The domain organization of *Nae*I endonuclease: Separation of binding and catalysis

JAMES D. COLANDENE AND MICHAEL D. TOPAL*

Lineberger Comprehensive Cancer Center and Department of Pathology, University of North Carolina Medical School, Chapel Hill, NC 27599-7295

ABSTRACT   *Nae*I is a remarkable type II restriction endonuclease. It must bind two recognition sequences to cleave DNA, forms a covalent protein–DNA intermediate, and is only 1 aa change away from topoisomerase and recombinase activity. The latter activities apparently derive from reactivation of a cryptic DNA ligase active site. Here, we demonstrate that *Nae*I has two protease-resistant domains, involving approximately the N-terminal and C-terminal halves of the protein, linked by a protease-accessible region of 30 aa. The domains were purified by cloning. The C-terminal domain was shown by gel mobility-shift assay to have approximately 8-fold lower DNA-binding ability than intact *Nae*I. Analytical ultracentrifugation showed this domain to be a monomer in solution. The N-terminal domain, which contains the catalytic region defined by random mutagenesis, did not bind DNA and was a mixture of different-sized complexes in solution implying that it mediates self-association. DNA greatly inhibited proteolysis of the linker region. The results identify the DNA-binding domain, imply that DNA cleavage and recognition are independent and separable, and lead us to speculate about a cleft-like structure for *Nae*I.

Restriction endonuclease *Nae*I appears to have special mechanistic and evolutionary importance. Showing its unusual heritage, *Nae*I must bind two DNA sequences to cleave DNA (1, 2). In addition, *Nae*I is the first restriction enzyme found to form a transient covalent intermediate with its newly cleaved substrate (3) much like the topoisomerases and recombinases. Finally, a single amino acid change, leucine to lysine at position 43 (L43K), radically transforms *Nae*I into a topoisomerase (3) and changes DNA recognition from double-stranded GCC/GGC to a preference for DNA containing unpaired bases (4). The L43K amino acid change is in a ligase-like active site that lies just upstream of the putative catalytic region (5). The ability of *Nae*I to act either as a restriction endonuclease or topoisomerase-recombinase depending on a single amino acid change is especially intriguing because *Nae*I has no detectable sequence homology with the topoisomerase or recombinase protein families. The unusual characteristics of *Nae*I suggest that it is either in the process of evolving from a topoisomerase to a restriction enzyme by abolishing its religation activity (3, 4) or may be able to serve both functions *in vivo*. *Nae*I may represent a missing link between some endonucleases and the topoisomerases and recombinases (4).

*Nae*I is a 317-aa polypeptide, has a molecular mass of 35 kDa (6), and forms a dimer in solution ($M_r = 70$) both in the absence and in the presence of cognate DNA (7). *Nae*I loops out DNA intervening distant *Nae*I DNA recognition sequences (8). This ability to loop out intervening sequences in DNA is highly unusual for a restriction enzyme but consistent with the requirement that *Nae*I bind two recognition sequences to activate DNA cleavage (1, 2).

Random mutagenesis shows that *Nae*I endonuclease has at least two functional regions: a region located within the N-terminal half of *Nae*I that mediates substrate binding and catalysis, and another region located within the C-terminal half of *Nae*I that binds DNA (5). We set out to map the protein-folding domains of *Nae*I for comparison with the regions defined by random mutagenesis, and to test whether the catalytic and DNA-binding functions could be separated by proteolysis. This knowledge is important because type II restriction endonucleases divide into subclasses in which sequence recognition and catalysis at the primary sequence level are either intertwined as in *Bam*HI, *Eco*RI, *Eco*RV, and *Pvu*II (reviewed in ref. 9) or independent as in *Fok*I (10). Knowledge of these domains will enable comparison with the known domain structures of restriction endonucleases and other protein families. The comparisons may offer clues to how these protein families are related.

Proteolytic cleavage provides a direct probe of protein folding. Regions of proteins accessible to proteases generally occur in extended linker regions or loops between tightly folded domains that are exposed on the surface of the protein (11). The domains can usually be purified for further analysis. These methods have been used to identify the DNA-binding domains of several sequence-specific DNA-binding proteins (12–20). Here, we use these methods to map the domain organization of *Nae*I. *Nae*I protein was overexpressed by using a maltose-binding protein (MBP) fusion system. The purified fusion protein had similar activity to wild-type *Nae*I. Limited proteolysis of *Nae*I apoenzyme demonstrated two major domains resistant to proteolysis, encompassing almost the entire N-terminal and C-terminal halves of the protein, linked by a region relatively accessible to cleavage by both trypsin and chymotrypsin. The two protease-resistant domains were cloned, expressed as MBP fusions, and analyzed for DNA cleavage activity and specific DNA binding. The C-terminal domain was found to be a monomer in solution and bound DNA containing cognate recognition sequence with a K_D 8-fold higher than that of *Nae*I. The N-terminal domain did not show DNA-binding ability and self-associated to give various-sized complexes in solution. The presence of DNA greatly reduced the susceptibility of the linker region to proteases. The change in susceptibility is the first indication of a protein conformational change associated with the switch from the inactive to active form of *Nae*I upon DNA binding (1).

