Analysis of DNA Binding by a Eubacterial Zinc Finger Transcription Factor

Victor J. McAlister† and Gail E. Christie*
Department of Microbiology and Immunology, Virginia Commonwealth University School of Medicine, Richmond, Virginia 23298-0678

Received 12 February 2009/Accepted 5 May 2009

The Serratia marcescens NucC protein is structurally and functionally homologous to the P2 Ogr family of eubacterial zinc finger transcription factors required for late gene expression in P2- and P4-related bacteriophages. These activators exhibit site-specific binding to a conserved DNA sequence, TGT-N$_3$-R-N$_2$-Y-N$_3$-aCA, that is located upstream of NucC-dependent S. marcescens promoters and the late promoters of P2-related phages. In this report we describe the interactions of NucC with the P2 FETUD late operon promoter P$_F$. NucC is shown to bind P$_F$ as a tetramer and to make 12 symmetrical contacts to the DNA phosphodiester backbone. The backbone contacts are centered on the TGT-N$_3$-R-N$_2$-Y-N$_3$-aCA motif. Major groove base contacts can be seen at most positions within the ~24-bp binding site. Minor groove contacts map to adjacent positions in the downstream half of the binding site, which corresponds to the area in which the DNA also appears to be bent by NucC binding. NucC binding provides a new example of protein-DNA interaction that is strikingly different from the DNA binding demonstrated for eukaryotic zinc-finger transcription factors.

The Serratia marcescens NucC protein was originally identified as a positive regulator of extracellular nuclease and bacteriocin 28b production (15, 23). NucC is a basic protein of 75 amino acids that is a member of the bacteriophage P2 Ogr family, a novel class of eubacterial zinc-binding transcription factors required for late gene expression in P2- and P4-related phages. All members of this family are characterized thus far at least somewhat functionally interchangeable and have significant amino acid sequence similarity in the N-terminal two-thirds of the protein. This similarity includes four essential Cys residues that are involved in coordination of zinc (17, 27, 30, 36), arranged in a Cys-X$_2$-Cys-X$_2$-Cys-X$_2$-Cys motif. The function(s) of the nonconserved C-terminal residues is unclear. Viable C-terminal deletions of P2 ogr and phage 186 B suggest that this region is dispensable, although it likely contributes to full activity (18, 36). Genetic evidence is consistent with a mechanism for transcription activation that involves a direct interaction with the C-terminal domain of the α subunit(s) of RNA polymerase (1, 45, 50). Purified NucC, RNA polymerase holoenzyme, and template DNA have been shown to be the only necessary components for open complex formation and transcription from the P2 late promoter P$_F$, and NucC induces a 90° bend in DNA containing the P$_F$ binding site (35).

Zinc fingers constitute a well-characterized motif for protein-DNA interaction, and structural data are available for a number of eukaryotic zinc finger transcription factors (reviewed in references 28, 29, and 32). Zinc finger proteins containing C$_{4-9}$ zinc coordination motifs represent the largest group of transcription factors in eukaryotes (38, 47). These proteins typically contain several zinc fingers that wrap around the DNA in a spiral manner. The zinc coordinating motif in each finger contains an α-helix preceded by a β-hairpin. Amino acid residues on the surface of the α-helix make contacts to adjacent bases in the major groove, and in most cases the β-hairpin makes phosphate backbone contacts. Two other types of zinc coordination motifs are exemplified by the yeast Gal4 protein and the hormone receptor family of transcriptional activators. Gal4 binds as a homodimer to the sequence CCG-N$_1$-CGG (4, 7, 19). The zinc coordinating motif is composed of a pair of α-helices that coordinate two zinc ions through six cysteine residues. The first helix contacts DNA bases in the major groove. The second helix contacts the phosphodiester backbone. Each zinc binding domain recognizes a CCG triplet (34). The hormone receptor family zinc-coordinating motif is characterized by two antiparallel α-helices capped by loops at their amino-terminal ends. Each helix-loop pair coordinates a single zinc ion. The first helix interacts with DNA bases in the major groove, while the loops and second helix make contacts to the phosphate backbone (41). The DNA-binding sites for the hormone receptors consist of two 6-bp consensus half sites (reviewed by Freedman and Luisi [16]). The orientation of the two half-sites and the spacing between them (3 to 6 bp) determines which subunits dimerize and the orientation between them (parallel or symmetrical).

