The Helix-Turn-Helix Motif of the Coliphage 186 Immunity Repressor Binds to Two Distinct Recognition Sequences*

Received for publication, August 13, 2001, and in revised form, October 26, 2001 Published, JBC Papers in Press, November 7, 2001, DOI 10.1074/jbc.M107740200

Keith E. Shearwin‡, Ian B. Dodd, and J. Barry Egan
From the Department of Molecular Biosciences, University of Adelaide, Adelaide, Australia 5005

The CI protein of coliphage 186 is responsible for maintaining the stable lysogenic state. To do this CI must recognize two distinct DNA sequences, termed A type sites and B type sites. Here we investigate whether CI contains two separate DNA binding motifs or whether CI has one motif that recognizes two different operator sequences. Sequence alignment with 186-like repressors predicts an N-terminal helix-turn-helix (HTH) motif, albeit with poor homology to a large master set of such motifs. The domain structure of CI was investigated by linker insertion mutagenesis and limited proteolysis. CI consists of an N-terminal domain, which weakly dimerizes and binds both A and B type sequences, and a C-terminal domain, which associates to octamers but is unable to bind DNA. A fusion protein consisting of the 186 N-terminal domain and the phage λ oligomerization domain binds A and B type sequences more efficiently than the isolated 186 CI N-terminal domain, hence the 186 C-terminal domain likely mediates oligomerization and cooperativity. Site-directed mutagenesis of the putative 186 HTH motif eliminates binding to both A and B type sites, supporting the idea that binding to the two distinct DNA sequences is mediated by a variant HTH motif.

DNA binding proteins are often modular in structure, with separate domains responsible for binding and oligomerization. Such an arrangement, even in a simple system such as the lysis-lysogeny switch of coliphage λ, permits remarkable control of gene expression through a series of thermodynamically linked protein-protein and protein-DNA interactions. The λ genetic switch, one of the most intensively studied systems in biology, has in many ways provided the basis for the study of switch biology in higher organisms (1, 2). Coliphage 186, a member of the P2 family of phage, shows essentially no similarity with λ at the protein or DNA level, and so the two have presumably evolved independently of each other. Nevertheless, the lysis-lysogeny switches of each show superficial similarity, and it is expected that a comparative analysis of the differences in detail between the two will improve overall understanding of switch operation, hence the present structure-function study of 186 CI repressor.

The immunity repressor, CI, of coliphage 186 is responsible for maintenance of the stable lysogenic state and achieves this by binding directly over and repressing two promoters; pR, the promoter of the early lytic operon and pB, the promoter for the late promoter activator gene B (Fig. 1). 186 contains a total of four binding sites for CI, including two sites (F1 and F2) that flank the lytic promoter (6). These flanking sites play a role fine-tuning CI-regulation of transcription from pR and from the lysogenic promoter p1. The flanking sites F1 and F2 each consist of an inverted repeat, while the CI binding site at the pR promoter contains a pair of inverted repeats. These four inverted repeats share sequence similarity and are separated in each case by a five-base pair A/T-rich spacer. These CI binding sites have been designated A type sites (Fig. 1). The CI binding site at pR, responsible for repressing the early lytic genes, consists of three inverted repeats. There is a central A type site, which has a four- rather than five-base pair spacing between conserved bases and is designated an A′ site. Situated on either side of the A′ site are inverted repeats, again with an A/T-rich spacer, but unrelated in sequence to the A type sites (see Ref. 6, Fig. 1). These alternative recognition elements have been termed B type sites. Hence, the CI binding site at pR has the arrangement B-A′-B. The recognition elements at pR all lie on the same face of the helix and are strongly supported by DNase I footprint data and by a library of 19 virulent (vir) mutations (6, 7). Thus, CI is able to recognize two distinct DNA sequences.

In the present work, we set out to determine whether there are two distinct DNA binding regions within the CI protein or whether there is just one motif that binds with relaxed specificity to the two different types of binding sites. To this end we have investigated the domain structure of CI by sequence analysis, linker insertion mutagenesis, and limited proteolysis. We have examined the self-association and DNA binding properties of the isolated domains and from the information so obtained carried out mutagenesis on residues predicted to be critical for DNA binding.

EXPERIMENTAL PROCEDURES

Bacteria BL21 (DE3) pLysS was used as the host for expression of CI and its variants from the T7 promoter of the pET3a vector (8). NK7049 (9) was used as the host for lacZ reporter genes. NK7049 (ARS45 pR short lacZ) was used to assay transposon insertion mutants. NK7049 (ARS45 pR HincII/SnaBI lacZ) was used to assay the ability of CI and its variants to repress pR, and NK7049 (ARS45 pR lacZ) was used to assay pR activity.

Plasmids, bacteriophage pRASI1 was created by amplifying the 186 cI gene, including the native ribosome binding site by PCR using primers 108 and 55, digesting the PCR product with EcoRI and BamHI and ligating into pBluescriptKS (Stratagene), which had been digested with the same enzymes.

pET3aHis6 is a derivative of pET3a, which is more sensitive to CI-encoded C-terminal thrombin cleavage sequence (LVPRGS) followed by a six-histidine affinity tag. pETCIHis6 was created by amplifying the cI gene from pET3aCI (10) with primers T7 and 129, digestion of the PCR product with Nde and Sau3A, and ligation into pET3aHis6.

This paper is available online at http://www.jbc.org

† To whom correspondence should be addressed: Dept. of Molecular Biosciences, Univ. of Adelaide, North Terrace, Adelaide, South Australia, Australia 5005. Tel.: 61-8-8303-5362; Fax: 61-8-8303-4348; E-mail: keith.shearwin@adelaide.edu.au.

