THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 278, No. 11, Issue of March 14, pp. 9733–9739, 2003
Printed in U.S.A.
© 2003 by The American Society for Biochemistry and Molecular Biology, Inc.

Amino Acid Substitutions at Position 43 of *Nae*I Endonuclease

EVIDENCE FOR CHANGES IN *Nae*I STRUCTURE*

Received for publication, September 9, 2002, and in revised form, December 31, 2002
Published, JBC Papers in Press, January 8, 2003, DOI 10.1074/jbc.M209192200

Kevin L. Carrick and Michael D. Topal‡

From the Lineberger Comprehensive Cancer Center and the Department of Biochemistry & Biophysics, University of North Carolina Medical School, Chapel Hill, North Carolina 27599-7295

Nae*I endonuclease contains a 10-amino acid region with sequence similarity to the active site KXDG motif of DNA ligase except for leucine (Leu-43) in *Nae*I (^{43}L*X*DG^{46}). Changing Leu-43 to lysine abolishes the *Nae*I endonuclease activity and replaces it with topoisomerase and recombinase activities. Here we report the results of substituting Leu-43 with alanine, arginine, asparagine, glutamate, and histidine. Quantitating specific activities and DNA binding values for the mutant proteins determined the range of amino acids at position 43 that alter *Nae*I mechanism. Substituting alanine, asparagine, glutamate, and histidine for Leu-43 maintained endonuclease activity, but at a lower level. On the other hand, substituting positively charged arginine, like lysine at position 43, converted *Nae*I to a topoisomerase with no observable double-strand cleavage activity. The specific activities of *Nae*I-43K and *Nae*I-43R and their relative sensitivities to salt, the topoisomerase-inhibiting drug *N*-[4-(9-acridinylamino)-3-methoxyphenyl]methane-sulfonamide (amsacrine) and single-stranded DNA showed that the two activities are similar. The effect of placing a positive charge at position 43 on *Nae*I structure was determined by measuring (for *Nae*I and *Nae*I-43K) relative susceptibilities to proteolysis, UV, circular dichroism spectra, and temperature melting transitions. The results provide evidence that a positive charge at position 43 induces dramatic changes in *Nae*I structure that affect both the Endo and Topo domains of *Nae*I. The identification of four putative DNA ligase motifs in *Nae*I leads us to speculate that structural changes that superimpose these motifs on the ligase structure may account for the changes in activity.

*Nae*I endonuclease (*Nae*I)[1] is a prototype for the type IIe (enhancer) (1, 2) restriction endonucleases, so named (3) because they require interactions with an enhancing DNA sequence. One DNA sequence acts as enhancer to induce cleavage of the other sequence (4–7). In solution, *Nae*I protein is a 70-kDa homodimer (7) composed of two 317-amino acid polypeptides (8, 9) that recognize and cleave at the arrow, the DNA sequence 5′-GCC ↑ GGC-3′ using only Mg^{2+} as a cofactor. Two-site binding gives *Nae*I a specificity of DNA recognition $\sim 10^4$-fold better than single-site-binding proteins that recognize similar sized sequences (3). The two-site binding loops out intervening DNA sequences (5), hinting at more complicated functions than monofunctional DNA cleavage. Substituting lysine for leucine at position 43 (L43K) in *Nae*I endonuclease abolishes restriction endonuclease activity and in its place gives topoisomerase and recombinase activities (10). In addition, substitution L43K results in a preference for binding of single-stranded DNA and a sensitivity to salt and intercalative topoisomerase-inhibiting drugs, such as *N*-[4-(9-acridinylamino)-3-methoxyphenyl]methane-sulfonamide (amsacrine), which is lacking in restriction endonucleases but characteristic of topoisomerases (11, 12).

*Nae*I position 43 is located near the C terminus of α-helix H2, which is part of the central hydrophobic core of the Endo domain (Ref. 13 and Fig. 1). Unlike the structures of most restriction endonucleases, *Nae*I contains two separate domains, both of which bind DNA (14). The N-terminal, Endo domain contains the restriction endonuclease cleavage motif found in restriction enzymes as well as repair nucleases mutH, Vsr, and λ exonuclease (14–17) and transposases (18). The C-terminal, Topo domain contains a CAP motif also found in topoisomerases IA and II (19). Position 43 is positioned at the bottom edge of the Endo domain almost between the Endo and Topo domains and at the boundary separating the two *Nae*I monomers (Fig. 1) The two Leu-43 residues lie toward the center of the *Nae*I dimer ~15 Å apart.

The *Nae*I position 43 lies within a 10 amino acid region with similarity to the conserved, active-site KXDG motif for DNA ligases, RNA ligases, and RNA-capping enzymes, which together make up the nucleotidyl transferase superfamily (for discussions, see Ref. 20). Nucleotidyl transferase catalysis involves three steps. First, the ligase is activated by the formation of a covalent protein-AMP intermediate with the AMP linked to the ϵ-amino group of lysine by a phosphoramidate bond. The conserved lysine in the sequence KXDG forms the adenylated intermediate using the high energy cofactors ATP (generally found in eukaryotes, viruses, and Archaebacteria) and NAD^+ (generally found in Eubacteria). Second, the AMP moiety is transferred from lysine to the 5′-phosphate at the nicked DNA. Finally, the DNA-free ends are joined in an enzyme-dependent reaction with loss of AMP. *Nae*I has leucine instead of the essential lysine at position 43 (^{43}L*X*DG^{46}). The topoisomerase activity of *Nae*I-43K is possibly the result of activating a cryptic ligase active site and thereby coupling restriction endonuclease cleavage with ligation. *Nae*I forms a covalent intermediate with a cleaved substrate (10), which may serve as

* This work was supported by National Institutes of Health Grant GM52123. 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.

‡ To whom correspondence should be addressed: Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC 27599-7295. Tel.: 919-966-8208; Fax: 919-966-3015; E-mail: mdtopal@med.unc.edu.

[1] The abbreviations used are: *Nae*I, *Nae*I endonuclease; CAP, catabolite-activating protein; ES, enzyme-substrate complex; CD, circular dichroism; oligonucleotides, oligodeoxyribonucleotides; Topo, topoisomerase; MBP, maltose-binding protein; ssDNA, single-stranded DNA; amsacrine, *N*-[4-(9-acridinylamino)-3-methoxyphenyl]methane-sulfonamide.

This is an Open Access article under the CC BY license.