In bacteria, transcription factors with zinc-binding motifs are uncommon. The Agrobacterium tumefaciens Ros protein contains a single Cys$_3$His$_2$ zinc finger domain and regulates expression of virulence genes and the plant oncogene ipt (8). Ros homologs have been identified primarily in other Rhizobiaceae. Although a plant homologue of Ros has not been found, phylogenetic analysis indicates that these proteins are probably of eukaryotic origin (3). The only other known example of a eubacterial zinc finger transcription factor is provided by the P2 Ogr family of proteins. Homologues can be identified by...
sequence similarity and by the presence of the conserved Cys-
X2-Cys-X2-Cys-X2-Cys zinc coordination motif. Most are
found in members of the Enterobacteriaceae (presumably in
resident P2-related prophages) and in P2- and P4-related
phages. Homologues have not been identified in other king-
doms. However, an internal region of the staphylococcal
cassette recombinase CcrC (22) appears to contain a similar
domains. However, an internal region of the staphylococcal

FIG. 1. DNA templates used for footprinting analyses and the primers used for PCR and 5’-end labeling. The amplified region of pM37 is
shown; the sense strand is the top strand of the double-stranded DNA sequence, and the template strand is the bottom strand. The TGT and CCA
hal-site (centered at positions −65 and −50, respectively) are indicated with inverted arrows, and the transcription start site (+1) and direction
is also indicated by an arrow. Template 1 was made by PCR using pM37 (55) and the primers 1204 M + n and M6. Template 2 was made by PCR
using pM37 and the primers 1204 M + n and M6-3. Template 3 was made by PCR using pLYU2 and primers 1204 M + n and M6-3. PCR
amplification adds two adenine residues (in primer M6-3) downstream of the P2 sequence to templates 2 and 3. Plasmid pLYU2 is a derivative
of pM37 with a T instead of a G at position −43 and a position C instead of an position A at −40 of the P2 sequence. Upstream of P2 nucleotide
−69 is 37 bp of vector derived DNA, which is indicated by italics.

MATERIALS AND METHODS

Proteins and DNA. NucC was purified as described previously (35). Three
DNA templates were used for footprinting and are shown in Fig. 1. The
templates were made by PCR amplification, using 5’-end-labeled primers
Taq polymerase. Primer end labeling was done with T4 kinase and

DNA footprinting and electrophoresis. The DNA binding conditions for pro-
tection and interference footprinting consisted of 1 pmol of template DNA (300,000 to 500,000 cpm), 50 mM Tris-HCl, 50 mM KCl, 10 mM MgCl2, 1 mM
dithiothreitol, 100 μg of bovine serum albumin/mL, and 10% glycerol in a 10-
to 20-μL volume, except where indicated otherwise. The amount of NucC used in
the binding reactions (typically 50 to 200 ng) was adjusted for each experiment.
Interference and protection products were isolated on polyacrylamide Tri-bo-
rate-EDTA gels. Free DNA and NucC-bound DNA were recovered from poly-
acrylamide gel slices as indicated for each experiment. The recovered DNA
was further modified as described below. Footprinting reactions were resolved on
6.8% to 8% polyacrylamide–Tris-borate-EDTA–7.2 M urea gels.

Methylation protection and interference. Methylation protection and interfer-
ence assays were performed as described by Shaw and Stewart (42). For both
assays, 20 μL of 0.01% dimethyl sulfoxide was added to 20 μL of template DNA
for 2 min. In the protection assay, the modification was done in the presence of
NucC. The reactions were stopped by the addition of 250 μM dithio-
threitol, DNA modified in the absence of NucC was ethanol precipitated with 0.3
M sodium acetate twice before use in the interference assay. Reaction products
were isolated on polyacrylamide gels, and the DNA was recovered from gel slices
to centrifugation at 4°C for 10 min. Depurination interference. A total of 20 μL of 4% (vol/vol) formic acid was added to 2 μg of template DNA to 200 μL of H2O and transferred to a 37°C
water bath for 20 min. The reaction was placed on ice and precipitated by adding 20 μL of 3 M sodium acetate (pH 7) and 600 μL of 95% ethanol. The DNA was
pellet was centrifuged at 10 min at room temperature. The pellets were
washed with 75% ethanol, air dried, resuspended in H2O, and used in the gel
shift assay. Reaction products were recovered from gel slices by electroelution
into saturated ammonium acetate and ethanol precipitated overnight at –80°C.