‡ This work was supported by the Australian Research Council. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

1 I. B. Dodd and J. B. Egan, manuscript in preparation.
Domain Structure and Function of Coliphage 186 Repressor

Fig. 1. Map of the control region of coliphage 186. The organization of the early control region of the 186 genome is shown, with the region from the PstI site (sequence coordinate 20,315) to the end of the cII gene (coordinate 23,943) enlarged to show details (3, 4, 5). Sequence numbering starts at the left cos end of the 186 genome. Genes are shown as gray boxes: B, activator of late transcription; 69, unknown function; int, integrase; cl, immunity repressor; apl, excisionase and transcriptional control; cII, establishment of lysogenicity. Promoters are shown as arrowheads, their transcripts as arrows, terminators as stem loops, and the phage attachment site attP as a solid box. The CI binding sites at pR, F1, pS1, and F4 are indicated as solid circles. The sequences of the inverted repeats from each of these sites are shown in the lower part of the figure. The diamonds indicate the center of symmetry of each inverted repeat. F1 and F4 each consist of one A type site, pS2 of two A type sites with the arrangement B-A′-B′-C. The consensus sequences for A type and B type sites are shown, where w = A or T, y = C or T and r = A or G. The central w in the A type consensus is optional, reflecting the alternate spacing of A and A′ type sites.

which had been digested with the same enzymes. pETCI (1–82)His6 and pETCI (83–196)His6, were constructed similarly to pETCIHis6, using primers designed to introduce Ndel and SacI sites at the appropriate positions.

The CI/(Hybrid)His6 was constructed by amplifying by PCR the region encoding amino acids 1–82 of 186 cl using a 3′-primer designed to introduce an SpHl site. The region of the λ cI gene encoding amino acids 93–236 of λ repressor was amplified by PCR using a 5′-primer designed to introduce an SpHl site. The 186-cl PCR product was digested with Ndel and SpHl, the λ cI PCR product with SpHl and SacII, and these fragments inserted into Ndel/SacII-digested pET3aHis, pETCIHT18, plasmid DNA isolated from the BL21 pLysS containing pRAS1 was used in an in vitro reaction (5 µl) containing target DNA (pRAS1, 20 ng), donor DNA (4.2-kb Transprimer pGP5 carrying the kanamycin resistance gene, 5 ng) and transposase protein. After a 1-h strand transfer reaction at 37 °C, the enzyme was inactivated by heating to 75 °C for 10 min. Aliquots of the mixture were transformed by electroporation into NK7049 (λ pR short lacZ) and plated on LB plates containing kanamycin (20 µg/ml), ampicillin (100 µg/ml), and X-gal (20 µg/ml) to select for transformants carrying the transposon within pRAS1. Blue colonies indicated the potential presence of a transposon within the CI gene, since CI was rendered unable to repress the pR lacZ reporter. These potential insertion mutations were further screened by a PCR-based assay. Plasmid DNA was isolated from those strains in which the transposon was confirmed as being within the CI gene. The plasmid DNA was digested with Pmel to remove the bulk of the transposon, the plasmids religated and retransformed into strain NK7049 (λ pR short lacZ) for genomic transplantation and survival of the resulting plasmids was determined by DNA sequencing. The effects of the mutations were assayed in the same strain by measuring β-galactosidase activity.

The abbreviations used are: PMSF, phenylmethylsulfonyl fluoride; X-gal, 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside; SFR, surface plasmon resonance; RU, response units; HTH, helix-turn-helix.

GTGAAATCTCGAGGGGGG. (CI binding sites are shown in bold). B32B top: AATCCGGTACTTTGCAAGGGGCGTACATTCGACAGGTCGTCTTCCGAGGATATCGATATCATGTAATACCTTGAAACCCACCCGCCAACATCTAG. (CI binding sites are shown in bold). B32B bottom biotin: BiotinTGAATGATGGGTTTATCTGAAAGCTCTGAGGGGCGTAATTCGACAGGTCGTCTTCCGAGGATATCGATATCATGTAATACCTTGAAACCCACCCGCCAACATCTAG. (CI binding sites are shown in bold). OLI-OLI: OLI-CCTGCTGATATCGATATCATGTAATACCTTGAAACCCACCCGCCAACATCTAG. (CI binding sites are shown in bold). OLI-OLI bottom biotin: BiotinTGAATGATGGGTTTATCTGAAAGCTCTGAGGGGCGTAATTCGACAGGTCGTCTTCCGAGGATATCGATATCATGTAATACCTTGAAACCCACCCGCCAACATCTAG.
**Domain Structure and Function of Coliphage 186 Repressor**

**Limited Proteolysis**—CHI68 (1.1 mg/ml) in TEG150 was digested at 37 °C at CI to protease molar ratios of 280 (subtilisin), 370 (papain), or 1100 (proteinas K). At appropriate time points, 5 μl of samples were diluted into an equal volume of 2× SDS loading buffer containing 20 mM PMFS, immediately heated to 95 °C for 1 min, and analyzed by SDS-PAGE. Samples to be analyzed were run on a 12% gel at a flow rate of 5 μl/min. Standards containing bovine serum albumin and myoglobin (20 μl) were stopped by the addition of PMFS and heating. Electrospray mass spectrometry was kindly performed by Dr. C. Bagley, Institute of Medical and Veterinary Science, Adelaide, Australia. Cleavage points were deduced from mass spectrometry results using the PAWS program.

**Monomeric Single Copy Lac Z Fusions—**Strain NK7049 transformed with the appropriate pMR98 derivative was used as the host for growth of the ARS45 phage vector. Phage stocks obtained were plated on NK7049, and single recombinant planks selected on the basis of color in the presence of X-gal and purified once by streaking across a lawn of NK7049. Independent blue lysogens from at least two recombinant planks were purified by restreaking. Single copy status of these lysogens was confirmed by PCR (12). For assay of βp, or βp(β-galactosidase activity, the appropriate CI expression plasmid (pETCHI68 or the parental pET3a plasmid) was transformed into these lysogens, and liquid cultures started from single colonies.