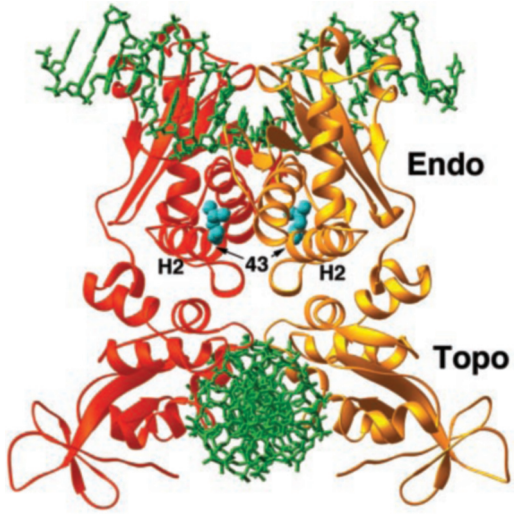

FIG. 1. ***Nae*I-DNA crystal structure (14).** Wild-type (*-L43*) *Nae*I protein dimer is shown as *red* and *gold ribbons*. DNA is shown in *green*. Leucine at position 43 is shown as *cyan ball* and *stick* and oversized for clarity. Endo (N-terminal) and Topo (C-terminal) domains are indicated.

the high energy intermediate needed for ligation, as is the case with the topoisomerases and recombinases. The amino acid that covalently links *Nae*I to its DNA substrate has not been identified. There is no similarity between the folds of *Nae*I (13, 14) and DNA ligase (21–24). Moreover, the K*X*DG-like motif in *Nae*I lies away from the endonuclease metal-binding site necessary for cleavage. Thus, the transformation to topoisomerase activity in L43K implies a conformational change in the ES complex that results in the K*X*DG-like motif lying closer to the phosphodiester scissile bond; the endonuclease fold may be altered to mimic aspects of the ligase fold found at the active sites of the DNA ligases.

To learn whether lysine at position 43 is unique in its ability to give *Nae*I topoisomerase activity, we substituted alanine, asparagine, glutamate, histidine, and arginine. We also substituted lysine for leucine at position 40. Alanine has a small, nonpolar, amino acid side chain. Asparagine has an uncharged, amide-bearing polar side chain. Like lysine, histidine, arginine, and glutamate all have charged, polar side chains depending on the pH. We report that alanine, asparagine, glutamate, and histidine at position 43 maintained *Nae*I wild-type endonuclease activity when substituted at position 43, but with significant decreases in DNA cleavage. Substitution with arginine, however, resulted in topoisomerase activity identical to that of *Nae*I-43K. *Nae*I and *Nae*I-43K susceptibility to protease, UV-circular dichroism (CD) spectra, and CD temperature melting transitions were determined. The results provide the first evidence that placing a positive charge at position 43 causes a dramatic change in *Nae*I folding. The finding of additional putative ligase motifs within the *Nae*I sequence offers a possible rationale for how the conformational changes may alter *Nae*I activity.

EXPERIMENTAL PROCEDURES

Materials—Escherichia coli strain CAA1 (*F*$^-$*e14*$^-$ (*mcrA*$^-$) *lacY1 or D* (*lac*)^6 *SupE44 galK2 galT22 mcrA rfbD1 mcrBa hsd* (r_k^- m_k^+) *M·Msp*I$^+$) and plasmid pNEB-786, containing the *Nae*IR gene, were obtained from New England Biolabs. Plasmid pMAL-C2 and amylose resin were purchased from New England Biolabs. Substrate pBR322 was purchased from Promega Corp. Amsacrine, a DNA-intercalating drug that inhibits topoisomerase activity was purchased from Topogen Inc. Oligodeoxyribonucleotides (oligonucleotides) were synthesized by the Nucleic Acid Core Facility at UNC. Cellulose phosphate, sp-Sepharose, and heparin resins were purchased from Sigma. Cognate (dTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCT) and noncognate (dTTTCTCGCCACGTTCGAAGAATTTCCCCGTCAAGCT) oligonucleotides were annealed to their complements to yield DNA fragments, which were gel purified.

NaeI Substitutions—The *Nae*IR gene was subcloned into the expression vector pMAL-C2, and site-directed mutagenesis was performed using the method of Clackson *et al.* (25). Mutated *Nae*IR genes were sequenced to confirm the mutation and ensure that no secondary mutations were generated.

Purification of NaeI Mutants—To express the fusion protein, pMAL-*Nae*I mutants (MBP·*Nae*I) were induced with 1 mM isopropyl-β-D-thiogalactoside (IPTG). Cells were harvested by centrifugation and cell pellets resuspended in 4 volumes of Tris column buffer (10 mM Tris-HCl, pH 7.4, 0.1 mM EDTA, 5% glycerol, 1.0 mM 2-mercaptoethanol, 50 mM NaCl) with 1 mM phenylmethylsulfonyl fluoride to inhibit serine proteases. The resuspended cell pellet was sonicated on ice for 1 min per 10 ml followed by centrifugation at 30,000 × *g* to remove cellular debris. The supernatant was subjected to amylose resin chromatography. The column was washed with column buffer containing 400 mM NaCl then equilibrated with column buffer containing 50 mM NaCl. Fusion protein was eluted in Tris column buffer containing 10 mM maltose. Maltose had no effect on *Nae*I and *Nae*I-43K activities.

NaeI—43R was cloned into pNEB786 (to give pNEB786-*Nae*I-43R) and isolated from cells containing this plasmid. Cell extracts showed topoisomerase activity absent from extracts from cells containing pNEB786. Phosphocellulose, sp-Sepharose, and heparin columns were used for purification. Protein elution was achieved by NaCl gradients in column buffer. Fractions were assayed for topoisomerase activity. Peak fractions containing Topo activity were pooled and dialyzed with column buffer containing 50 mM NaCl after each column. Heparin fractions were dialyzed against 50 mM NaCl in column buffer for storage. *Nae*I-43R identity was confirmed by Western analysis using an affinity-purified antibody prepared against wild-type *Nae*I. The final purities of the proteins was estimated to be >90% based on optical density measurements of the protein resolved by SDS-PAGE by standard procedures.

Restriction Endonuclease Activity—Restriction endonuclease activities relative to wild-type *Nae*I were determined from measurements of steady-state DNA cleavage rates. Protein concentrations were titrated while keeping reaction time (30 min) and DNA concentration (500 nM) constant. Reactions were prepared in 15-μl total volume to contain 10 mM Tris-HCl, pH 8.0, 20 mM NaCl, 5 mM $MgCl_2$, 5 mM 2-mercaptoethanol, and bovine serum albumin (0.1 mg/ml). Reactions were incubated at 37 °C for 30 min. Substrate was cognate DNA radiolabeled with [α-^{32}P]ATP using T4 polynucleotide kinase. Reactions were stopped by addition of EDTA (40 nM) and glycerol (10%). Reaction products were separated on 8% polyacrylamide gels and analyzed using a Molecular Dynamics Storm 540 PhosphorImager. The cleaved and uncleaved gel bands were quantified using Imagequant 5.0 from Molecular Dynamics.

Gel Mobility Shift Assays—Single-stranded oligonucleotides were radiolabeled and annealed to complementary oligonucleotides to give probe. Protein and probe were incubated in a 20-μl volume containing 10 mM Tris-HCl (pH 8.0), 10 mM $CaCl_2$, 20 mM NaCl, 10% glycerol, and bovine serum albumin (0.1 mg/ml). Reactions were incubated at 25 °C for 20 min. Reaction products were analyzed by PAGE (6%). Apparent K_D values (defined by the protein concentrations necessary to shift half the amount of probe) were determined with cognate probe. Nonspecific binding was tested using noncognate probe.