Piperidine strand scission was performed as described for the methylation assay.
Uracil substitution and missing thymine interference. Uracil was randomly
substituted for thymine by supplementing the PCR with 5 μM dUTP (200 μM
concentrations of other deoxynucleoside triphosphates and 1.5 mM MgCl2). For
the uracil substitution assay the DNA was recovered from the polyacrylamide gel
used in the gel shift assay by passive diffusion, treated with uracil DNA glyco-
sylase, phenol-chloroform extracted, and ethanol precipitated. Piperidine strand
scission was performed as described for the methylation assay. For the missing
thymine assay, the uracil substituted template was treated with uracil DNA glyco-
sylase, phenol-chloroform extracted, and ethanol precipitated before the
gel shift assay. The DNA was recovered from gel shift slices by electroelution and
precipitated overnight at –80°C.

Ethylation interference. Ethylation interference was performed as described
by Manfield and Stockley (33). The DNA was resuspended in 100 μL of sodium
cacodylate. After a brief incubation at 30°C, 100 μL of ethynitrosourea-saturated
ethanol was added, and incubation at 50°C was continued for 30 min. The
ethynitrosourea DNA was precipitated by adding 4 μL of 4 M NaCl, 30 μL of Escherichia
coli tRNA, and 150 μL of ethanol. The reactions were incubated on ice for 10
min, followed by centrifugation for 15 min at room temperature to recover the
modified DNA, which was washed once with 70% ethanol, air dried, and dis-
solved in H2O for use in a gel shift assay. Bound and unbound DNA was
recovered from gel slices by electroelution into saturated ammonium acetate and
ethanol precipitated overnight at –80°C. The precipitated DNA was resus-
pended in 15 μL of 10 mM sodium phosphate (pH 6.8). For strand scission, 2.5
μL of 1 M NaOH was added, and the DNA was incubated at 90°C for 30 min.
The reactions were placed on ice and then precipitated by adding 2.5 μL of 1 M acetic
acid, 2 μL of 4 M NaCl, and 70 μL of ethanol. After a 10-min incubation on ice and
a 15-min room temperature centrifugation, the pellet was washed with 70% ethanol and air dried.

Hydroxyl radical protection. Hydroxyl radical cleavage reactions were per-
formed as described by Schickor and Heumann (40). Aliquots (3.5 μL of 20 mM
sodium ascorbate, 0.6% hydrogen peroxide, and 2 mM iron(II) sulfate hexa-
hydrate–4 mM EDTA were added to a 10-μL NucC binding reaction in the absence of
glycerol. The cleavage reaction was stopped after 30 s by adding 1 volume
of 20% glycerol, and the bound and free products were isolated by polyacrylamide
gel electrophoresis (PAGE). The DNA was recovered from gel slices by passive
diffusion at 4°C in H2O, ethanol precipitated, washed with 80% ethanol, and air dried.

Molecular mass estimation of protein-DNA complexes by native gel elec-
rophoresis. The molecular mass of the NucC-DNA complex was measured by using
the method of Ferguson (14). Native gel electrophoresis was performed in a
buffer containing 50 mM Tris base, 40 mM boric acid, and 10 mM EDTA.
Oligonucleotide 5′-GTT GTG CTC TGG ATT AGA CAA C-3′ and its 22-bp
complement were annealed to generate the NucC binding site. The migration of

Downloaded from http://jb.asm.org/ on August 15, 2017 by guest
protein standards (Sigma), the DNA and the NucC-DNA complexes were analyzed on a series of gels containing different concentrations of polyacrylamide. The relative mobility ($R_f$) was calculated as the ratio of the mobility of a given protein to the mobility of bromophenol blue dye. For each protein, a derived value, $100\log (R_f)/100$, was plotted as a function of the acrylamide concentration (9, 21). The slopes of the lines generated for each standard protein were determined by linear regression analysis and plotted as a function of the logarithm molecular mass. This calibration curve, a Ferguson plot, was applied to determine the molecular mass of the NucC-DNA complex. The following polyacrylamide concentrations were used: 4, 5.5, 6, 7, 8, 9 and 10% for bovine serum albumin; 6, 7, 8, 9, 10, 11.5, and 12% for chicken egg albumin; 6, 7.5, 8.5, 9, 10, 11.5, and 13% for bovine erythrocyte carbonic anhydrase; 7, 11, 11.5, 12, and 13% for bovine milk $\alpha$-lactalbumin; and 6, 7.5, 8.5, 9, 10, 11.5, and 12% for the NucC-DNA complex. Migration of the unbound DNA was measured in the same gel as the NucC-DNA complexes. Protein standards were visualized by Coomassie blue staining. The NucC-DNA complexes were visualized by ethidium bromide staining.