Kinetic LacZ assays were done in 96-well microtiter plates by an extensively modified Miller method (13). Fresh colonies on selective LB plate were grown in 200 μl of TEG150 + antibiotic. Plates were sealed and incubated at 37 °C for ~16 h without shaking. These cultures were subcultured by diluting 2 μl into 98 μl of fresh medium and incubated with rotation to an A600 of 0.2–1.2 (log phase). A600 was measured using a Labsystems Multiskan Ascent plate reader with a 620-nm filter; the A620 values were converted to A600 (1-cm path length) values using an empirically derived relationship and adjusted for light-scattering non-linearity according to (14). Cells were chilled and then permeabilized with polyvinyl B (15) by adding 20 μl of culture + 30 μl of LB to 150 μl of lysis buffer in a microtiter plate. Lysis buffer was T28 (100 mM Tris-HCl, pH 8.0, 1 mM MgSO4, 10 mM KC1) + 2.7 μl/million 2-mercaptoethanol and 50 μg/ml polyvinyl B. The presence of detergents and chelating agents did not improve the assay. A higher pH value of 9.5 improved the activity of β-galactosidase in the absence of NaCO3 added to stop the reaction. Assays were performed at 28 °C and were initiated by addition of 40 μl of o-nitrophenyl-β-D-galactoside (4 mg/ml in T28). The A414 of the reaction was read every 2 min for 1 h, and enzyme activity determined as the slope of the line of best fit of A414 versus minutes (readings with A414 > 2.5 were ignored). Enzyme activity found was to be directly proportional to the A600 of the culture and the volume of culture added to the assay (V in μl). LacZ units were calculated as 200,000 × (A414/min)/A600 × V and were roughly equivalent to standard Miller units.

**Analytical Ultracentrifugation**—Ultracentrifugation experiments were performed in a Beckman XL-A analytical ultracentrifuge using absorbance optics and a four-hole An60Ti rotor. Approximately 100 μl of sample and 105 μl of reference solution were loaded in the sectors of the epoxy centerpieces. Following 24 h of centrifugation, scans were compared at 3-h intervals to ensure that equilibrium had been reached. Data were collected at 280 nm at a spacing of 0.003 cm. The buffer for all experiments was TEG150. Protein was prepared for centrifugation by exhaustive dialysis against TEG150, and the dialyze used as the reference solution for centrifugation. Buffer density (ρ) was measured in an Anton-Paar precision density meter to be 1.03553 g/ml at 5 °C and 1.03644 g/ml at 20 °C. The partial specific volumes (τ) were calculated (using the SEDNTERP program) as 0.727 ml/mg for CIHis6, 0.712 ml/g for CI (83–192)His6, and 0.736 for CI (83–192)His 6.

**RESULTS**

**Homologs of 186 CI**—The first step in investigating the structure-function relationship of 186 CI was to search for homologs. A number of proteins related to the 186 CI repressor were identified by BLAST (16) (Fig. 2). The 186 CI amino acid sequence was used initially to search the protein data base, and four prophage proteins (repressors from phage φ67, Hemophilus influenzae phages HP1 and S2, and Vibrio cholerae phage K139) with homology to CI repressor were identified. The unfinished microbial genomes data base was also searched with 186 CI as the input sequence, and two additional proteins were found. The first was from Klebsiella pneumoniae (WUSTL Genome Sequencing Center) and, judging from other sequence similarities, appears to be the CI homolog of a phage closely related to 186, present as a prophage. The second was a putative phage repressor from Salmonella typhi (CT18 phage) (Sanger Center, Cambridge, UK). Several partial sequences related to 186 CI were also evident in the unfinished genomes of other Salmonella subspecies (typhimurium, paratyphi, enteriditis) but have not been included here. A block alignment of these 186-like proteins was then used to search the BLOCK data base (24). This more powerful search method detected, in addition to those proteins already identified, the CI protein from φ68, a lambdaoid phage, as being related to the 186 like repressors. Alignment of the 186 CI repressor amino acid sequence with those of the seven related repressor proteins reveals two blocks of homology, one of ~70 amino acids at the N terminus, and a second block of ~60 amino acids at the C terminus (Fig. 2). The two blocks are separated by a low homology region of 40–50 amino acids. This non-conserved region may represent an unstructured linker joining two more highly structured domains. In the case of the φ68 repressor, homology was less evident at the C-terminal end. The C termini of the lambdoid repressors form part of the RecA recognition site, with cleavage of the repressor occurring within the central linker (25, 26). 186 CI is not RecA-sensitive (7).

Each CI-like protein in Fig. 2 was examined for the presence of protein motifs using a number of search methods. Potential helix-turn-helix motifs were identified by both the Dodd and Egan (22) weight matrix method and the GYM2 pattern recognition method (23) in some of the proteins. The S.D. scores obtained for each protein by the Dodd and Egan method are shown in parentheses following each sequence in Fig. 2. An S.D. score above 2.5 is considered good evidence for a HTH; likely HTH motifs were identified in some of the proteins, and both search methods always identified the same location as the potential HTH region (solid line in Fig. 2). Thus, although a number of the proteins including CI itself have poor S.D. scores, the alignment of conserved amino acids within the N-terminal block of homology coincides with the position of the predicted HTH motifs in each case. We take this as evidence that 186 CI very likely contains an N-terminal helix-turn-helix motif.

**Domain Structure of CI**—We have used two techniques, linker insertion mutagenesis and limited proteolysis, to investigate the domain structure of the 186 CI repressor protein, with the aim of determining whether one or both of the putative

---

Downloaded from http://www.jbc.org/ by guest on July 23, 2018
amino acids had quite different effects, depending on the posi-
role. The majority of the transposon sequence is then re-
ferred to as having homology to the coliphage 186 CI repressor. The abbreviations used are as follows: Kleb, presumptive phage protein from unfinished sequence of Klebsiella pneumoniae (WUSTL genome sequencing center), HP1, repressor protein from phage HP1 of Hemophilus influenzae (17), S2, repressor protein of phage S2 of Hemophilus influenzae (18), CT18, repressor protein from a phage within Salmonella typhi CT18 (Sanger center), EC67, putative repressor from retrophage Ec67 (19), K139, repressor protein of phage K139 of Vibrio cholerae (20), φ80, CI repressor protein of coliphage φ80 (21). Where at least five of the eight amino acids are identical or conserved, they are shown on a black or gray background, respectively. Each protein sequence was also examined for motifs. The Dodd and Egan weight matrix method (22) and the GYM2 pattern recognition algorithm (23) both detected potential HTH motifs. The S.D. scores obtained for each protein by the Dodd and Egan method are given in parentheses after the sequences. A score of 2.5 or greater indicates a likely HTH motif. Both methods always identified the same sequence within each protein the most likely to contain a HTH motif and these sequences, indicated by the black line, are all aligned in the multiple sequence alignment.

domains have the potential to bind DNA.