MATERIALS AND METHODS

Materials. Amylose resin and protein molecular weight markers were purchased from New England Biolabs. G-50 Sephadex, S-Sepharose, Q-Sepharose, and Phosphocellulose resin were purchased from Sigma. *Escherichia coli* strain

*To whom reprint requests should be addressed at: Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC 27599-7295. e-mail: mdtopal@med.unc.edu.

3532 Biochemistry: Colandene and Topal

CAA1 (F^- $e14^-$ ($mcrA^-$) lacY1 or $D(lac)^6$ SupE44 galK2 galT22 mcrA rfbD1 mcrBa hsd(rk^-mk^+) M·$MspI^+$) and plasmid pNEB786, containing the *naeIR* gene, were obtained from New England Biolabs. Bovine factor Xa was obtained from Haematologic Technologies (Burlington, VT).

Protein Purification. The *NaeIR* gene and two *NaeIR* gene fragments were amplified from pNEB786 by PCR (21) and inserted into the plasmid vector pMALc2 (New England Biolabs) to give pMALc2:*Nae*I, pMALc2:*Nae*IΔC172, pMALc2:*Nae*IΔN168. The insertions were downstream from the *malE* gene, which encodes MBP and a factor Xa cleavage site so that either MBP::*Nae*I or, after cleavage with factor Xa, *Nae*I proteins could be isolated. Plasmid pMALc2:*Nae*I was transformed into *E. coli* strain CAA1 and selected on Luria–Bertani plates containing ampicillin (100 μg/ml). Twelve-liter cultures were grown from single colonies with aeration at 25–30°C. MBP::*Nae*I expression was induced with isopropyl β-D-thiogalactoside (0.5 mM), grown for an additional 6–8 hr, and harvested by centrifugation. Cell pellets were resuspended, washed in 0.9% NaCl, collected by centrifugation, and either processed or stored frozen. Cell pellets were resuspended in 4 volumes of column buffer (10 mM Tris·HCl (pH 7.4), 0.1 mM EDTA, 5% glycerol, 1 mM 2-mercaptoethanol, 50 mM NaCl). Phenylmethylsulfonyl fluoride (PMSF) was added to 1 mM, to inhibit serine proteases. The resuspended cells were sonicated on ice (1 min per each 10 ml) and immediately centrifuged at 30,000 × g for 30 min to remove cellular debris.

MBP::*Nae*I represented about 30% of total protein. Cell-free extract was passed over amylose resin equilibrated in column buffer plus 350 mM NaCl. Bound protein was eluted with 10 mM maltose in column buffer (fraction 2), which contained >70% of total protein as MBP::*Nae*I. Fraction 2 was further purified by phosphocellulose chromatography using a NaCl gradient (50 mM–1 M NaCl in column buffer). Active fractions (225 mM NaCl) were pooled (fraction 3). For free *Nae*I protein, fraction 3 was digested with factor Xa (0.2% of total protein by weight) at 4°C for 48 hr (fraction 4), dialyzed in *Nae*I storage buffer (column buffer without glycerol), and purified by S-Sepharose column chromatography by using storage buffer and a NaCl gradient (50 mM–1 M). Active fractions were pooled (fraction 5) and a minor amount of uncleaved MBP::*Nae*I protein was removed by using amylose resin (fraction 6). By these methods, *Nae*I was purified to >95%. MBP::*Nae*I(1–145) and MBP::*Nae*I(169–317) were purified by amylose resin chromatography. *Nae*I(169–317) was released from the fusion protein using factor Xa and purified to greater than 90% using phosphocellulose for further study.

Limited Proteolysis of *Nae*I. Trypsin and chymotrypsin (Boehringer Mannheim) digestions were performed by adding at 37°C either 32 ng of trypsin or 1.6 μg of chymotrypsin in 16 μl of protease reaction buffer (10 mM Tris·HCl, pH 8.0/0.1 mM EDTA, pH 8.0/1 mM 2-mercaptoethanol/100 mM NaCl/5% glycerol/5 mM CaCl₂) to 200 μl of protease reaction buffer containing 160 μg of purified *Nae*I. At various time points, 31-μl aliquots were removed and added to 0.3 μl of PMSF (100 mM in pure ethanol) to stop the reaction. SDS was added to 1%, the aliquots were heated to 100°C for 5 min, and reaction products were resolved by SDS/PAGE using 16% polyacrylamide gels. Gels were either stained with Coomassie brilliant blue-G250 or electroblotted onto poly(vinylidene difluoride) (PVDF) membrane for protein sequencing (22).

Amino Acid Sequence Analysis. Polypeptide fragments were visualized as bands on PVDF membrane by staining with Ponceau S in water. Individual bands were cut from the membrane and subjected to N-terminal amino acid sequencing by the University of North Carolina Micro Protein Chemistry Facility using a Perkin–Elmer/ABI model 492 Precise instrument. The sequence of the first seven amino acid residues was determined for each fragment analyzed.

DNA Cleavage Activity Assays. DNA cleavage activities of *Nae*I constructs were routinely monitored by examining their ability to cleave 0.5 μg of pBR322 as described elsewhere (3).