Assay of NaeI-43R Topoisomerase Activity—Plasmid pBR322 (11.6 nM) and *Nae*I-43R (0.21 μM) were incubated at 37 °C for 30 min in 10 μl containing 10 mM Tris-HCl (pH 8.0), 20 mM NaCl (except for the assay of NaCl dependence), 5 mM $MgCl_2$, bovine serum albumin (0.1 mg/ml), and 5 mM β-mercaptoethanol. Its weak binding to DNA (12) necessitated the relatively high concentration of *Nae*I-43R. The DNA binding affinity was similar to that of *Nae*I-43K. Reactions were stopped by addition of SDS to 1%. Products were resolved on 1.0% agarose gels containing 0.5 μg/ml ethidium bromide in the gel and in the running buffer. Assays for the effects on *Nae*I-43R activity of single-stranded DNA, NaCl, and amsacrine varied the concentrations of these, as indicated.

Limited Proteolysis—Trypsin (2.3 ng) was added to 7 μg of *Nae*I or *Nae*I-43K, and 1-μg aliquots were removed at the times indicated. Phenylmethylsulfonyl fluoride (1 mM) and 1.0% SDS were added to the aliquots, which were then heated to 100 °C for 10 min. Reaction products were resolved by SDS-PAGE (5% stacking and 15% resolving). The gel was stained with Coomassie Blue and photographed.

*Size-exclusion Chromatography—Nae*I and *Nae*I-43K were sized relative to known molecular weight proteins using chromatography through Sephacryl S-200 resin (32-cm column) pre-equilibrated with phosphate column buffer (20 mM potassium phosphate (pH 7.4), 0.1 mM

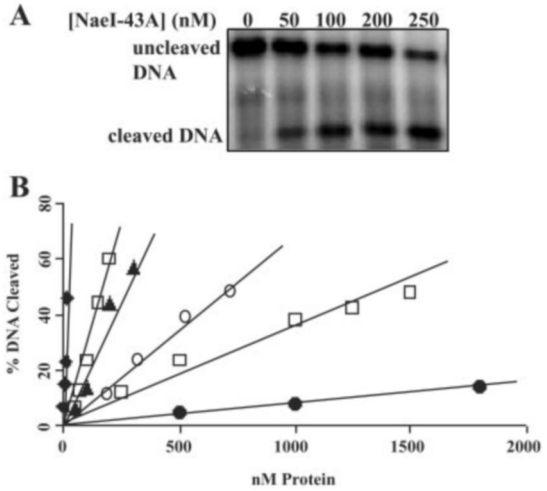

FIG. 2. **Increase in restriction endonuclease DNA cleavage activity with increase in MBP·*Nae*I mutant protein concentrations.** *A*, MBP·*Nae*I-43A was incubated with 500 nM DNA substrate for 30 min. Reaction products were separated by 8% PAGE and visualized by phosphorimager. *B*, plots of % DNA substrate cleaved *versus* nanomolar *NaeI* mutant protein for position 43 substitutions, except for *Nae*I-43K and -43R, which do not have detectable activities as fusion proteins and have DNA relaxation activities, but no restriction enzyme activities when fused to MBP. Wild type (♦), 43A (□), 43H (▲), 43N (■), 43E (●).

EDTA, 5% glycerol, 1.0 mM 2-mercaptoethanol, 50 mM NaCl). Protein (50 μl at 0.1 mg/ml) was loaded and eluted at 0.15 ml/min. Absorbance was monitored at 280 nm and the elution volume determined. The void volume was determined using blue dextran. Chromatography was performed with a Bio-Rad Biologic Chromatography System.

Circular Dichroism—UV-CD spectra were measured using an Applied Photophysics PiStar-180 spectrometer. *Nae*I and *Nae*I-43K were extensively dialyzed into 10 mM phosphate, pH 7.0 (buffer conditions that showed similar specific activities to that measured using the above assay conditions). CD spectra were measured for *Nae*I, *Nae*I-43K, and buffer between wavelengths 185 nm and 260 nm at 25 °C. Buffer spectrum was subtracted from *Nae*I and *Nae*I-43K spectra. The concentrations of both *Nae*I and *Nae*I-43K were 0.1 mg/ml as determined from absorbance measurements at 280 nm. CD measurements were also made at a wavelength of 208 nm while increasing the temperature from 10 to 90 °C.

RESULTS

Activity of NaeI Mutants—All of the mutants were expressed as fusions with maltose-binding protein. The effects of substitutions at position 43 were determined. DNA specific activities for each mutant were determined from the slope of the line defined by amount cleaved per amount of protein (Fig. 2). The plots of DNA cleavage *versus* amount of protein were reproducible and linear over the entire range of protein concentration (Fig. 2 and Table I). Substitutions with alanine, asparagine, glutamate, and histidine retained endonuclease activity. The cleavage patterns of the fusion protein mutants were identical to that of wild-type *Nae*I fused to maltose-binding protein (MBP·*Nae*I). The specific activities of the mutants, however, were reduced compared with that of MBP·*Nae*I. When *Nae*I-43K and -43R were expressed as protein fusions with maltose-binding protein neither restriction endonuclease activity nor topoisomerase activity was detected. *Nae*I, expressed from this vector, showed no topoisomerase activity. Therefore, *Nae*I-43K and -43R were expressed in *E. coli* from pNEB786 where they both demonstrated topoisomerase activities. Therefore, all of the mutants were studied as fusion proteins (MBP·*Nae*I-43A, −43H, −43E, −43N), except for *Nae*I-43K and *Nae*I-43R.

Relative DNA Binding—The apparent K_D values were determined for *Nae*I mutant interactions with DNA from the amount of protein necessary to shift half of the cognate DNA

TABLE I
Effects of amino acid substitutions for Leu-40 and Leu-43 on endonuclease activity

Mutant[a]	Sp. activity	Wild type	app K_D	Wild type
	(×1000)	*%*	*nM*[b]	*%*
Wt	30 ± 3[c]	100	10.0 ± 0.2	100
L40K	0.8 ± 0.05	3 ± 1	1250 ± 75	0.8 ± 0.6
L43A	3.7 ± 0.01	12 ± 1	10 ± 0.2	100 ± 4
L43E	0.1 ± 0.1	0.3 ± 0.4	225 ± 5	4.4 ± 0.6
L43H	1.5 ± 0.05	5 ± 1	10 ± 0.2	100 ± 4
L43N	0.5 ± 0.3	2 ± 1	125 ± 2	8.0 ± 0.3
L43K	ND[d]			
L43R	ND			

[a] All mutants were expressed as fusions with MBP.
[b] Apparent K_D is abbreviated app K_D.
[c] Errors are variation from average of two determinations. Errors for normalized % values are based on propagation of errors.
[d] ND, not detectable. L43K and L43R showed no detectable endonuclease activity when expressed either as protein fusions or unfused. Unfused they demonstrated DNA relaxation activities.