In linker insertion mutagenesis (27, 28) a short stretch of amino acids is inserted into the protein of interest. The effect of the inserted amino acids on the activity of the protein is dependent upon the location of the insertion. For example, an insertion located on a surface loop of the protein or in a relatively unstructured region would be expected to have a minor effect on protein function. In contrast, an insertion within a tightly folded or buried region is more likely to interfere with protein function, whether by disrupting protein structure, protein folding, or through an effect on protein stability. To define the regions of CI that are either tolerant or intolerant to insertions, we have used the Genome Priming System-Linker Scanning system (New England Biolabs, Beverly, MA). In this system, a modified Tn7-based transposon is used in an in vitro reaction to make random (1.7 kb) insertions into the gene of interest. The majority of the transposon sequence is then removed by restriction digest and religation, leaving a 15-base pair insertion. In four of the six possible reading frames, this insertion results in a five-amino acid linker, while the other two reading frames generate stop codons. Thus, two sets of CI mutants were generated; (i) a set of truncated proteins and (ii) a set of proteins containing randomly located five-amino acid insertions. The activities of these CI mutants were measured by their ability to repress a single copy, chromosomally inserted pR lacZ reporter, NK7049 (ARS45 pga0000 lacZ). In this system, unrepressed pga0000 lacZ gave 839 (± 77) units, while wild type CI from pRAS1 repressed pga0000 lacZ to 0.7 (± 0.8) units.

Transposon insertions that resulted in truncated protein products occurred at amino acids 36, 61, 74, 103, 106, 107, 110, 112, 123, 173, and 185. These truncated CI proteins invariably lost the ability to repress the pga0000 lacZ reporter. Even a protein truncated at amino acid 185, resulting in just a seven-amino acid C-terminal deletion, was inactive, indicating that these amino acids are required for CI function or stability.

Among the second set of mutants (Fig. 3), insertions of five amino acids had quite different effects, depending on the pos-

\[ \text{FIG. 2. Alignment of proteins with homology to 186 CI.} \]

\[ \text{Alignment by ClustalW of proteins identified by BLAST as having homology to the coliphage 186 CI repressor.} \]

\[ \text{The abbreviations used are as follows: Kleb, presumptive phage} \]

\[ \text{protein from unfinished sequence of Klebsiella pneumoniae (WUSTL genome sequencing center), HP1, repressor protein from phage} \]

\[ \text{HP1 of Hemophilus influenzae (17), S2, repressor protein of phage S2 of Hemophilus influenzae (18), CT18, repressor protein from a phage} \]

\[ \text{within Salmonella typhi CT18 (Sanger center), EC67, putative repressor from retrophage Ec67 (19), K139, repressor protein of phage} \]

\[ \text{K139 of Vibrio cholerae (20), φ80, CI repressor protein of coliphage φ80 (21). Where at least five of the eight amino acids are identical or conserved, they are shown on a black or gray background, respectively. Each protein sequence was also examined for motifs. The Dodd and Egan weight matrix method (22) and the GYM2 pattern recognition algorithm (23) both detected potential HTH motifs. The S.D. scores obtained for each protein by the Dodd and Egan method are given in parentheses after the sequences. A score of 2.5 or greater indicates a likely HTH motif. Both methods always identified the same sequence within each protein the most likely to contain a HTH motif and these sequences, indicated by the black line, are all aligned in the multiple sequence alignment.} \]

\[ \text{FIG. 3. Linker insertion mutagenesis.} \]

\[ \text{A series of repressor mutants were generated in which 15 bp of DNA were randomly inserted into the plasmid (pRAS1)-encoded CI gene. Depending on the reading frame of this 15-bp insertion gave rise to either a truncated protein or a protein with a five-amino acid insertion. The activities of the various CI proteins were assayed in NK7049 (ARS45 pga0000 lacZ). In this system, unrepressed pga0000 lacZ gave 839 (± 77) β-galactosidase units, while wild type CI (supplied from pRAS1) repressed pga0000 lacZ to 0.7 (± 0.8) units. None of the 11 truncated proteins (see "results") were able to repress the single copy pga0000 lacZ reporter gene. The locations of the five amino acid insertions within CI are shown. Insertions shown to the left of the figure reduced or eliminated the ability of CI to repress the pga0000 lacZ reporter gene; the activities of each of these mutants are shown as error bars.} \]

\[ \text{The two shaded areas represent the blocks of homology described in Fig. 2.} \]

\[ \text{[Image 51x552 to 553x729]} \]

\[ \text{[Image 364x186 to 499x447]} \]
mutation of the insertion. β-galactosidase units for the inactive or partially active mutants are given to the right of Fig. 3; active mutants were defined as those that gave less than two units of β-galactosidase activity and are shown to the left of Fig. 3. Insertions within the N-terminal region, with one exception, abolished the ability of CI to repress \( p_R \). The exception was the insertion at amino acid 5, which lies just outside the conserved N-terminal region and which remained fully active. Western blotting of cell extracts prepared from the inactive mutants indicated that the inactivity of two of the mutants (insertions at amino acids 15 and 66) reflected a lack of CI in the soluble fraction (data not shown). Together these data are consistent with the idea that insertions within the conserved HTH-containing N-terminal region disrupt folding and/or protein stability and are thus detrimental to DNA binding. In contrast, mutants having insertions within the putative linker region remained able to fully repress \( p_R \), suggesting that the central region of CI is relatively unstructured or forms part of a surface loop and is thus tolerant to insertions. Only three insertions were obtained within the C-terminal region. The insertion at amino acids 167 had no effect on CI activity, suggesting that this amino acid may also be located on the surface of the protein. Insertions at amino acids 139 and 156 reduced but did not eliminate the ability of CI to repress \( p_R \). One possibility is that insertions in this region disrupt protein-protein association or cooperativity, although we have not tested this explicitly.