Gel Mobility-Shift Assays. Radiolabeled single-stranded deoxyribooligonucleotide (oligonucleotide) was annealed to complementary oligonucleotide to give cognate and noncognate probe DNAs. Probes were incubated with protein in a 20-μl reaction volume containing 10 mM Tris·HCl (pH 8.0), 10 mM CaCl₂, 20 mM NaCl, 10% glycerol, and BSA (0.1 mg/ml) for 20 min at room temperature. Cognate duplex DNA sequence was TTTCTCGCCACGTTCGCCGGCTTTC-CCCGTCAAGCT; recognition sequence (underlined) was substituted by GAAGAA to give noncognate DNA. Reactions were immediately loaded onto 6 or 10% polyacrylamide nondenaturing gels and fractionated by electrophoresis as previously described (7). Results were analyzed by using a PhosphorImager (Molecular Dynamics).

Effect of DNA Binding on Proteolysis of *Nae*I. To assess the effect of DNA binding on proteolysis of *Nae*I, 80 μg of *Nae*I in 100 μl of protease reaction buffer was incubated on ice for 30 min in the presence and absence of 107 μg of cognate DNA (TTGGGT**GCCGGC**AGGGTC). Protease digestion at 37°C was initiated by addition of 80 ng of trypsin or 800 ng of chymotrypsin. At timed intervals, 14 μl aliquots of reaction was mixed with 0.3 μl PMSF to terminate the reaction and fractionated by SDS/PAGE. To determine whether DNA was acting as a general protease inhibitor, this procedure was repeated using 80 μg of mammalian cell cycle protein p27 (23), which does not bind DNA, in place of *Nae*I.

Analytical Ultracentrifugation. The oligomeric states of MBP-fusion proteins were determined by sedimentation velocity measurements in a Beckman XL-A analytical ultracentrifuge. Fusion proteins were dialyzed in centrifugation buffer containing 10 mM Tris (pH 7.4), 0.1 mM EDTA, 1 mM 2-mercaptoethanol, 100 mM NaCl, 10 mM maltose, and 5% glycerol. Maltose ensures against oligomerization of MBP (24). Samples (A_{280} = 1.5) were filtered (0.45 μm) and centrifuged (12,000 rpm) to eliminate particulates. Protein samples (400 μl) were centrifuged at 50,000 rpm. Sedimentation boundary was monitored by A_{280}. Buffer density (1.0188 g/ml) was measured by specific gravity meter (DA-110M). Buffer viscosity (0.0114 poise) was determined from composition. Protein partial specific volumes were determined from amino acid composition (25). Molecular mass was derived from the Svedberg equation using the sedimentation coefficient (S) and diffusion coefficient (D) determined by the second boundary method (26), assuming globular spherical proteins with no hydration. S and D were corrected to standard state (25) using the AXIAL program (Les Holladay, Alza Corp., Palo Alto, CA). Molecular mass was also determined by direct fit of sedimentation velocity profile (27).

RESULTS

Proteolysis of *Nae*I Endonuclease. The absence of structural information for *Nae*I led us to map the domain structure of *Nae*I. By domains, we mean tightly folded regions of the protein resistant to proteolysis with trypsin and chymotrypsin. *Nae*I, purified as described in *Materials and Methods*, was incubated with limiting amounts of trypsin and chymotrypsin under conditions that digested the majority of full-length *Nae*I within 90 min (Fig. 1). Digestion with trypsin gave essentially three resistant fragments corresponding to molecular masses of 16.6, 17.3, and 19.1 kDa, which we refer to as fragments T16.6, T17.3, and T19.1 respectively. Digestion with chymotrypsin gave only two resistant fragments of approximately 13.5 (C13.5) and 19.1 (C19.1).

The N-terminal sequences of the stable fragments were determined by automated protein sequencing (Table 1). The protein fragment was mapped to the *Nae*I polypeptide by the
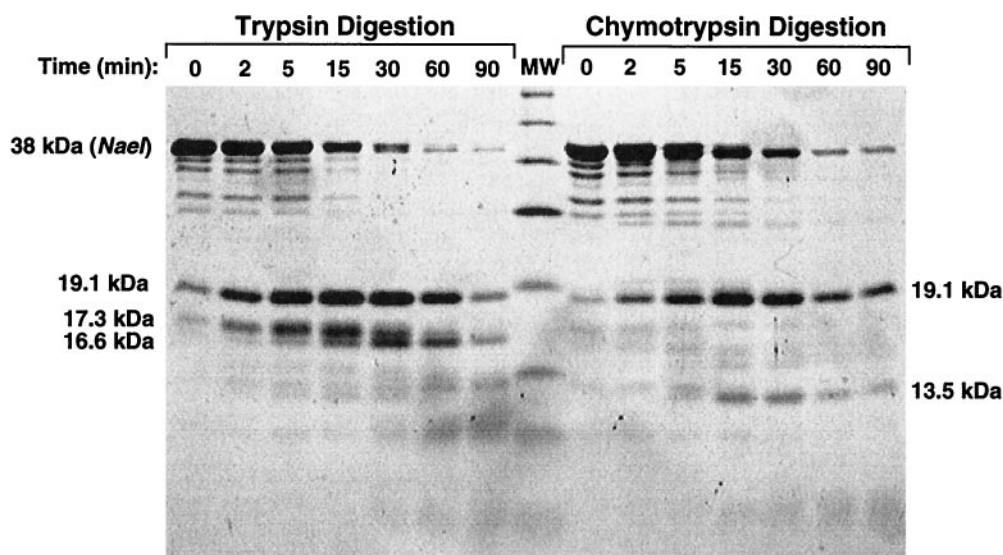
Fig. 1. Coomassie brilliant blue stained SDS-polyacrylamide gel showing pattern of polypeptide fragments produced by limited trypsin and chymotrypsin digestion of *Nae*I protein. The time of digestion is indicated along the tops of the lanes. The M_r of protease-resistant fragments analyzed in this study are indicated alongside the gel image and are based on the molecular weight (MW) markers glutamic dehydrogenase (55,561 Da), maltose-binding protein (42,710 Da), lactate dehydrogenase (36,487 Da), triosephosphate isomerase (26,625 Da), trypsin inhibitor (20,040–20,167 Da), lysozyme (14,313 Da), and aprotinin (6,517 Da). The digestion pattern at 0 min is caused by the amount of digestion taking place during mixing and sampling.