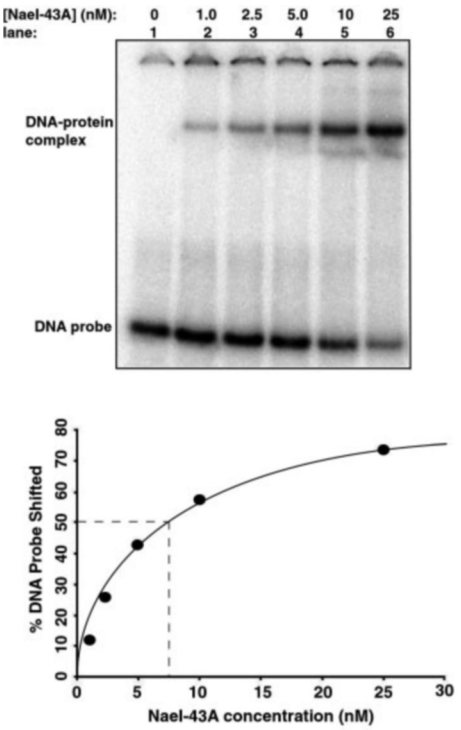

FIG. 3. **Gel mobility-shift assay of *Nae*I-43A.** *Upper panel*, PAGE (6%) analysis of complexes formed from incubation of increasing concentration of *Nae*I-43A with radiolabeled DNA probe. Bands were visualized by phosphorimager and quantitated by Imagequant software. *Lower panel*, plot of percent DNA shifted *versus* increasing *Nae*I-43A protein concentration. Apparent K_D is defined as the protein concentration required to shift half of the starting amount of DNA probe.

probe during PAGE (Fig. 3 and Table I). DNA-binding results are the average of two determinations. The apparent K_D values varied significantly from wild type only for MBP·*Nae*I-43N, -43E, and -40K, which gave values of 125 ± 2, 225 ± 5, and 1250 ± 75 (nM), respectively. The gel-shift results show a small amount of density in the wells probably due to a small amount of aggregated protein that binds DNA. The small amount of protein in the wells is counted in the determination of relative K_D values.

Comparison of NaeI-43R and NaeI-43K Activities—*Nae*I-43K has topoisomerase activity rather than endonuclease activity, which is sensitive to salt, amsacrine, and ssDNA (10–12). Incubating similar concentrations of *Nae*I-43R and *Nae*I-43K with pBR322 resulted in almost identical banding ladders

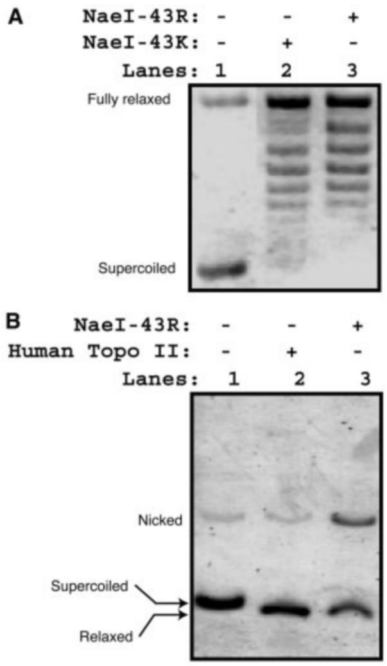

FIG. 4. **Topoisomerase activity of *Nae*I-43R.** *A*, *Nae*I-43K and *Nae*I-43R proteins (0.21 μM) were incubated with pBR322 (11.6 nM) for 30 min. at 37 °C. The products were resolved by 1% agarose gel electrophoresis. The products were visualized by ethidium bromide staining. *B*, reaction products from above reactions were analyzed as above but gel contained ethidium bromide (0.5 μg/ml). Commercially available human Topo II was used in a control reaction.

characteristic of topoisomerase activity (Fig. 4*A*). The differences between the ladders in Fig. 4*A* can be attributed to small differences in specific activities between the two protein preparations.

Nicking *versus* relaxation by *Nae*I-43R was assayed to determine the relative amounts of nicked *versus* covalently closed, fully relaxed products produced (Fig. 4*B*). About half the final products were the latter. *Nae*I-43K results in a similar amount of nicked *versus* relaxed DNA (26).

The effects of varying salt concentration on the relaxation reaction were determined (Fig. 5). The optimum NaCl concentration was below 30 mM. At 210 mM salt, relaxation was completely inhibited. The *Nae*I-43R mutation also made *Nae*I sensitive to the intercalative drug amsacrine (Fig. 5*B*). Inhibition of *Nae*I-43R Topo activity by amsacrine was apparent at a concentration of 5 μM with near total inhibition at 10 μM. Single-stranded DNAs, containing the *Nae*I cognate or noncognate sites, were tested for their ability to inhibit *Nae*I-43R activity. Nearly complete inhibition of *Nae*I-43R occurred with 10 nM ssDNA with or without cognate recognition sequence (Fig. 5, *C* and *D*). The results are identical to those for *Nae*I-43K.

Effect of L43K Substitution on NaeI Structure—To determine the effect of a positive charge at position 43 on *Nae*I structure, *Nae*I and *Nae*I-43K conformations were probed using size-exclusion chromatography, limited proteolysis, and circular dichroism. *Nae*I-43K elutes by size-exclusion chromatography at the same volume as *Nae*I, indicating that it is a dimer of about 70 kDa (Fig. 6). Ultracentrifugation (27), gel filtration (7), and crystallization (13, 14) show that the preferred structure of *Nae*I in solution is a dimer. Trypsin was used to probe the domain structure of *Nae*I-43K relative to *Nae*I. The sites accessible to trypsin cleavage were clearly different between *Nae*I-43K and *Nae*I (Fig. 7). *NaeI* showed stable domains at molecular sizes of 16 and 19 kDa. *Nae*I-43K showed no stable domains and a significantly different initial banding pattern

