1. *Virology* 1976;70:17–27.
- Kahmann R, Seiler A, Wulczyn FG, Ptaff E. The *mom* gene of bacteriophage Mu: a unique regulatory scheme to control a lethal function. *Gene* 1985;39:61–70.
- Balke V, Nagaraja V, Gindlesperger T, Hattman S. Functionally distinct RNA polymerase binding sites in the phage Mu *mom* promoter region. *Nucleic Acids Res* 1992;20:2777–2784.
- Sun W, Hattman S. Bidirectional transcription in the *mom* promoter region of bacteriophage Mu. *J Mol Biol* 1998; 284:885–892.
- Harley CB, Reynolds RP. Analysis of *E. coli* promoter sequences. *Nucleic Acids Res* 1987;15:2343–2361.
- De A, Paul BD, Ramesh V, Nagaraja V. Use of protein A gene fusions for the analysis of structure-function relationship of the transactivator protein C of bacteriophage Mu. *Protein Eng* 1997;10: 935–941.
- Ramesh V, De A, Nagaraja V. Engineering hyperexpression of bacteriophage Mu C protein by removal of secondary structure at the translation initiation region. *Protein Eng* 1994;8:1053–1057.
- De A, Ramesh V, Mahadevan S, Nagaraja V. Mg^{2+} mediated sequence-specific binding of transcriptional activator protein C of bacteriophage Mu to DNA. *Biochemistry* 1998;37:3831–3838.
- Ramesh V, Nagaraja V. Sequence-specific DNA binding of the phage Mu C protein: footprinting analysis reveals altered DNA conformation upon protein binding. *J Mol Biol* 1996;260:22–33.
- Sun W, Hattman S, Kool E. Interaction of the bacteriophage Mu transcriptional activator protein, C, with its target site in the *mom* promoter. *J Mol Biol* 1997;273:765–774.
- Basak S, Nagaraja V. Transcriptional activator C protein-mediated unwinding of DNA as a possible mechanism for *mom* gene activation. *J Mol Biol* 1998;284:893–902.
- Zhou YH, Zhang XP, Ebright RH. Random mutagenesis of gene-sized DNA molecules by use of PCR with *Taq* DNA polymerase. *Nucleic Acids Res* 1991;19:6052.
- Miller J. *Experiments in molecular genetics*. New York: Cold Spring Harbor Laboratory Press; 1972.
- Rost B, Sander C. Combining evolutionary information and neural

- networks to predict protein secondary structure. *Proteins* 1994;19:55–72.
17. Alexandrov NN, Nussinov R, Zimmer RM. Fast protein fold recognition via sequence to structure alignment and contact capacity potentials. In: Hunter L, Klein TE, editors. Pacific Symposium on Biocomputing '96, Singapore: World Scientific Publishing Co.; 1996. p 53–72.
 18. Rost B. TOPITS: threading one-dimensional predictions into three-dimensional structures. In: Rawlings C, Clark D, Altman R, Hunter L, Lengauer T, Wodak S, editors. The Third International Conference on Intelligent Systems for Molecular Biology (ISMB), Cambridge, England, Menlo Park, CA: AAAI Press; 1995. p 314–321.
 19. Salwinski L, Eisenberg D. Motif-based fold assignment. *Protein Sci* 2001;10:2460–2469.
 20. Kelley LA, MacCallum RM, Sternberg MJE. Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J Mol Biol* 2000;299:499–520.
 21. Fischer D, Barret C, Bryson K, Elofsson A, Godzik A, Jones D, Karplus KJ, Kelley LA, MacCallum RM, Pawowski K, Rost B, Rychlewski L, Sternberg MJ. CAFASP-1: critical assessment of fully automated structure prediction methods. *Proteins* 1999; suppl 3:209–217.
 22. Kelley LA, MacCallum R, Sternberg MJE In: Istrail S, Pevzner P, Waterman M, editors. RECOMB 99, Proceedings of the Third Annual Conference on Computational Molecular Biology. New York: The Association for Computing Machinery; 1999. p 218–225.
 23. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* 1995;247:536–540.
 24. Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* 1970;48:443–453.
 25. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22:4673–4680.
 26. Myers EW, Miller W. Optimal alignments in linear-space. *Comput Appl Biosci* 1988;4:11–17.
 27. Nevill-Manning CG, Huang CN, Brutlag DL. Pairwise protein sequence alignment using Needleman-Wunsch and Smith-Waterman algorithms. Personal communication, 1997.
 28. Xu W, Rould MA, Jun S, Demplan C, Pabo CO. Crystal structure of a paired domain-DNA complex at 2.5 Å resolution reveals structural basis for Pax developmental mutations. *Cell* 1995;80:639–650.