N-terminal sequence of the fragment. Protein fragment length was estimated from the molecular mass of the fragment as determined by SDS/PAGE (Fig. 1). From the fragment's N-terminal sequence and apparent molecular mass we mapped its location on the *Nae*I amino acid sequence as shown in Fig. 2. Based on this analysis, *Nae*I consists of two major protease-resistant domains linked by a region susceptible to protease cleavage. The protease-resistant domains reside within the C-terminal and N-terminal halves of the protein. To analyze the domains for DNA cleavage and binding, an attempt was made to separate the fragments by using ion-exchange chromatography (S-Sepharose, Q-Sepharose, and Phosphocellulose). The attempts failed to separate the two *Nae*I fragments and may indicate strong interactions between the two fragments. To overcome this problem, we cloned fragments of the gene, representing approximately the N-terminal (amino acids 1–145) and C-terminal (amino acids 169–317) domains, fused to MBP as described in *Materials and Methods* to give MBP::*Nae*I(1–145) and MBP::*Nae*I(169–317), respectively.

Proteins MBP::*Nae*I(1–145), MBP::*Nae*I(169–317), and MBP::*Nae*I were purified >70% over amylose resin columns. DNA cleavage activities were examined by incubating up to 1

μg of each protein with pBR322. The specific activity of MBP::*Nae*I was similar (within 2-fold) to that of native *Nae*I purified by other methods (28), indicating that MBP fusion has little effect on *Nae*I activity. Neither MBP::*Nae*I(169–317) nor MBP::*Nae*I(1–145) alone nor when combined showed detectable DNA cleavage activity.

DNA Binding. Gel mobility-shift assays (29) were used as described in *Materials and Methods* to determine whether MBP::*Nae*I(1–145) and MBP::*Nae*I(169–317) (100 μM each) could bind DNA. MBP::*Nae*I(169–317) showed strong binding affinity for DNA with *Nae*I cognate recognition sequence. The fraction of DNA bound was similar to that of the full-length *Nae*I fusion protein. No DNA-binding activity (specific or nonspecific) was detected for MBP::*Nae*I(1–145) at protein concentrations of up to 1.1 μM (data not shown).

The ability of *Nae*I(169–317) to bind DNA was quantitated and compared with that of *Nae*I after isolating the two polypeptides free of MBP as described in *Materials and Methods*. Apparent DNA-binding coefficients (K_D), or concentration at which 50% of the DNA was bound by protein, for *Nae*I and *Nae*I(169–317) binding to a 36-bp cognate DNA were approximately 5 nM for *Nae*I and 40 nM for *Nae*I(169–317) (Fig. 3).

Table 1. Polypeptide fragments of *Nae*I produced by limited proteolysis

Fragment*	Sequence of N terminus†	Sequence position of polypeptide fragments‡
T$_{19.1}$	LTPEGRA	151
T$_{17.3}$	MTELPLQ	1
T$_{16.6}$	MTELPLQ	1
C$_{19.1}$	LTPEGRA	151
C$_{13.5}$	QFAEPDD	7

*Fragments are named according to protease treatment (T, trypsin; C, chymotrypsin) and apparent molecular size of fragment (in kDa) based on relative mobility during SDS/PAGE (e.g., T$_{19.1}$ = fragment of approximately 19.1 kDa produced by trypsin digestion).

†Sequences shown were determined by automated amino acid sequencing of proteolytic reaction products as described in *Materials and Methods*.

‡The sequence position of each polypeptide fragment is given as the position of the first amino acid residue of the N-terminal end of the fragment relative to the amino acid sequence of *Nae*I.
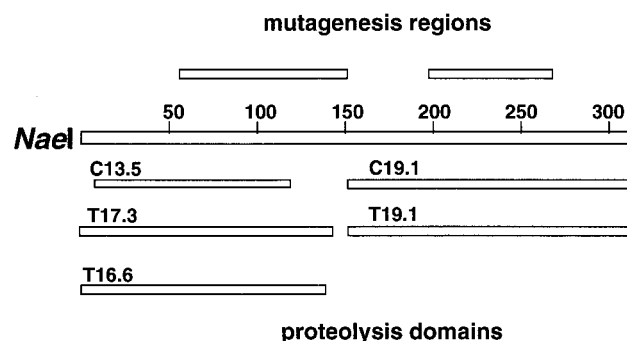

Fig. 2. Map of the locations of the protease-resistant fragments relative to full-length *Nae*I polypeptide. Also shown is the location of the catalytic and DNA-binding regions mapped by random mutagenesis (5).