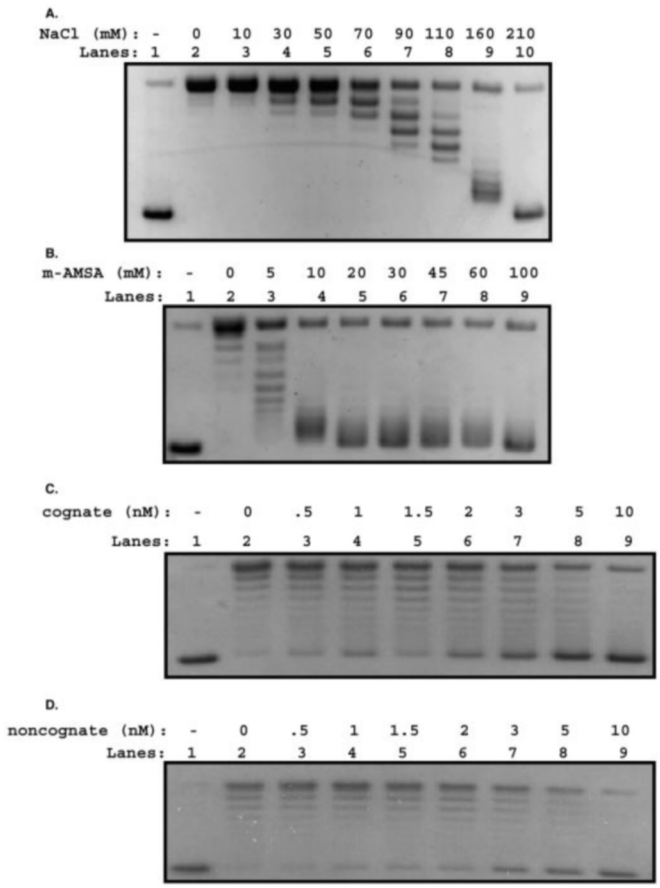

FIG. 5. **Effects of NaCl (*A*), amsacrine (*B*), and single-stranded (*ss*) DNA (*C* and *D*) on the topoisomerase activity of *Nae*I-43R.** Reaction conditions were as defined in the Fig. 4 legend and under "Experimental Procedures." Reaction products were resolved by 1% agarose gel and stained with ethidium bromide.

from *Nae*I at 5-min digestion with trypsin (Fig. 7). Additionally, *Nae*I was more resistant to trypsin cleavage than *Nae*I-43K.

The protease digestion experiments implied conformational differences between *Nae*I and *Nae*I-43K. To confirm this, we measured the UV CD spectra of both proteins. CD measurements showed distinct differences between the CD spectra of *Nae*I and *Nae*I-43K (Fig. 8*A*). Most notably, the minima between 205 and 230 nm were larger for *Nae*I-43K, with a distinct difference at the α-helical characteristic wavelength of 222 nm. The CD curves were reproducible and overlapped above λ of 250 nm, which indicated no significant concentration differences. CD was also used to monitor the temperature melting profiles of the two proteins to determine their relative T_m values. The thermal transition point was determined at λ of 208 nm, which is within the wavelength area where the CD values are most sensitive to protein structure (28). *Nae*I and *Nae*I-43K gave well defined melting points of 56 $\pm$ 0.3 °C and 59 $\pm$ 0.3 °C, respectively (Fig. 8*B*). The melting point values are the averages of two determinations.

Visual Inspection of the NaeI Amino Acid Sequence—Initial inspection of the *Nae*I sequence led to the discovery of a 10 amino acid sequence in *Nae*I with similarity to motif I of DNA ligase, the KXDG motif (10). Visual inspection, taking into consideration the secondary structures of the DNA ligase motifs, identified three additional regions of *Nae*I protein sequence (Fig. 9) with similarity to three additional motifs that define the DNA ligase enzyme family (20–24, 29). Fig. 9 shows the four ligase motifs for several of the DNA ligases and for

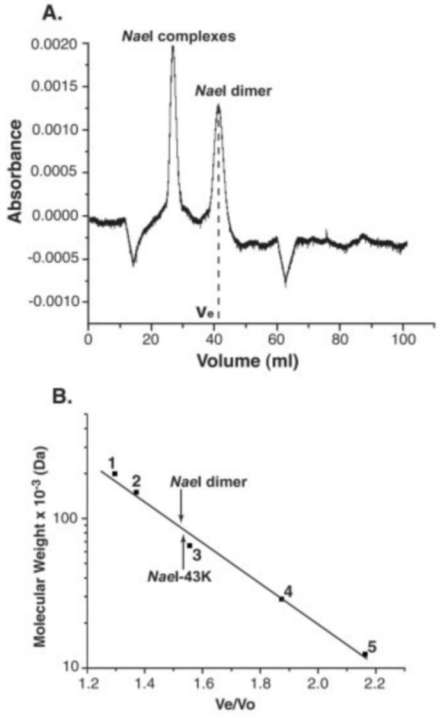

FIG. 6. **Determination of the native molecular weight of *Nae*I-43K by size-exclusion chromatography.** *A*, chromatograph for *Nae*I on Sephacryl S-200 (see "Experimental Procedures"). Ultracentrifugation shows *Nae*I in solution to be a mixture of higher order complexes and dimer (27). Thus, the first peak is assigned to the higher order complexes of *Nae*I, which eluted at the void volume (V_o). The two V-shaped dips were probably caused by bubbles. The elution volume (V_e) for the *Nae*I dimer is defined as the volume at which the peak maximum eluted. *B*, (V_e)/(V_o) is plotted against the log of the molecular weight for: 1) sweet potato β-amylase (200 kDa), 2) yeast alcohol dehydrogenase (150 kDa), 3) bovine albumin (66 kDa), 4) bovine carbonic anhydrase (29 kDa), and 5) horse cytochrome *c* (12.4 kDa). *Nae*I (70 kDa) is a dimer (7, 13).

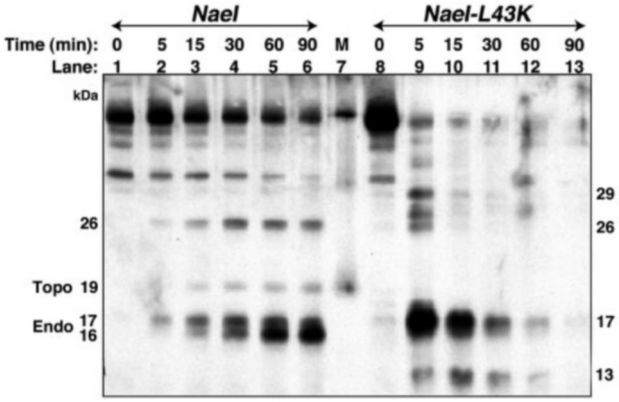

FIG. 7. **Limited proteolysis of *Nae*I and *Nae*I-43K.** Products of incubation of substrate protein (7 μg) with trypsin (2.3 ng) SDS-PAGE (15%) for the time intervals indicated. Gel was stained with Coomassie Blue. Stable Endo and Topo domains of *Nae*I are indicated.

*Nae*I. The *Nae*I amino acid sequence regions are shown at the top of Fig. 9 in *bold*. Amino acids in the DNA ligase motifs in the same exchange group (30) as the corresponding amino acid in *Nae*I are shown in *bold* and *underlined*.