Limited proteolysis was used to further probe the domain structure of CI, the principle being that structured regions of the protein will be more resistant to low levels of protease than an unstructured linker region. Purified CI\(_{\text{His6}}\), which we have shown to be equivalent to wild type CI ("Experimental Procedures"), was digested with low levels of protease (papain, proteinase K or subtilisin), aliquots removed over the course of the digestion, and the reactions quenched with PMSF. These samples were analyzed by SDS-PAGE using 10% Tris-Tricine gels (29) (Fig. 4). At early time points, a stable fragment of \(-14\) kDa was observed for each of the proteases employed. A fragment representing the remainder of the protein (expected size \(-8.5\) kDa) was not observed. In the case of the subtilisin and proteinase K digests, the 14-kDa CI fragment was further digested at later time points to a stable 8- to 9-kDa fragment.

The boundaries of these stable fragments were determined by analyzing samples from each of the digests using electrospray mass spectrometry. The results are summarized in the lower part of Fig. 4. All three proteases cleaved within the presumably unstructured C-terminal six-histidine affinity tag of CI. The 14-kDa fragment from the subtilisin digest represents the C-terminal region of CI, cut primarily at amino acid 79, along with some minor products digested within a few amino acids either side of residue 79, consistent with the nonspecific nature of this protease. The smaller subtilisin fragment(s) obtained at later time points result from further digestion at both ends of the larger fragment to give a minimal polypeptide consisting of residues 116–198. This result suggests that the C-terminal region is a stable, folded domain. The absence of a stable N-terminal fragment suggests that the N-terminal region is at least partly unstructured in the absence of DNA and so is susceptible to proteolysis. However, repetition of the proteolysis experiment in the presence of an oligonucleotide containing the \( p_R \) CI binding site gave an identical pattern of cleavage, suggesting that either the N-terminal region remains susceptible to proteolysis when bound to DNA or that upon cleavage elsewhere within the bound protein, the N-terminal fragment dissociates from the DNA and is then susceptible to the protease.

Limited proteolysis with papain gave fragments cleaved primarily around residues 79 and 80, again with some other minor products cleaved at nearby residues. Digestion of CI\(_{\text{His6}}\) with proteinase K also gave cleavage at residues 77–79 at early time points, followed by cleavage primarily at residue 110 at later time points (Fig. 4). This protease-sensitive region (approximately amino acids 77–116) of CI is consistent with the central non-conserved region shown in Fig. 2.

**Properties of CI Domains**—To further examine their biochemical properties in comparison with full-length repressor, residues 1–82 (N-terminal region) and residues 83–204 (linker plus C-terminal region) were cloned, expressed, and purified using a C-terminal six-histidine affinity tag. Both fragments were soluble and obtained in milligram amounts.

Full-length wild type CI repressor associates in solution in an equilibrium between monomers, dimers, tetramers, and octamers (10). The His\(_{\text{6}}\) affinity-tagged CI also associates to octamers in solution (data not shown), with dimers the predominant species at the concentration range in which DNA binding first occurs. The oligomeric state of the N-terminal and C-terminal fragments of CI were assessed by sedimentation equilibrium (Fig. 5). For CI (1–82)His\(_{\text{6}}\), data were obtained at three different loading concentrations (cell 1, 5 \( \mu \)M; cell 2, 16 \( \mu \)M, and cell 3, 32 \( \mu \)M) at a rotor speed of 24,000 rpm (Fig 5a). Initially, the individual scans were analyzed in terms of Equation 1 to obtain whole cell molecular weights. These ranged from 15,870 (± 420) for cell 1 to 18,660 (± 100) for cell 3, values that approach twice that of the monomer molecular weight (10,549). The data for all three cells were then fitted globally to a number of association schemes, with the monomer molecular weight fixed at 10,549. The best fit (as judged by the sum of squares of the residuals) was to a monomer-dimer equilibrium,
with an association constant of $2.5 \times 10^5 \text{M}^{-1}$ ($\Delta G = -6.9$ kcal/mol). When the molecular weight of monomer ($M_1$) was included as an additional fitting parameter, the association constant was unchanged, and a value of 10,770 ($\pm 170$) was obtained for $M_1$. There was no evidence for species beyond dimer. Thus, the N-terminal fragment is able to form stable dimers in solution, albeit with an association constant at least $10^4$-fold weaker than the full-length protein (10). It seems reasonable to now refer to the CI (1–82)His6 fragment as a domain, even though it is not highly resistant to proteolysis.

Data for CI (83–192)His6 were obtained at three loading concentrations (cell 1, 19 μM; cell 2, 50 μM; cell 3, 77 μM) and two rotor speeds (12,000 rpm (Fig. 5b) and 18,000 rpm (Fig. 5c)). Analysis in terms of Equation 1 gave whole cell molecular weights in the range 57,900 ($\pm 1200$) to 85,760 ($\pm 1700$), indicating association to a species at least 6.3-fold larger than the monomer ($M_1 = 13,490$). All six data sets were then fitted globally to a number of models. The best fit (Fig. 5, b and c) was to a dimer to octamer association with an association constant of $1.4 \times 10^{15} \text{M}^{-5} \cdot \text{cm}^3$ ($\Delta G = -19.2$ kcal/mol). It was not possible to obtain data on the monomer-dimer association, which was essentially complete over the concentration range accessible in the sedimentation experiments. Nor was it possible to obtain information about tetramer formation, since the tetramer to octamer transition is a concerted (energetically favored) process and tetramer is not a significantly populated species. However, the free energy of association per dimer for the dimer to octamer transition is $-4.8$ kcal/mol for CI (83–192)His6, compared with $-5.3$ kcal/mol for wild type CI (10). This calculation suggests that the majority of the free energy for CI association is derived from interactions between C-terminal domains.