RESULTS AND DISCUSSION

Base contacts. (i) Methylation protection and interference. The alkylating reagent dimethyl sulfate (DMS) primarily methylates the N7 position of guanine in the major groove of DNA and the N3 position of adenine in the minor groove (43) and can be used to identify contacts in protection and pre-modification interference assays (42). Purine methylation results in the loss of a hydrogen bond acceptor and addition of a positive charge on the purine ring. These changes and steric blockage can affect protein binding. In addition to blocking the accessibility of the modified guanine and adenine atoms to protein, the DNA and the NucC-DNA complexes were analyzed on a series of gels containing different concentrations of polyacrylamide. The relative mobility ($R_f$) was calculated as the ratio of the mobility of a given protein to the mobility of bromophenol blue dye. For each protein, a derived value, $100\log (R_f)/100$, was plotted as a function of the acrylamide concentration (9, 21). The slopes of the lines generated for each standard protein were determined by linear regression analysis and plotted as a function of the logarithm molecular mass. This calibration curve, a Ferguson plot, was applied to determine the molecular mass of the NucC-DNA complex. The following polyacrylamide concentrations were used: 4, 5.5, 6, 7, 8, 9 and 10% for bovine serum albumin; 6, 7, 8, 9, 10, 11.5, and 12% for chicken egg albumin; 6, 7.5, 8.5, 9, 10, 11.5, and 13% for bovine erythrocyte carbonic anhydrase; 7, 11, 11.5, 12, and 13% for bovine milk $\alpha$-lactalbumin; and 6, 7.5, 8.5, 9, 10, 11.5, and 12% for the NucC-DNA complex. Migration of the unbound DNA was measured in the same gel as the NucC-DNA complexes. Protein standards were visualized by Coomassie blue staining. The NucC-DNA complexes were visualized by ethidium bromide staining.

SDS-PAGE analysis. NucC was compared to protein molecular mass standards by SDS-PAGE with a Tricine buffer system and an 18% polyacrylamide gel. The two samples of NucC that were compared were stored at $-80^\circ\text{C}$ and at

FIG. 2. Major groove contacts to guanine N7 atoms and minor groove contacts to adenine N3 atoms probed by methylation protection and interference. (A) Methylation protection. Template 1 DNA was end labeled on either the sense strand or the template strand. DNA bound to NucC was isolated by gel shift to avoid footprinting of nonspecific complexes. Unbound or NucC-bound DNA, as indicated, was cleaved by using a reaction specific for methylated guanines prior to analysis on a polyacrylamide-urea sequencing gel (described in Materials and Methods). Asterisks designate two adenine residues that were also cleaved by DMS in this assay. Guanines protected from methylation by NucC binding are indicated by closed diamonds; open diamonds indicate guanines that are hypermethylated when NucC is bound. A G+A sequencing ladder of the same DNA was used as a marker. (B) Methylation interference. Template 1 DNA was end labeled on either the sense strand or the template strand and methylated prior to NucC binding. Bound and unbound fragments were isolated by gel shift and cleaved by using a reaction specific for methylated guanines and adenines (described in Materials and Methods) prior to analysis on a polyacrylamide-urea sequencing gel. Closed circles indicate nucleotides that decreased NucC binding when methylated; open circles indicate residues where methylation enhanced NucC binding. (C) Summary of the contacts mapped in the methylation protection and interference assays; symbols as described above. The conserved bases in the TGt-N12-cCA motif are boxed.
DMS, protein binding can alter the methylation pattern of DNA by changing the conformation of the DNA and by creating areas of altered hydrophobicity. For the protection assay it was important to isolate the bound DNA before strand scission to avoid footprinting nonspecific NucC-DNA complexes.