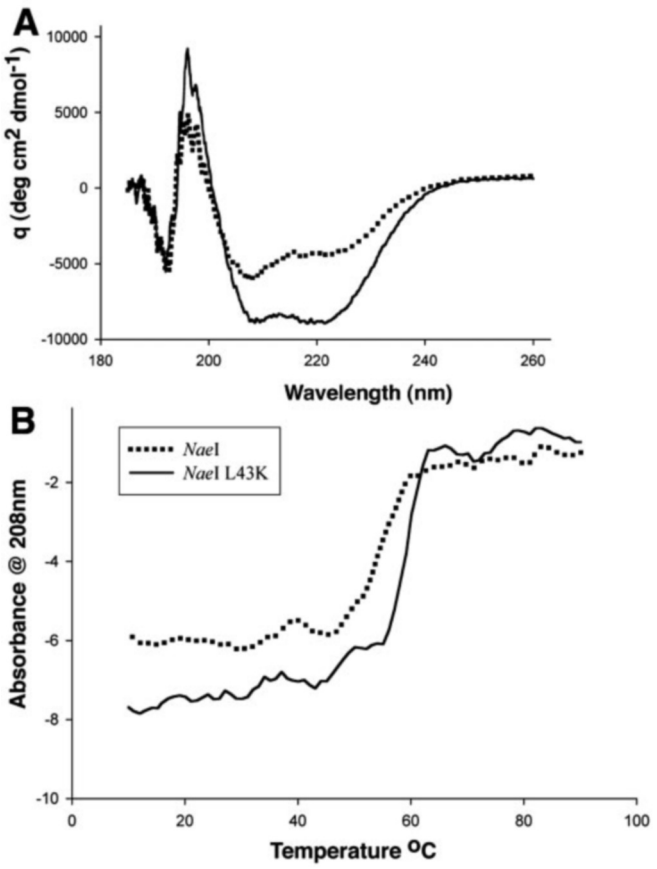

FIG. 8. **CD measurements of *Nae*I and *Nae*I-43K.** *A*, UV-CD spectra for *Nae*I (—) and *Nae*I-43K (—) at 25 °C. *B*, CD measurements at λ of 208 nm (chosen because it lies within the wavelength region where the CD shows maximum dependence on protein structure, Refs. 28 and 31) for *Nae*I (—)and *Nae*I-43K (—) as a function of temperature from 10 to 90 °C. Protein concentrations were 0.1 mg/ml, as determined by absorbance at 280 nm.

	I	III	IIIa	IV
NaeI-43K	$_{43}$KYDGQR -34-	$_{83}$FETDYEIA -35-	$_{126}$CAWTAGLVKVIP -33-	$_{171}$ENLLL
Hu1 ATP	KYDGQR -42-	FILDTEAV -32-	CLYAFDLIYLNG -52-	EGLMV
Hu3 ATP	KYDGER -41-	MILDSEVL -28-	CLFVFDCIYFND -52-	EGLVI
Spo ATP	KYDGER -42-	FILDCEAV -32-	CLFAFDILYLNG -52-	EGLMV
Sce ATP	KYDGER -42-	LILDCEAV -32-	CLFAFDILCYND -52-	EGLMV
Vac ATP	KYDGER -41-	IVLDSEIV -27-	CLFVFDCLYFDG -52-	EGLVL
ChV ATP	KIDGIR -29-	EGSDGEIS -24-	SYYWFDYVTDDP -55-	EGVMI
T7 ATP	KYDGVR -48-	FMLDGELM -49-	HIKLYAILPLHI -60-	EGLIV
Eco NAD	KLDGLA -47-	LEVRGEVF -45-	TFFCYGVGVLEG -52-	DGVVI
Bst NAD	KIDGLA -44-	LEARGEAF -45-	DLFVYGLADAEA -51-	DGIVI
Tth NAD	KVDGLS -45-	LEVRGEVY -47-	TFYALGLGLEEV -53-	DGVVV

FIG. 9. **Nucleotidyl transferase motifs in the *Nae*I protein sequence.** Five collinear sequence elements, designated nucleotidyl transferase motifs I, III, IIIa, IV, and V, are conserved in ATP-dependent DNA ligases (*ATP*) and NAD-dependent DNA ligases (*NAD*) (20–24, 29). The amino acid sequences shown aligned with *Nae*I-43K (*top*, *bold*) include: the ATP-dependent ligases human ligase I (*Hu 1*), human ligase 3 (*Hu 3*), *Schizosaccharomyces pombe* (*Spo*), *Saccharomyces cerevisiae* (*Sce*), vaccinia virus (*Vac*), *Chlorella* virus (*ChV*), and bacteriophage T7 (*T7*); and the NAD-dependent ligases *Escherichia coli* (*Eco*), *Bacillus stearothermophilus* (*Bst*) and *Thermus thermophilus* (*Tth*). The number of amino acids separating the motifs is indicated by (-n-). Amino acids in the ligase motifs that are in the same exchange group (30) as the corresponding *Nae*I amino acid are shown in *bold* and *underlined*. *Nae*I was compared with the DNA ligases by visual inspection of amino acid sequences keeping in mind the secondary structures of the individual ligase motifs (21–24).

DISCUSSION

NaeI Position 43—Substituting lysine for leucine at position 43 abolishes *Nae*I endonuclease activity and replaces it with topoisomerase and recombinase activities (10). Alanine, arginine, asparagine, glutamate, and histidine were substituted at position 43 and lysine at position 40 and the activities of the respective N-terminal MBP fusion proteins quantitated to test the chemical characteristics that lead to topoisomerase activity. All substitutions, except arginine and lysine substituted at

position 43, retained sequence-specific endonuclease activity, albeit at lower levels (Table I). Substitution of alanine at position 43 retained the most cleavage activity, whereas inserting lysine at position 40 and the negatively charged glutamate at position 43 reduced cleavage ~33- and 100-fold, respectively. Alanine and histidine substitutions at position 43 retained wild-type levels of DNA binding. Substituting lysine for leucine at position 40 and glutamate for leucine at position 43, however, reduced DNA binding 125-fold and 23-fold, respectively. Deletion of the entire Endo domain only reduces DNA binding for *Nae*I 8-fold (27) because of the strong DNA binding of the Topo domain (6, 14, 27). Thus, placing a charge within the core hydrophobic region of the Endo domain appears to have effects that reach beyond the local environment of the Endo domain.

*Nae*I-43K lacks activity when expressed either as an N-terminal or as a C-terminal MBP fusion protein (not shown). *Nae*I-43R fused at its N terminus to MBP also lacks activity. The effect of fusion at the C terminus was not determined. When expressed without MBP, on the other hand, both mutant proteins gave identical topoisomerase activities. *Nae*I has endonuclease activity either when expressed alone or as an N-terminal MBP fusion protein (27). No activity is recovered from *Nae*I expressed as a C-terminal fusion protein. The apparent loss of activity when fused at the C terminus can be rationalized from the crystal structure of the *Nae*I-DNA complex (14): The C terminus of each *Nae*I monomer lies over the DNA binding pocket. A bulky protein fused in this position would block access of DNA to the *Nae*I Topo domain. On the other hand, the *Nae*I N terminus lies on the outside of *Nae*I away from the binding faces of the protein (14). Thus, the loss of activity when MBP is fused to the N terminus of *Nae*I-43K and -43R, but not when fused to the N terminus of *Nae*I, suggests an altered structure for *Nae*I-43K and -43R.