The functions of the domains were tested in two ways: (i) the ability to repress a reporter gene under control of the early lytic promoters $p_R$ or $p_B$ and (ii) binding in vitro to A type and B type sites. The ability of these protein fragments to repress a single copy chromosomally inserted $p_R$ lacZ and $p_B$ lacZ reporter in vivo was tested (Table I). CI and its variants were expressed from the T7 promoter of pET3a-based plasmids in a strain lacking T7 polymerase. There was sufficient “leakage” of expression in this system to give approximately the same level of CI expression as that found in a 186 lysogen (data not shown).

### Table I

| Plasmid$^{a}$ | $p_R$ lacZ | $p_B$ lacZ |
|--------------|-------------|-------------|
| pET3a        | 139 ± 4.6 (n = 35) | 130 ± 4.8 (n = 36) |
| pETCI        | 4.6 ± 0.6 (n = 22)  | 4.8 ± 0.8 (n = 23)  |
| pETCI(His6)  | 5.3 ± 0.4 (n = 30)  | 5.3 ± 0.7 (n = 29)  |
| pETCI(1–82)His6 | 143 ± 6.4 (n = 22) | 145 ± 7.2 (n = 23) |
| pETCI(83–192)His6 | 137 ± 6.4 (n = 22) | 137 ± 7.2 (n = 23) |
| pETCI(hybrid)His6 | 547 ± 23 (n = 29)  | 547 ± 23 (n = 29)  |
| pETCI(HTH)His6 | 135 ± 4.9 (n = 22) | 135 ± 4.9 (n = 22) |

$^{a}$ Plasmids present in strain NK7049 (ARS45 $p_R$ HinClI/SnaBI lacZ) or NK7049 (ARS45 $p_B$ lacZ).

The strong $p_R$ promoter was repressed ~230-fold, from 534 units in the absence of CI to 2.3 units in the presence of full-length CIHis6. The $p_B$ promoter, which at 139 units is 4-fold weaker than the $p_R$ promoter, is also not as strongly repressed by CI, retaining 4.6 units of activity in the presence of full-length CIHis6, a 30-fold repression. This weaker repression of $p_B$ probably reflects the number and arrangement of the CI operators at the respective promoters (Fig. 1). Neither CI (1–82)His6 or CI (83–192)His6 were able to repress $p_R$ in this system (Table I). Similarly, the CI (1–82)His6 and CI (83–192)His6 domains had no effect on repression of the $p_B$ lacZ reporter. Thus, at least at the concentrations of protein generated in this assay system, repressive capacity of the CI fragments was lost. This is consistent with the inability of the CI truncation mutants (Fig. 3a) to repress $p_R$, even when expressed from the high copy number pRAS1 plasmid. It is possible, however, that the isolated CI domains may be able to bind the CI operators, but be unable to bring about repression of the promoter.

The ability of the N and C-terminal fragments of CI to bind DNA was measured in vitro by SPR. This technique measures binding between macromolecules by detecting changes in refractive index at the surface of a sensor chip, the response being proportional to the mass of macromolecule bound. While a pair of A type CI recognition sites are found at $p_B$, B type sites only occur in combination with an A’ sites at $p_R$. Since we wished to differentiate between the ability of CI to bind to A type and B...
During this titration, nonspecific binding of CI became apparent, but only if the DNA contained the residues necessary for DNA binding to both types of sites (not shown), even when the mutated protein was unable to repress either lacO or lacZ reporters in vivo (Table I) and gave no response in Biacore experiments to either A or B type sites (Fig. 7, b and c). The hybrid repressor bound to both A and B type sites, although with somewhat lower affinity than the wild type 186 repressor. Control Biacore experiments showed that full-length λ rep had no affinity for either A or B type 186 sequences. Taken together, these results indicate that at least some of the binding determinants for both A and B type sequences are located in the 186 CI N-terminal (amino acids 1–82) region. The loss of some binding affinity of the hybrid compared with the wild type repressor is presumably due to less than optimal cooperativity between adjacent bound dimers, since operator to operator spacings differ between λ and 186.

**Mutagenesis of Helix-Turn-Helix**—To test whether the determinants for DNA binding to A and B type sites are both located in the same DNA binding motif of 186 CI, critical residues in the predicted helix-turn-helix motif were mutated. Residues to be mutated were chosen on the basis that they should change the sequence away from the 186-like repressor consensus, but not disrupt the structure of the protein (Fig. 8). Residues 12 and 13 of the ITH motif are commonly involved in sequence-specific interaction with the DNA (34). The serines at these positions in 186 CI (amino acids 37 and 38 of CI) were mutated to arginine and glutamic acid, respectively, amino acids occurring frequently at these positions in other HTH motifs. These changes actually improve the match to the Dodd and Egan (22) HTH master set consensus (S.D. score = 0.5 for wild type, 1.4 for mutant). We expected that these changes would not disrupt the ITH motif but would alter its DNA binding specificity. This protein, CI/HTH His6, was purified in milligram amounts and was shown by sedimentation equilibrium to self-associate to octamers, similar to the wild type protein. This is good evidence that the mutations do not have a large effect on protein folding and are specifically affecting DNA recognition.

The mutated protein was unable to repress either pR or pB lacZ reporters in vivo (Table I) and gave no response in Biacore experiments to either A or B type sites (not shown), even when used at a concentration of 10 μM. In addition, when wild type 186 phage was plated on a strain carrying pETCI(HTH–His6), the resulting plaques were less turbid than those obtained by plating on a control (pET only) strain, suggesting that
FIG. 8. Mutation of the helix-turn-helix motif. The proposed structure the 186 CI HTH motif is shown (33, 34). Shaded residues are those proposed to be involved in sequence specific DNA recognition. The mutations made at serine 12 and serine 13 of the motif (amino acids 37 and 38 of CI) are indicated. The first and last residues of the helices are numbered, as is residue 9, located at the turn and most often a glycine or alanine residue. The bulkier aspartate residue present in CI will most likely result in restricted stereochemistry. A common, but not essential, helix-helix interaction between residues 5 and 15, which helps to orient the helices, is indicated by a dashed line.