NucC binding affected the methylation of eleven guanines and two adenines within the binding site (Fig. 2A). NucC binding decreased methylation at positions −68, −65, −60, and −57 on the sense strand and positions −62, −58, −50, and −47 on the template strand. NucC increased the methylation of guanines at positions −69 and −52 on the sense strand and position −51 on the template strand. Although methylated adenines are not usually resolved by piperidine strand scission, two bands corresponding to the adenines at positions −55 and −54 are apparent in the footprint. The adenine at position −54 was slightly hypermethylated in the presence of NucC, while the adenine at position −55 was slightly protected. NucC binding also affected the methylation of several bases that are outside of the area predicted to be contacted by the protein. These modifications are consistent with a NucC-induced structural alteration that changes the conformation of the major groove in these areas.

In the interference assay, a strand scission reaction that equally resolves guanine and adenine methylation was used. Many of the modifications detected in the interference assay were also detected in the protection assay. Methylation of ten guanines and two adenines affected NucC binding (Fig. 2B). Methylation of guanines at positions −68, −65, −60, and −52 on the sense strand and at positions −62, −50, and −47 on the template strand decreased the binding, and methylation of the guanines at positions −69 and −63 on the sense strand and at position −46 on the template strand increased binding. Methylation of adenines at position −53 on the sense strand and at position −55 on the template strand also increased NucC binding.

(ii) Missing purine interference. Methylation protection and premodification interference assays provide very specific and limited information regarding the chemical features involved in DNA recognition. For instance, contacts to guanine 6-carbonyl and adenine N7 and 6-methyl groups are not expected to be detected in this assay. We addressed additional contributions of purine samples by using a less specific assay, depurination interference (5).

Depurination at most positions within the predicted binding site decreased NucC binding (Fig. 3). The two positions where binding was not affected were −69 and −54. Purine removal at position −55 increased binding. The results at positions −54 and −55 stand in contrast to the results obtained in the methylation interference assay, where methylation of the adenine at position −54 decreased binding while methylation of adenine at position −55 increased binding. The apparent discrepancy between the results of the depurination and methylation assays at positions −55 and −54 likely reflect a local distortion of the DNA helix that accompanies NucC-induced bending.

(iii) Uracil substitution and missing thymine interference. Substitution of uracil for thymine in DNA results in the loss of the thymine 5-methyl group. The uracil substitution assay (37) was used to further probe areas of the major groove contacted by NucC. Uracil substitution at positions −67, −66, −64, and −59 on the sense strand and at positions −56, −53, −49, and −48 on the template strand decreased NucC binding (Fig. 4A). Uracil substitution in the template strand also interferes with NucC binding at what appears to be positions −96 and/or −99, which are well upstream of the region predicted to be contacted by NucC. We do not have an explanation for this result. This area of the probe does not correspond to Ps DNA and was derived from the plasmid template used to generate the probe.

The only positions within the NucC binding site that were not affected by uracil substitution are −61, −55, and −54. Uracil substitution does not provide information regarding contacts to the uracil 4-carbonyl group in the major groove or to the uracil 2- and 4-carbonyl groups in the minor and major grooves, respectively. To address the potential for these and other thymine contacts and to examine the effects of DNA...
flexibility on NucC binding, we used the missing thymine interference assay (11). The results (Fig. 4B) are identical to the results of the uracil substitution assay at every position except two; removal at positions −55 and −54 on the sense strand increased NucC binding. These are two of the three positions within the NucC binding site that were not detected as contacts in the uracil substitution assay.

**Phosphodiester backbone contacts: ethylation interference.** Modification of DNA with the alkylating reagent ethylnitrosourea results in the addition of ethyl groups to the nonesterified oxygens in the phosphodiester backbone (43), which can be detected as contacts in an interference assay (see reference 33 and references therein). Interference with protein binding occurs when the addition of an ethyl group causes steric blockage or disrupts electrostatic interaction with the phosphodiester backbone.

NucC ethylation interference occurred at positions corresponding to phosphodiester groups between positions −69 and −68, −66 and −65, −61 to −59, and −53 to −51 on the sense strand and between positions −64 and −61, −56 and −54, −50 and −49, and −47 and −46 on the template strand (Fig. 5). Phosphodiester contacts at these positions were confirmed in multiple experiments. Other positions where ethylation appears to affect binding in the gel shown could not be reproduced on a consistent basis. These may result from base mod-