NaeI-43R Has Topoisomerase Activity—*Nae*I-43R and *Nae*I-43K were found to be similar with respect to specific activity, NaCl dependence, amsacrine inhibition, and ssDNA inhibition. Amsacrine and ssDNA have no affect on *Nae*I, but inhibited the topoisomerase activities of *Nae*I-43K and *Nae*I-43R, demonstrating the equivalency of the L43K and L43R substitutions. To probe whether introducing a positive charge into the *Nae*I hydrophobic core could explain the switch to topoisomerase activity, lysine was substituted for leucine at position 40 in *Nae*I. Leucine 40 lies in the same α-helix, H2, as leucine 43 and is also part of the hydrophobic core of *Nae*I. Substitution of lysine for Leu-40 of *Nae*I decreased *Nae*I endonuclease activity, relative to wild type, but did not produce topoisomerase activity (Table I). The decrease in binding caused by L40K was similar to that seen when a negative charge (on glutamate) was inserted at position 43. The result implies that the topoisomerase activities in *Nae*I-43K and *Nae*I-43R depend on chemistry that involves more than the disruption caused by the introduction of a positive charge into the hydrophobic core. It is also possible, however, that a positive charge placed at position 40 causes a different disruption to folding than when placed at position 43.

Positive Charge at Position 43—The addition of a positive charge in the form of lysine or arginine is sufficient to convert *Nae*I from a restriction endonuclease to a topoisomerase. *Nae*I position 43 is located near the interface between the two subunits that compose the homodimer (Fig. 1). The addition of a positive charge could interfere with *Nae*I dimerization, but size-exclusion chromatography demonstrated that *Nae*I-43K is a dimer. Limited proteolysis of *Nae*I shows two distinct and stable domains as products (Ref. 27 and Fig. 7). Limited proteolysis of *Nae*I-43K showed a very different initial pattern of digestion products (Fig. 7), indicating that in *Nae*I-43K *versus* *Nae*I, different arginines and lysines are accessible to trypsin. It is clear from the proteolysis results that *Nae*I-43K is less stable to trypsin than *Nae*I. At 90 min, a large portion of full-length *Nae*I remained undigested, whereas the full-length *Nae*I-43K was almost completely digested in 60 min and by 90 min showed no domain remaining resistant to trypsin. The most resistant band from digestion of *Nae*I-43K with trypsin was at 17 kDa. Peptide sequencing showed that this product and the 17-kDa product of *Nae*I (27) come from the Endo domain. Thus, the L43K substitution, which lies in the Endo domain, has significant effect on trypsin digestion of both the Topo and Endo domains. This implies significant refolding of the protein.

The protein UV-CD spectrum is determined by the protein conformation (28). The UV-CD spectra of *Nae*I and *Nae*I-43K indicate significant structural differences (Fig. 8*A*). The CD spectrum for *Nae*I-43K has the classical minima at 208 and 222 nm indicative of significant α-helical structure (31), whereas in *Nae*I they are not as prominent. The CD spectra were analyzed using the computer program Continll (28). *Nae*I-43K showed higher α-helical content than wild-type *Nae*I: 40 *versus* 15% α-helix, 33 *versus* 42% β-sheet, 17 *versus* 24% turn, and 10 *versus* 20% unordered, respectively. The errors associated with these values are all better than $\pm$ 1% as determined from the best-fit results. CD analysis of protein structure is not very accurate. Here, the CD results underestimate the amount of *Nae*I α-helix; the crystal structure of *Nae*I shows that ~30% of the protein is α-helical. We assume that the α-helical content of *Nae*I-43K was underestimated to the same extent. Thus, by CD analysis, *Nae*I-43K has an altered structure relative to wild-type *Nae*I.

The presence of the positive charges in *Nae*I-43K and *Nae*I-43R could destabilize the proteins. To measure relative stability, CD measurements at λ of 208 nm as a function of increasing temperature were used to monitor *Nae*I and *Nae*I-43K temperature denaturation (T_m) profiles (Fig. 8*B*). *Nae*I-43K was more heat-stable. The 3 °C increase in T_m value for *Nae*I-43K relative to *Nae*I was surprising considering the location of position 43 in the hydrophobic core. L43K substitution causes a significant change in *Nae*I structure. The change may move lysine 43 out of the hydrophobic core so that it can participate in electrostatic interactions that help stabilize *Nae*I-43K. We speculate that the change in structure is required for the switch from endonuclease to topoisomerase activity.

Changes in Structure and Activity—It is reasonable to assume that a significant refolding of *Nae*I is necessary to convert its activity from endonuclease to that of topoisomerase and recombinase. If this is true, then the significant global conformational changes demonstrated here may be the result of this refolding. How the changes in protein folding lead to the changes in protein activity is unknown. We can speculate about a possible mechanism, however, based on sequence homologies found by visual comparisons of the *Nae*I protein sequence with the sequences of the DNA ligases (Fig. 9). There are five conserved nucleotidyl transferase motifs that define the ligase/capping enzyme superfamily (20–24, 29). The five collinear sequence elements, designated nucleotidyl transferase motifs I, III, IIIa, IV, and V, are conserved in ATP-dependent DNA ligases, mRNA-capping enzymes, and NAD-dependent DNA ligases. Four of five sequence motifs were detected in *Nae*I by visual inspection (Fig. 9). The *Nae*I sequences are collinear with the conserved nucleotidyl transferase motifs, and the sequence distances between the *Nae*I sequences are approximately within the ranges predicted from that between the corresponding nucleotidyl transferase motifs. The motifs are spread over both domains of *Nae*I, and the *Nae*I structure cannot be superimposed on the DNA ligase structure. It is

interesting to speculate that the refolding of *Nae*I caused by the placement of a positive charge at position 43 causes at least some of these motifs to superimpose. Active site residues in *Nae*I bind divalent metal, which stabilizes the DNA binding and the pentavalent transition state. DNA ligase motif III contains two conserved residues: one is conserved among all the DNA ligases, the other is conserved among the ATP ligases (Fig. 9). The two conserved amino acids in motif III contribute to the third (nick-joining) step of DNA ligation. For example, in Chlorella virus DNA ligase, Asp-65 and Glu-67 enhance the rate of step 3 phosphodiester formation by 20- and 1000-fold, respectively (32). Asp-65 overlaps with *Nae*I Asp-86, an essential metal ion-binding amino acid in *Nae*I (Fig. 9), which enhances DNA cleavage by 50-fold.[2] Mutations of Glu-88 in *Nae*I have not been studied because this position is not conserved among restriction enzymes. Study of the effects on topoisomerase activity of mutation of *Nae*I-43K amino acids with similarity to the ligase motifs shown in Fig. 9 are in progress. The presence of additional ligase motifs in *Nae*I would imply homology with the DNA ligase family and support the notion that *Nae*I topoisomerase activity arises through linkage of DNA cleavage and ligation activities through the *Nae*I-DNA covalent intermediate (10).