Effect of DNA Binding on Proteolysis of *Nae*I. To gain insight into the domains that either interact with DNA in the intact *Nae*I molecule or undergo a conformational change upon binding DNA, we compared the trypsin and chymotrypsin digestion pattern in the presence and absence of DNA. We looked for regions of *Nae*I made either more sensitive or more resistant by DNA binding. The DNA was shown not to be a general inhibitor of either protease by incubation with mammalian cell cycle protein p27 (23) that does not bind DNA. Protein p27 was found to be equally susceptible to protease cleavage whether or not the DNA fragment was added to the reaction (results not shown).

The protease digestion pattern of *Nae*I was determined in the presence and absence of cognate DNA as described in *Materials and Methods*. The digestion reactions were analyzed by SDS/PAGE (Fig. 4). Digestion with both trypsin and chymotrypsin was greatly attenuated by *Nae*I interaction with DNA, and the chymotrypsin digestion pattern was altered as well. Whereas the 13.5-kDa band was formed to a similar extent whether or not DNA was present, formation of the prominent 19.1-kDa band (corresponding to the C-terminal domain) was inhibited when DNA was included in the reaction. This indicates that the C terminus of the linker region is less accessible after DNA binding than its N terminus. This conclusion is consistent with the complete loss of cleavage by trypsin because trypsin only cleaves near the C terminus of the linker region (see Fig. 2).

Self-Association. *Nae*I self-associates to form a dimer in solution (7). Analytical ultracentrifugation was used to determine the abilities of the *Nae*I domains to self-associate. From the determination of sedimentation and diffusion coefficients, the apparent molecular masses of MPB::*Nae*I, MPB::*Nae*I(1–145), and MPB::*Nae*I(169–317) were calculated assuming spherical globular proteins with no hydration. Sedimentation of the two domains

and wild-type protein fused to MBP are shown in Fig. 5. MBP::*Nae*I centrifugation shows the presence of two major species (Fig. 5*A*). One, a rapidly sedimenting complex, is apparently a mixture of various-sized complexes of *Nae*I. The other major species showed a sedimentation coefficient of 5.125×10^{-13} sec, diffusion coefficient of 3.46×10^{-7} cm^2/sec, and molecular mass of 147 kDa as determined by direct fitting of the sedimentation velocity profile (27). Similar values were obtained by other methods as described in *Materials and Methods*. The molecular mass determined is approximately two times the molecular mass of the monomeric MBP::*Nae*I protein (78 kDa) determined from its amino acid composition. Thus MBP::*Nae*I, like *Nae*I (7) is a dimer in solution, and also gave a rapidly sedimenting mixture of higher-ordered complexes that could be resuspended by mixing and recentrifuged to give the same sedimentation pattern. MBP::*Nae*I(169–317) showed the presence of a single species (Fig. 5*B*) with sedimentation coefficient of 3.151×10^{-13} sec, diffusion coefficient of 6.53×10^{-7} cm^2/sec, and molecular mass of 47 kDa. The apparent molecular mass value of 47 kDa is less than that determined from its amino acid composition (59 kDa); we infer that MBP::*Nae*I(168–317) is a monomer in solution. MBP::*Nae*I(1–145) showed a pattern corresponding to that expected for a mixture of higher-ordered complexes similar to that seen in the preparation of MBP::*Nae*I (Fig. 5*C*).