CI(HTH\(^-\))His\(_F\) acts as a dominant negative mutant in vivo to wild type phage-derived repressor. This further supports the idea that the mutant is correctly folded, is able to heteroassociate with wild type repressor subunits, and is unable to bind CI operators. We conclude that serines 37 and 38 within the putative HTH motif of 186 CI are necessary for binding to both A and B type sites.

DISCUSSION

Coliphage 186, like the intensively studied bacteriophage \(\lambda\), has evolved to enable it to follow two distinct but interchangeable developmental pathways. As 186 and \(\lambda\) are almost unrelated at the DNA and protein level, the focus of this laboratory has been to study the genetic switch of 186, since it represents an independently evolved solution to a common problem. One aspect of this study has been to investigate the properties of the 186 lysogenic repressor (3, 4, 6, 10). We have shown previously (6) that the 186 CI repressor binds to four sites within the early control region of the phage and that, among these four sites, are two distinct types of inverted repeat operator sequences, termed A type sites and B type sites. The operators are arranged in the order AA-A-BA B-A, where the A' operator has a four rather than five base pair spacing between half-sites. Since 186 CI needs to recognize two different types of sequences, the possibility existed that CI does this using two distinct DNA binding motifs. Such an arrangement is found for example in members of the integrase family of proteins, which recognize core type sites and arm type sites (35, 36). In another example, the recent crystal structure of the bacterial repressor MarA, an araC family member, shows that it contains two HTH motifs, which together bind an asymmetric, degenerate sequence (37). Here we have investigated the structure-function relationship of CI and show that there is one DNA binding motif that recognizes both types of site, rather than two distinct DNA binding motifs. We have also shown that CI consists of two domains, an N-terminal domain (nominally amino acids 1–82), which contains a putative helix-turn-helix motif, forms weak dimers in solution, and is responsible for sequence-specific DNA binding, and a C-terminal domain which, together with the linker region, forms octamers in solution and has no capacity for DNA binding.

In terms of domain structure, 186 CI is similar to the lambdoid repressors, which also consist of an N-terminal DNA binding domain, and a C-terminal domain, which mediates dimerization as well as cooperative interactions between adjacent bound dimers (32, 38). Indeed this arrangement of domains is common among many prokaryotic and eukaryotic transcriptional regulators where protein association is linked to DNA binding (39). There are several lines of evidence that 186 CI also utilizes cooperative interactions. The C-terminal domain alone can associate strongly to octamers. Like \(\lambda\) repressor (40), full-length 186 CI exists in solution in an equilibrium between monomers, dimers, tetramers, and octamers, and both proteins have similar free energies of association (10). CI binding sites are arranged such that they are on the same face of the helix, spaced two or three turns of the helix apart (6). Gel mobility shift experiments show only one retarded species, whether there are one (\(\Lambda\)), two (AA) or three (BA B) sites present on the DNA (6). Mutations (vir mutants) in one or two of the inverted repeats at \(p_R\) diminishes overall binding affinity, yet the same retarded complex is observed (6). We also have preliminary evidence that CI bound at \(p_R\) can interact with CI bound at the flanking sites, similar to the looping observed between \(\lambda\) CI bound at the \(O\_R\) and \(O\_L\) operators (41). Taken together, these points suggest that cooperativity between DNA bound dimers of CI may be important for the existence of a stable lysogenic state. The recent crystal structure of the \(\Lambda\) C-terminal domain has provided a model for cooperative binding between dimers bound to adjacent sites, as well as suggesting a mechanism for tetramer-tetramer interactions (42). The availability of mutants unable to associate were important in confirming the validity of the proposed models. Sequence alignment of the 186 CI-like repressors shown in Fig. 1 with a set of lambda phage repressors was attempted, however, no obvious homologies across families were found, with the exception of the lambdoid \(\phi\)80 phage repressor (Fig. 1), where homology was primarily at the N-terminal region. One approach to isolating C-terminal association mutants (whether monomer-monomer, dimer-dimer, or tetramer-tetramer) of 186 CI would be to select mutants of CI(HTH\(^-\)) that no longer display a dominant negative phenotype.

CI repressor must recognize alternate spacing between the A type half-sites, five base pairs at A sites, four base pairs at the central A' site of \(p_R\). Members of the araC family are able to recognize different half-site spacings by utilizing a flexible linker between the DNA binding domain and dimerization domain (43). Three lines of evidence suggest the presence of a flexible linker joining two domains in 186 CI: (i) in the alignment of 186-like phages (Fig. 2), the two blocks of homology are separated by a region (amino acids 74–135) containing little sequence homology, (ii) insertions of five amino acids within this region (between amino acids 87–124) did not affect the ability of CI to repress \(p_R\) (Fig. 3), and (iii) limited proteolysis of 186 CI (Fig. 4) resulted in cleavage around amino acid 80 and, at later time points, around amino acid 116, while retaining a stable C-terminal fragment. Thus the presence of a linker between 40 and 60 amino acids in length may allow CI to recognize the variable (4 or 5 bp) half-site spacing found in A type sites and also may be important in higher order association.

It is apparent that A and B type sequences are not recognized equally well by 186 CI. Both full-length CI and the N-terminal domain bind to a pair of A sites at least 10-fold more strongly than to a pair of B sites separated by the same distance. Strong repression of the \(p_R\) promoter by CI is essential for maintaining the lysogenic state and indeed, in a lysogen, \(p_R\) is repressed at least 300-fold (4). Strong cooperativity between adjacent bound dimers of CI is likely responsible for the overall tight binding at \(p_R\), as the individual sites have relatively poor affinity for repressor (6). This strong cooperativity is also manifested, at least in vitro, in the observation of spreading of CI binding along the DNA from specifically bound sites (Ref. 6;
Domain Structure and Function of Coliphage 186 Repressor

Acknowledgments—We thank Peter Brautigan for preliminary protolysis experiments, Lyle Carrington for assistance with analytical ultracentrifugation, and members of the Egan laboratory for valuable discussions and reagents. We acknowledge the Genome Sequencing Center, Washington University, St. Louis, MO for communication of DNA sequence data prior to publication and the Sanger Center at Cambridge and Del Pickard at Imperial College for use of CT18 sequencing data.