Comparison with Nucleotidyl Transferases—The nucleotidyl transferase K*X*DG-like motif in *NaeI*-43K appears to provide another example of the motif retaining some activity after substitution of the active site K*X*DG lysine with arginine but not when substituted with other amino acids. Two nucleotidyl transferases that manifest similar behavior are vaccinia virus RNA-capping enzyme and T4 RNA ligase (33, 34). Active-site lysine to arginine substitution in vaccinia virus RNA-capping enzyme gives low overall activity, whereas other mutations give no activity (33). The equivalent substitution in T4 RNA ligase gives intermediate levels of activity in all three steps of nucleotidyl transfer (34), whereas asparagine substitution gives no activity in any of the three steps (34).

*Nae*I-43K and *Nae*I-43R can catalyze many cycles of DNA relaxation in the absence of ATP and NAD^+ (10). Therefore, they do not require adenylation to relax DNA. Rather they form a covalent enzyme-DNA complex (10). The energy from the covalent complex is used for ligation, replacing the need for an enzyme-adenylate intermediate. This suggests that *Nae*I-43K and -43R require only the third step of the nucleotidyl transferase reaction, strand closure, for topoisomerase activity. Studies of vaccinia virus RNA-capping enzyme (33), T4 RNA ligase (34), and Chlorella virus DNA ligase (35) show that the K*X*DG lysine contributes to strand closure. In T4 RNA ligase, substitution of the active-site lysine with asparagine inactivates the strand-closure step (34). Lysine is not absolutely required for this step by RNA-capping enzyme and DNA ligase but contributes 16-fold to the rate of closure (35). Thus, the importance of the active site lysine for strand closure may vary with the protein being considered. In *Nae*I, arginine and lysine substitutions at position 43 contribute a positive charge that alters the structure of *Nae*I and gives it the ability to catalyze strand closure of newly cleaved substrate. The result of this cleavage and ligation, when linked by the enzyme-DNA covalent intermediate, is relaxation of supercoiled DNA.

[2] K. Carrick and M. Topal, unpublished results.

REFERENCES

1. Oller, A. R., Vanden Broek, W., Conrad, M., and Topal, M. D. (1991) *Biochemistry* **30,** 2543–2549
2. Reuter, M., Kupper, D., Pein, C.-D., Petrusyte, M., Siksnys, V., Frey, B., and Krüger, D. (1993) *Anal. Biochem.* **209,** 232–237
3. Yang, C. C., Baxter, B. K., and Topal, M. D. (1994) *Biochemistry* **33,** 14918–14925
4. Conrad, M., and Topal, M. D. (1989) *Proc. Nat. Acad. Sci. U. S. A.* **86,** 9707–9711
5. Topal, M. D., Thresher, R. J., Conrad, M., and Griffith, J. (1991) *Biochemistry* **30,** 2006–2010
6. Yang, C. C., and Topal, M. D. (1992) *Biochemistry* **31,** 9657–9664
7. Baxter, B. K., and Topal, M. D. (1993) *Biochemistry* **32,** 8291–8298
8. Holtz, J. K., and Topal, M. D. (1994) *J. Biol. Chem.* **269,** 27286–27290
9. 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., and Guthrie, E. P. (1995) *Gene* (*Amst.*) **155,** 19–25
10. Jo, K., and Topal, M. D. (1995) *Science* **267,** 1817–1820
11. Jo, K., and Topal, M. D. (1996) *Biochemistry* **35,** 10014–10018
12. Jo, K., and Topal, M. D. (1996) *Nucleic Acids Res.* **24,** 4171–4175
13. Huai, Q., Colandene, J. D., Chen, Y., Luo, F., Zhao, Y., Topal, M. D., and Ke, H. (2000) *EMBO J.* **19,** 3110–3118
14. Huai, Q., Colandene, J. D., Topal, M. D., and Ke, H. (2001) *Nat. Struct. Biol.* **8,** 665–669
15. Ban, C., and Yang, W. (1998) *EMBO J.* **17,** 1526–1534
16. Tsutakawa, S. E., Muto, T., Kawate, T., Jingami, H., Kunishima, N., Ariyoshi, M., Kohda, D., Nakagawa, M., and Morikawa, K. (1999) *Mol. Cell* **3,** 621–628
17. Kovall, R. A., and Matthews, B. W. (1999) *Curr. Opin. Chem. Biol.* **3,** 578–583
18. Hickman, A. B., Li, Y., Mathew, S. V., May, E. W., Craig, N. L., and Dyda, F. (2000) *Mol. Cell* **5,** 1025–1034
19. Berger, J. M., Fass, D., Wang, J. C., and Harrison, S. C. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 7876–7881
20. Shuman, S., and Schwer, B. (1995) *Mol. Microbiol.* **17,** 405–410
21. Subramanya, H. S., Doherty, A. J., Ashford, S. R., and Wigley, D. B. (1996) *Cell* **85,** 607–615
22. Singleton, M. R., Håkansson, K., Timson, D. J., and Wigley, D. B. (1999) *Structure* **7,** 35–42
23. Lee, J. Y., Chang, C., Song, H. K., Moon, J., Yang, J. K., Kim, H.-K., Kwon, S.-T., and Suh, S. W. (2000) *EMBO J.* **19,** 1119–1129
24. Odell, M., Sriskanda, V., Shuman, S., and Nikolov, D. B. (2000) *Mol. Cell* **6,** 1183–1193
25. Clackson, T., Gussow, D., and Jones, P. T. (1991) in *PCR-A Practical Approach* (Quirke, P., and Taylor, G. R., eds) pp. 187–214, Oxford University Press, New York
26. Jo, K., and Topal, M. D. (1998) *Nucleic Acids Res.* **26,** 2380–2384
27. Colandene, J. D., and Topal, M. D. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 3531–3536
28. Provencher, S. W., and Glockner, J. (1981) *Biochemistry* **20,** 33–37
29. Sriskanda, V. Schwer, B., Ho, C. K., and Shuman, S. (1999) *Nucleic Acids Res.* **27,** 3953–3963
30. Dayhoff, M. O., Schwartz, R. M., and Orcott, B. L. (1978) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., ed) Vol. 5, Suppl. 3, pp. 345–352, National Biomedical Research Foundation, Washington, D. C.
31. Cantor, C. R., and Schimmel, P. R. (1980) *Biophysical Chemistry, Part II: Techniques for the study of biological structure and function*, pp. 425–428, W. H. Freeman and Co., San Francisco, CA
32. Sriskanda, V., and Shuman, S. (2002) *Nucleic Acids Res.* **30,** 903–911
33. Cong, P., and Shuman, S. J. (1993) *J. Biol. Chem.* **268,** 7256–7260
34. Heaphy, S., Singh, M., and Gait, M. J. (1987) *Biochemistry* **26,** 1688–1696
35. Sriskanda, V., and Shuman, S. (1998) *Nucleic Acids Res.* **26,** 4618–4625