![FIG. 4. Uracil substitution and thymine removal interference. (A) Uracil substitution interference. Uracil was randomly substituted for thymine in template 2 by PCR amplification with Taq polymerase and dUTP. The interference products (bound and unbound) for DNA end labeled on each strand were isolated by gel shift prior to strand scission at the positions of substituted uracil and analysis on the polyacrylamide-urea sequencing gel. Individual thymine residues that interfered with NucC binding when replaced with uracil are indicated. (B) Thymine removal interference. Uracil was randomly substituted for thymine in template 2 by PCR. The substituted uracil bases were removed by using uracil deglycosylase prior to isolating the interference products by gel shift, strand cleavage, and analysis on the polyacrylamide-urea sequencing gel. Individual thymine residues that affected NucC binding when absent are indicated. (C) Summary of the contacts mapped in the uracil substitution and thymine removal assays. Closed circles indicate residues where both uracil substitution and thymine removal decreased NucC binding. Open circles indicate residues where thymine removal increased binding. The conserved bases in the TGt-N12-cCA binding site motif are boxed.](http://jb.asm.org/)

![FIG. 5. NucC phosphate backbone contacts. Template 2 was ethylated with ethylnitrosourea, and interference products were isolated by gel shift prior to strand cleavage at the ethylated positions and analysis of the bound and unbound fractions, as indicated, on a polyacrylamide-urea sequencing gel. Fragments were radiolabeled at one 5′ end, on the indicated strand. A G+ A sequencing ladder of the same end-labeled DNA was used as a marker. The sequence of each strand of the binding site is shown alongside the gel. Positions where ethylation of the phosphate backbone interferes with NucC binding are indicated by an “i.” The conserved bases in the TGt-N12-cCA motif are boxed.](http://jb.asm.org/)
fications that also affect binding but are poorly resolved under the experimental conditions.

Minor groove interactions: hydroxyl radical protection. Hydroxyl radical footprinting is a way to measure the solvent accessibility of sugar protons in DNA at high resolution (46). In canonical B-DNA, H-4' and H-5' are the sugar protons most accessible to solvent and are located on the minor groove face of the DNA. Changes in the reactivity of these protons to hydroxyl radicals can be attributed to protein contacts that block access or to protein-induced structural alterations to the DNA (6, 12).

There are variations in the pattern of DNA reactivity to hydroxyl radicals in the presence of NucC (Fig. 6). Areas of decreased reactivity were centered at approximately positions −66 and −54 on the sense strand and at positions −64 and −47 on the template strand. These results are consistent with NucC binding to this region, but it is unclear whether these changes are a direct result of close interactions of NucC within the minor groove. Changes in the pattern of reactivity are seen outside of the NucC binding site as well, and these changes as well as those within the binding site are consistent with NucC-induced structural alterations of the DNA.

Binding stoichiometry. (i) Molecular mass analyses of NucC-DNA complexes by native gel electrophoresis (Ferguson plot) analysis. Protein standards and NucC bound to a 22-bp DNA binding site were run on a series of gels with different polyacrylamide concentrations. The relative mobility ($R_f$) of each species was calculated as the ratio of mobility to the mobility of bromophenol blue and a derived value, $100\log(R_f \times 100)$, was plotted as a function of acrylamide concentration (data not shown). The slopes of the lines generated were determined by linear regression analysis and plotted as a function of the molecular mass to generate the calibration curve (Ferguson plot). Data points: 1, unbound DNA (13.47 kDa); 2, $\alpha$-lactobumin (14.2 kDa); 3, carbonic anhydrase (29 kDa); 4, egg albumin (45 kDa); 5, NucC-DNA complex; 6, bovine albumin monomer (66 kDa); 7, bovine albumin dimer (132 kDa). The $R_f$ of the NucC-DNA complex was intermediate to that of the egg albumin charge isomers over the range of polyacrylamide gel concentrations measured.

FIG. 8. SDS-PAGE analysis of purified NucC. Samples, stored as indicated, were prepared for electrophoresis as described in Materials and Methods and run on an 18% polyacrylamide gel. NucC stored at 4°C progressively lost DNA-binding activity (data not shown).
DNA complex was 42.1 kDa. NucC has a predicted molecular mass of 8.38 kDa, and the molecular mass of the DNA fragment is 13.47 kDa. Using these values, 3.42 molecules of NucC are calculated to bind per molecule of DNA.