DISCUSSION

Domain Organization of *Nae*I. Limited digestion with trypsin and chymotrypsin reveals that *Nae*I is composed of at least two major domains that include amino acids 1–120 and 151–317. These two domains are linked by a short protease-sensitive region that includes amino acids 121–150. The boundaries of the protease-resistant domains approximately parallel
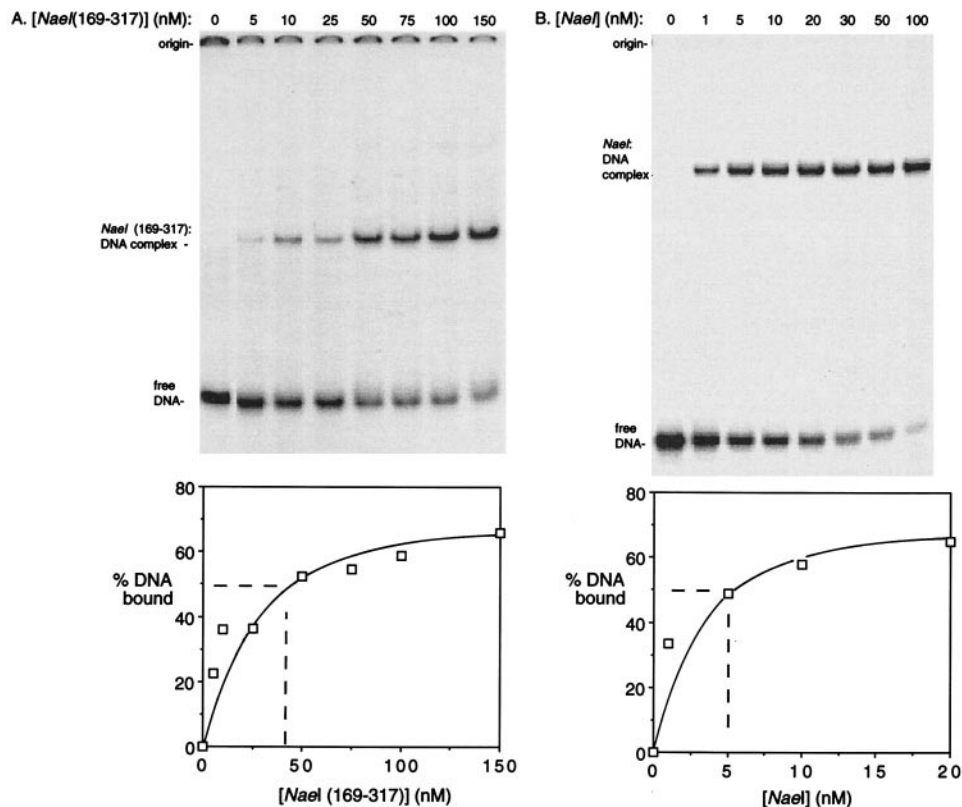

FIG. 3.   Determination of apparent K_D for *Nae*I and *Nae*I(169–317) binding to DNA using gel mobility-shift assay. DNA probe was cognate 36-mer double-stranded DNA (0.2 nM). The protein concentrations used in each reaction are shown above each lane. The reaction conditions are described in *Materials and Methods*. The band intensities as a function of protein concentration were quantitated by densitometry and are plotted in the graphs at the bottom.
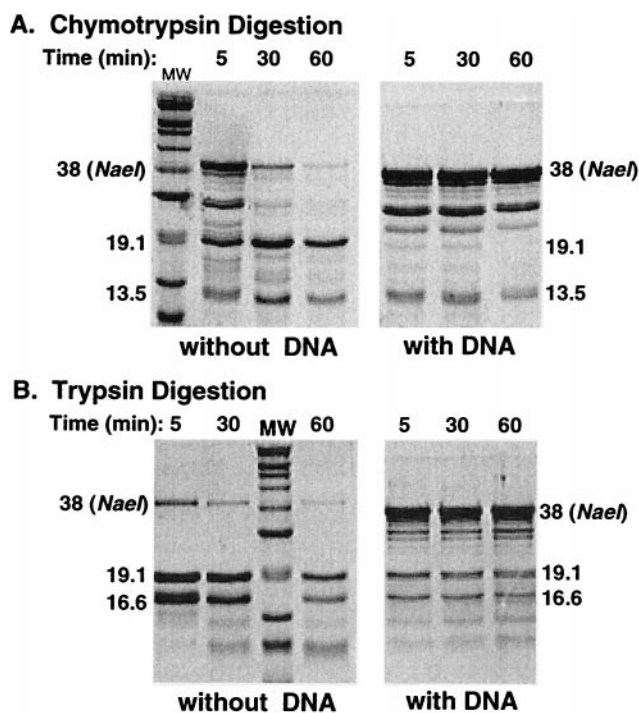
FIG. 4. Coomassie brilliant blue stained SDS-polyacrylamide gel showing pattern of polypeptide fragments produced by limited chymotrypsin (*A*) and trypsin (*B*) digestion of *Nae*I protein in the presence or absence of DNA. Digestion times are shown above the lanes. The M_r of protease-resistant fragments are indicated alongside the gel image and are based on the molecular weight (MW) markers described in the legend to Fig. 2.

(Fig. 2) the two functional regions defined by random mutagenesis (5). Cloning and characterization of the domains demonstrated that only the C-terminal domain detectably binds DNA. Thus, the separable N-terminal and C-terminal domains appear to have independent functions. Mutagenic studies imply that the N-terminal domain mediates catalysis: The L43K substitution gives *Nae*I topoisomerase activity (3) and T60I and E70K substitutions drastically reduce cleavage activity without reducing DNA binding (5). Mutagenic studies and the above DNA-binding results demonstrate that the C-terminal domain mediates sequence-specific DNA binding. Although it is clear from these studies that the C-terminal domain is mainly responsible for DNA recognition, it is not possible to fully rule out the presence in this domain of amino acids that contribute to catalytic function. No essential catalytic amino acids were located in this region by random mutagenesis, but the level of mutagenesis was not saturating.

The apparent independence of the catalytic and DNA-binding domains at the gene level is different from the type II restriction endonucleases for which information is available. For example, recognition and catalytic function are intertwined for *Bam*HI, *Eco*RI, *Eco*RV, and *Pvu*II (reviewed in ref. 9)—enzymes that recognize and cleave within a specific sequence. On the other hand, *Fok*I binds a specific recognition sequence but cleaves nearby the sequence and has separable recognition and catalytic domains (although not in the same order as *Nae*I) (10). *Nae*I, therefore, may be unusual because it cleaves within its recognition sequence, yet appears to have separable functions. The domain structure of *Nae*I suggests it may be more related to the IIs type endonucleases like *Fok*I than the type II enzymes like *Eco*RI.