REFERENCES

1. Ptashne, M. (1986) A Genetic Switch, Cell Press and Blackwell Scientific, Cambridge

2. Friedman, D. I., and Court, D. L. (2001) Curr. Opin. Microbiol. 4, 201–207

3. Kalionis, B., Dodd, I. B., and Egan, J. B. (1986) J. Mol. Biol. 191, 199–209

4. Dodd, I. B., Kalionis, B., and Egan, J. B. (1990) J. Mol. Biol. 214, 27–37

5. Dodd, I. B., Reed, M. R., and Egan, J. B. (1993) Mol. Microbiol. 10, 1139–1150

6. Dodd, I. B and Egan, J. B. (1996) J. Biol. Chem. 271, 1152–11540

7. Lamont, I., Brumby, A. M., and Egan, J. B. (1989) Proc. Natl. Acad. Sci. 86, 5492–5496

8. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) Methods Enzymol. 185, 60–89

9. Simons, R. W., Homan, P., and Kleekne, N. (1987) Gene 53, 85–96

10. Shearwin, R. E., and Egan, J. B. (1996) J. Biol. Chem. 271, 11525–11531

11. Neufing, P. J., Shearwin, K. E., Camerotto, J., and Egan, J. B. (1996) Mol. Microbiol. 21, 751–761

12. Powell, B. S., Court, D. L., Nakamura, Y., Rivas, M. P., and Turnhough, C. L., Jr. (1994) Nucleic Acids Res. 22, 5765–5766

13. Miller, J. H. (1972) Experiments in Molecular Genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

14. Bipatnath, M., Dennis, P., and Bremer, H. (1998) J. Bacteriol. 180, 265–273

15. Schupp, J. M., Travis, S. E., Price, L. B., Shand, R. F., and Keim, P. (1995) BioTechniques 19, 18–20

16. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Nucleic Acids Res. 25, 3389–3402

17. Esposito, D., Fitzmaurice, P. W., Benjamin, R. C., Goodman, S. D., Waldman, A. S., and Scooca, J. J. (1996) Nucleic Acids Res. 24, 2360–2368

18. Skowronek, K., and Baranowski, S. (1997) Gene 196, 139–144

19. Hsu, M. Y., Inouye, M., and Inouye, S. (1990) J. Biol. Chem. 265, 18459–18464

20. Ogawa, T., Ogawa, H., and Tomizawa, J. (1988) J. Mol. Biol. 202, 537–550

21. Dodd, I. B., and Egan, J. B. (1990) Nucleic Acids Res. 18, 5019–5026

22. Gao, Y., Mathee, K., Narasimhan, G., and Wang, X. (1999) Proc. 6th SPIRE Conference September 21–24, 1999, pp. 63–72. Institute of Electrical and Electronics Engineers Cancun, Mexico

23. Henikoff, J. G., Henikoff, S., and Pietrokovski, S. (1999) Nucleic Acids Res. 27, 226–228

24. Roberts, J. W., Roberts, C. W., and Mount, D. W. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 2283–2287

25. Sauer, R. T., Ross, J. M., and Ptashne, M. (1982) J. Biol. Chem. 257, 4458–4462

26. Goff, S. P., and Prasad, V. R. (1991) Methods Enzymol. 208, 586–603

27. Biery, M. C., Stewart, F. J., Stellwagen, A. E., Raleigh, E. A., and Craig, N. L. (1994) J. Mol. Biol. 244, 199–228

28. Biery, M. C., Stewart, F. J., Stellwagen, A. E., Raleigh, E. A., and Craig, N. L. (1994) J. Mol. Biol. 244, 199–228

29. Gao, Y., Mathee, K., Narasimhan, G., and Wang, X. (1999) Proc. 6th SPIRE Conference September 21–24, 1999, pp. 63–72. Institute of Electrical and Electronics Engineers Cancun, Mexico

30. Carlson, N. G., and Little, J. W. (1993) Biochemistry 32, 1108–1130

31. Kalionis, B., Dodd, I. B., and Egan, J. B. (1989) Proc. Natl. Acad. Sci. U.S.A. 86, 5492–5496

32. Shearwin, R. E., and Egan, J. B. (1996) J. Biol. Chem. 271, 11525–11531

33. Neufing, P. J., Shearwin, K. E., Camerotto, J., and Egan, J. B. (1996) Mol. Microbiol. 21, 751–761

34. Powell, B. S., Court, D. L., Nakamura, Y., Rivas, M. P., and Turnhough, C. L., Jr. (1994) Nucleic Acids Res. 22, 5765–5766

35. Schupp, J. M., Travis, S. E., Price, L. B., Shand, R. F., and Keim, P. (1995) BioTechniques 19, 18–20

36. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Nucleic Acids Res. 25, 3389–3402

37. Esposito, D., Fitzmaurice, P. W., Benjamin, R. C., Goodman, S. D., Waldman, A. S., and Scooca, J. J. (1996) Nucleic Acids Res. 24, 2360–2368

38. Skowronek, K., and Baranowski, S. (1997) Gene 196, 139–144

39. Hsu, M. Y., Inouye, M., and Inouye, S. (1990) J. Biol. Chem. 265, 18459–18464

40. Senear, D. F., Laue, T. M., Ross, J. B. A., Waxman, E., Eaton, S., and Rusinova, E. (1993) Biochemistry 32, 6179–6189

41. Revert, E., van Wilchen-Bergmann, E., Bessert, H., Barker, A., Muller-Hill, B. (1999) Curr. Biology 9, 151–154

42. Bell, C. E., Frescura, P., Hochschild, A., and Lewis, M. (2000) Cell 101, 801–811

43. Eustance, R. J., Bustos, S. A., and Schlief, R. F. (1994) J. Mol. Biol. 242, 330–338

44. Brumby, A. M., Lamont, I., Dodd, I. B., and Egan, J. B. (1996) Virology 219, 105–114

45. Shearwin, R. E., Brumby, A. M., and Egan, J. B. (1998) J. Biol. Chem. 273, 5708–5715