(ii) SDS-PAGE analysis of NucC. As has been reported previously for the related PSP3 Pag and dR73 Delta proteins (24), NucC monomers migrate on SDS-polyacrylamide gels with an apparent molecular mass of 6.2-kDa (Fig. 8). The reason for this aberrant migration is not understood. Surprisingly, freshly purified NucC migrates as an apparent dimer on an SDS-polyacrylamide gel (Fig. 8). When the protein is stored at 4°C for several days, it assumes the electrophoretic mobility of monomeric NucC but loses its DNA-binding activity. The related Delta protein of satellite phage P4 contains two head-to-tail NucC-like domains in tandem, both of which are required for activity (26). These results, in conjunction with the dyad symmetry present in the binding site, strongly imply that the binding stoichiometry is constrained to an even number. Based on the molecular mass determined for the protein-DNA complexes, we conclude that four molecules of NucC are binding to the DNA.

There are several reported examples of proteins that form noncovalent oligomers or complexes resistant to SDS denaturation, even after boiling, including streptavidin (2), ubiquitin-like proteins (31), prionlike proteins (44), and membrane proteins (39). Such tight noncovalent associations are generally thought to be due to strong hydrophobic interactions between protein molecules. Further studies of NucC structure and function will be required to determine what part of the protein is involved in this unusual dimerization.

(iii) Compilation and structural modeling of DNA contacts. Specific features of the DNA recognized by NucC are modeled on the B-form DNA and summarized in Fig. 9. The DNA is

![Diagram](image-url)
displayed as four 90° rotations of the molecule. For clarity, nitrogen modifications detected in the methylation assays are shown in blue on the model.

The 12 phosphodiester contacts form a symmetrical distribution. Major groove contacts are mapped at 18 positions within the binding site. Minor groove contacts are mapped to adenine N3 atoms at positions −55, −54, and −53. Removal of thymine bases at positions −55 and −54 increases binding, indicating that this area of the DNA is bent. A NucC-induced Pp DNA bend of 90° has been measured by gel shift analysis (35). A distortion of the DNA helix could result in extrahelical base conformations or non-Watson-Crick base pairing at positions −55 and −54, with adenine N3 atoms facing the major groove. Another feature in the downstream half of the binding site is the degree of hypermethylation; methylation increases NucC binding or NucC binding increases methylation at positions −55 to −51. Outside of this cluster, hypermethylation is only detected at three positions: −69, −63, and −46. The interface between the protein and DNA appears to be more hydrophobic in the downstream half of the binding site.

The conserved trinucleotide inverted repeat in the NucC binding site is separated by 12 nonconserved bases. A contributing role for these nonconserved bases in the Pp binding site has been demonstrated in previous genetic studies (10). In addition to showing that the conserved trinucleotide repeats in each half of the binding site are absolutely essential for binding and activity, this genetic analysis also indicated smaller effects for nucleotide substitutions in the intervening sequence. Furthermore, complementary base substitutions in the upstream half of the central bases of Pp (positions −63 to −58) were shown to cause a −25% greater decrease in transcriptional activation by P4 Delta than complementary base substitutions at the corresponding positions in the downstream half (positions −57 to −52). The asymmetric effects of these substitutions can now be interpreted in light of the current data, in which we have shown that the area from positions −55 to −53 is characterized by putative minor groove contacts, increased binding upon base removal at positions −55 and −54 and a more hydrophobic protein-DNA interface. It appears likely that NucC is interacting with the DNA differently in this area of the binding site.

These studies demonstrate that DNA binding by NucC differs in important respects from the DNA binding demonstrated for eukaryotic zinc finger transcription factors. The yeast zinc finger transcriptional activator Gal4 binds as a dimer to the consensus binding site CCG-N11-CCG (4, 7, 13, 19), and the only direct base contacts in the Gal4 crystal structure are to the conserved CCG repeats (34). In contrast, NucC binds as a tetramer and major groove contacts have been mapped throughout the NucC binding site. Thus, NucC and the related proteins in the P2 Ogr family of eubacterial zinc finger transcription factors provide a new paradigm for protein-DNA interaction. Further comparisons to other transcriptional activators will require structural studies of the NucC-DNA complex. Of particular interest will be the positions of the individual NucC subunits and the distribution of the DNA bend.

ACKNOWLEDGMENTS

This study was supported by National Science Foundation grant MCB9988252 (to G.E.C.) and by National Institutes of Health MARC Fellowship F31 GM17594 (to V.J.M.). Plasmid pLY2 was constructed by a laboratory rotation student, Ying Liu.

REFERENCES