DNA Binding. DNAs encoding the two protease-resistant *Nae*I domains were cloned and overexpressed in *E. coli*. Isolation of the peptides and analysis of their DNA-binding abilities demonstrate that only the C-terminal domain has
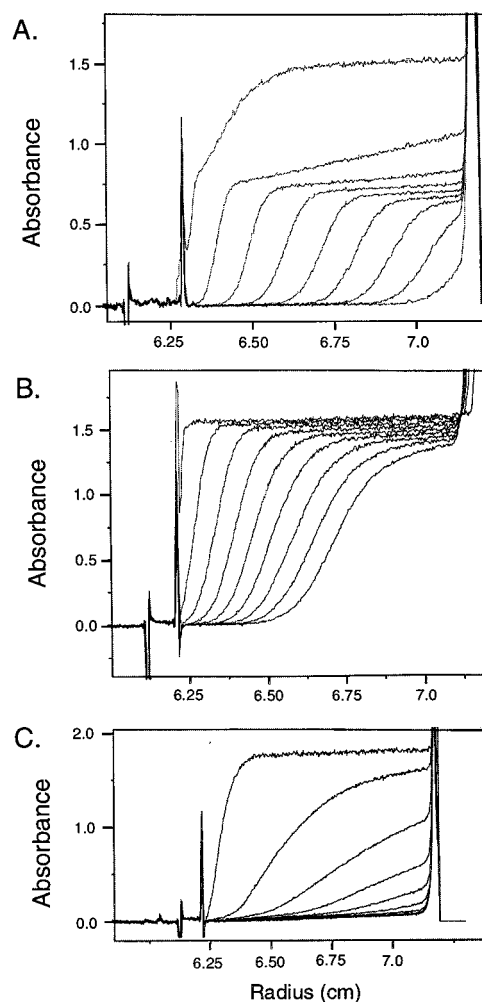

FIG. 5. Analytical ultracentrifuge determination of molecular weight of maltose-binding protein fusions with *Nae*I, and its N-terminal and C-terminal domains. Centrifugation was at 50,000 rpm; A_{280} boundary profiles are shown at 5, 10, 15, 20, 25, 30, 35, 40, and 45 min (from left to right). *x*-axis shows the radius of the sample cell (from left to right is the direction of centrifugal force); *y*-axis shows the A_{280} of the solution, measured through the sample cell. (*A*) MBP::*Nae*I; (*B*) MBP::*Nae*I(169–317); (*C*) MPB::*Nae*I(1–145).

detectable DNA-binding activity in solution. Measurements of apparent K_D for DNA binding show that the C-terminal domain accounts for approximately an eighth of the DNA-binding ability of intact *Nae*I. The C-terminal domain binds, however, as a monomer. Cooperative binding as a dimer in the wild-type protein could provide the missing binding energy.

The same DNA-binding measurements indicate that the N-terminal domain does not detectably bind DNA. Care must be taken, however, in concluding that the N-terminal domain has no DNA-binding capability. Mutagenesis studies (5) showed that some mutations in the N-terminal region modified DNA binding. These effects could be caused by disruptions to protein conformation that interfere with DNA binding to the C-terminal domain. Also likely, weak stabilizing interactions with DNA coming from the N-terminal domain, which must interact with DNA to catalyze cleavage in the native enzyme, may be undetectable when the N-terminal domain is separated from the major DNA-binding C-terminal domain.

Self-Association. How *Nae*I dimerizes is important because none of the known protein dimerization motifs are obvious from the *Nae*I amino acid sequence. Analytical ultracentrifugation demonstrated that the isolated N-terminal domain is a mix of various-sized complexes. The ability of the N-terminal

domain of *Nae*I to form various-sized complexes in solution implies that this domain mediates dimerization. Consistent with this conclusion, the C-terminal domain was shown to be a monomer in solution.

It is possible that self-association of the N-terminal domain contributes to the lack of DNA binding by this domain. We constructed other mutants with changes in the N-terminal domain that also form various-sized complexes in solution. The mutants bind DNA only if the C-terminal domain is intact (J.D.C. and M.D.T., unpublished results). The results show that formation of the various-sized complexes does not block *Nae*I accessibility to DNA.

Effects of DNA Binding on Proteolysis of *Nae*I. Proteolysis of *Nae*I was strongly inhibited by DNA. The inhibition can, in principle, be explained by either of two models. In the first model, DNA directly binding to the protease-sensitive linker region blocks access of the protease and thus inhibits cleavage. In the second model, binding of DNA to another region of the protein induces a conformational shift in *Nae*I that reduces access or sensitivity of the linker region to cleavage. The linker region of *Nae*I (amino acids 121–150) has a concentration of positive charges (5 of 12 residues) located toward its C terminus. The involvement of this region in DNA direct binding could explain why the addition of DNA more effectively inhibits cleavage of the C terminus of the linker region than it does the N terminus.