 29. Wilson DS, Guenther B, Desplan C, Kuriyan J. High resolution crystal structure of a paired (Pax) class cooperative homeodomain dimer on DNA. *Cell* 1995;82:709–719.
 30. Baikalov I, Schroder I, Kaczor-Grzeskowiak M, Grzeskowiak K, Gunsalus RP, Dickerson RE. Structure of the *Escherichia coli* response regulator NarL. *Biochemistry* 1996;35:11053–11061.
 31. Bansal M, Bhattacharyya D, Ravi B. NUPARM and NUCGEN: software for analysis and generation of sequence dependent nucleic acid structures. *CABIOS* 1995;11:281–287.
 32. Perlman DA, Case DA, Caldwell JW, Ross, WS, Cheathan TE III, Ferguson DM, Seibal GL, Singh P, Weiner P, Kollaman PA. AMBER Program Package, San Francisco, CA: UCSF.
 33. Brennan RG, Matthews BW. The helix-turn-helix DNA binding motif. *J Biol Chem* 1989;264:1903–1906.
 34. Dodd IB, Egan JB. Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. *Nucleic Acids Res* 1990;18:5019–5026.
 35. Dixon WJ, Hayes JJ, Levin JR, Weidner MF, Dombroski BA, Tullius TD. Hydroxyl radical footprinting. *Methods Enzymol* 1991;208:380–413.
 36. Lupas A, Van Dyke M, Stock J. Predicting coiled coils from protein sequences. *Science* 1991;252:1162–1164.
 37. van Heeckeren WJ, Sellers JW, Struhl K. Role of the conserved leucines in the leucine zipper dimerization motif of yeast GCN4. *Nucleic Acids Res* 1992;20:3721–3724.
 38. Kouzarides T, Ziff E. The role of the leucine zipper in the fos-jun interaction. *Nature* 1988;336:646–651.
 39. Maxon ME, Wigboldus J, Brot N, Weissbach H. Structure-function studies on *Escherichia coli* MetR protein, putative prokaryotic leucine zipper protein. *Proc Natl Acad Sci USA* 1990;87:7076–7079.
 40. Schagger H, von Jagow G. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Biochem* 1987;166:368–379.
 41. Ptashne M. A Genetic switch: phage λ and higher organisms. Cambridge, MA: Cell-Blackwell Scientific; 1994.
 42. Wintjens R, Rooman M. Structural classification of HTH DNA-binding domains and protein-DNA interaction modes. *J Mol Biol* 1996;262:294–313.
 43. Harrison SC, Aggarwal AK. DNA recognition by proteins with the helix-turn-helix motif. *Annu Rev Biochem* 1990;59:933–969.
 44. Siegele DA, Hu JC, Walter WA, Gross CA. Altered promoter recognition by mutant forms of the sigma 70 subunit of *Escherichia coli* RNA polymerase. *J Mol Biol* 1989;206:591–603.
 45. Coppard JR, Merrick MJ. Cassette mutagenesis implicates a helix-turn-helix motif in promoter recognition by the novel RNA polymerase sigma factor σ 54. *Mol Microbiol* 5;1991:1309–1317.
 46. Helmann JD, Chamberlin MJ. Structure and function of bacterial sigma factors. *Annu Rev Biochem* 1988;57:839–872.
 47. McLeod S., Xu J, Cramton SE, Gaal T, Gourse RL, Johnson RC. Localization of amino acids required for Fis to function as a class II transcriptional activator at the RpoS-dependent proP P2 promoter. *J Mol Biol* 1999;294:333–346.
 48. Li M, Moyle H, Susskind MM. Target of the transcriptional activation function of phage λ cI protein. *Science* 1994;263:75–77.
 49. Sun W, Hattman S, Fujita N, Ishihama A. Activation of bacteriophage Mu *mom* transcription by C protein does not require specific interaction with the carboxyl-terminal region of the α or σ^{70} subunit of *Escherichia coli* RNA polymerase. *J Bacteriol* 1998;180:3257–3259.
 50. Szalewska-Palesz A, Wegrzyn A, Blaszcak A, Taylor K, Wegrzyn G. DnaA-stimulated transcriptional activation of *ori λ* : *Escherichia coli* RNA polymerase β subunit as a transcriptional activator contact site. *Proc Natl Acad Sci USA* 1998;95:4241–4246.
 51. Miller A, Wood D, Ebright RH, Rothman-Denes LB. RNA polymerase β' subunit: a target of DNA binding-independent activation. *Science* 1997;275:1655–1657.
 52. Baikalov I, Schroder I, Kaczor-Grzeskowiak M, Cascio D, Gunsalus RP, Dickerson RE. NarL dimerization? Suggestive evidence from a new crystal form. *Biochemistry* 1998;37:3665–3676.