A Low-Resolution Model. Proteolysis demonstrates that *Nae*I has two tightly folded domains linked by a hinge region. The results imply independence of DNA binding and cleavage at the gene sequence level and show large effects of DNA on proteolysis of the hinge region. The requirement that both domains interact with DNA for sequence-specific cleavage leads us to speculate that docking of DNA in a crevice formed by these domains rearranges the linker region and makes it less accessible to protease. This low-resolution structure is consistent with the recently reported structure for *Fok*I in which recognition and catalysis reside in separate domains (10); *Fok*I has become a target for attempts to design new cleavage specificities based on domain swapping between proteins. Perhaps *Nae*I will also be useful for this purpose. In *Fok*I, the cleavage domain is folded away from the DNA by protein–protein interactions with the recognition domain (10). Apparently, this provides a mechanism for avoiding DNA cleavage until interaction with cognate recognition sequence when DNA interactions cause the cleavage domain to swing over to the DNA for cleavage (10, 30). Such a mechanism could also apply to *Nae*I with its separate cleavage and recognition domains apparently linked by a switch that is toggled by DNA binding.

1. Conrad, M. & Topal, M. D. (1989) *Proc. Natl. Acad. Sci USA* **86,** 9707–9711.
2. Yang, C. C. & Topal, M. D. (1992) *Biochemistry* **31,** 9657–9664.
3. Jo, K. & Topal, M. D. (1995) *Science* **267,** 1817–1820.
4. Jo, K. & Topal, M. D. (1996) *Nucleic Acids Res.* **21,** 4171–4175.
5. Holtz, J. K. & Topal, M. D. (1994) *J. Biol. Chem.* **269,** 27286–27290.
6. Taron, C. H., Van Cott, E. M., Wilson, G. G., Moran, L. S., Slatko, B. E., Hornstra, L. J., Benner, J. S., Kucera, R. B. & Guthrie, E. P. (1995) *Gene* **155,** 19–25.
7. Baxter, B. K. & Topal, M. D. (1993) *Biochemistry* **32,** 8291–8298.
8. Topal, M. D., Thresher, R. J., Conrad, M. & Griffith, J. (1991) *Biochemistry* **30,** 2006–2010.
9. Cheng, X., Balendiran, K., Schildkraut, I. & Anderson, J. E. (1994) *EMBO J.* **13,** 3927–3935.
10. Wah, D. A., Hirsch, J. A., Dorner, L. F., Schildkraut, I. & Aggarwal, A. K. (1997) *Nature (London)* **388,** 97–100.
11. Price, N. C. & Johnson, C. M. (1989) in *Proteolytic Enzymes: A Practical Approach*, eds. Beynon, R. J. & Bond, J. S. (IRL, Oxford), pp. 163–180.
12. Abdel-Meguid, S. S., Grindley, N. D. F., Templeton, N. S. & Steitz, T. A. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 2001–2005.
13. Smith, D. R., Jackson, I. J. & Brown, D. D. (1984) *Cell* **37,** 645–652.
14. Marzouki, N., Camier, S., Ruet, A., Moenne, A. & Sentenac, A. (1986) *Nature (London)* **323,** 176–178.
15. Huet, J. & Sentenac, A. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 3648–3652.
16. Nakayama, C., Teplow, D. B. & Harshey, R. M. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 1809–1813.
17. de Vargas, L. M., Pargellis, C. A., Hasan, N. M., Bushman, E. W. & Landy, A. (1988) *Cell* **54,** 923–929.
18. Boulanger, P. A., L'Etoile, N. D. & Berk, A. J. (1989) *Nucleic Acids Res.* **17,** 7761–7770.
19. Pan, H., Clary, D. & Sadowski, P. D. (1991) *J. Biol. Chem.* **266,** 11347–11354.
20. Adams, G. M. & Blumenthal, R. M. (1997) *Biochemistry* **36,** 8284–8292.
21. Clackson, T., Gussow, D. & Jones, P. T. (1991) in *PCR—A Practical Approach,* eds. Quirke, P. & Taylor, G. R. (Oxford Univ. Press, New York), pp. 187–214.
22. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith J. A. & Struhl, D., eds. (1987) *Current Protocols in Molecular Biology* (Wiley, New York).
23. Toyoshima, H. & Hunter, T. (1994) *Cell* **78,** 67–74.
24. Richarme, G. (1982) *Biochem. Biophys. Res. Commun.* **105,** 476–481.
25. Laue, T. M., Shah, B. D., Ridgeway, T. M. & Pelletier, S. L. (1992) in *Analytical Ultracentrifugation in Biochemistry and Polymer Science*, eds. Harding, S. E., Rowe, A. J. & Horton, J. C. (R. Soc. Chem., Cambridge, U.K.), pp. 90–125.
26. Goldberg, R. J. (1953) *J. Phys. Chem.* **57,** 194–202.
27. Philo, J. S. (1994) in *Modern Analytical Ultracentrifugation*, eds. Schuster, T. M. & Laue, T. M. (Birkhauser, Basel), pp. 157–170.
28. Yang, C. C., Baxter, K. B. & Topal, M. D. (1994) *Biochemistry* **33,** 14918–14925.
29. Fried, M. G. (1989) *Electrophoresis* **10,** 366–376.
30. Waugh, D. S. & Sauer, R. T. (1994) *J. Biol. Chem.* **269,** 12298–12303.