

Structure of *NaeI*-DNA Complex Reveals Dual-Mode DNA Recognition and Complete Dimer Rearrangement

Q. Huai,¹ J.D. Colandene,² M.D. Topal¹⁻⁴ and H. Ke^{1,2}

¹Department of Biochemistry and Biophysics (Email: hke@med.unc.edu)

²Lineberger Comprehensive Cancer Center,

³Department of Pathology (Email: mdtopal@med.unc.edu)

⁴Curriculum in Genetics and Molecular Biology, The Univ. of North Carolina, Chapel Hill, NC

For many biological processes, such as transcription and recombination, various proteins have to recognize specific sequences of DNA before interacting with them. *NaeI* is a restriction endonuclease that finds two -GCC-GGC- patterns in the DNA chain, cleaves these sequences in the middle, loops out the intervening DNA sequence, and forms a covalent bond with the cleaved substrate. To accomplish this, two domains of the protein, the so-called endo and topo domains, have to recognize and bind with this sequence in the DNA molecule. In this work, the structure of DNA bound to *NaeI* has been determined, and it presents the first example of novel recognition of two copies of one DNA sequence by two different amino acid sequences and two different structural motifs in one polypeptide.

NaeI, a novel DNA endonuclease, shows topoisomerase and recombinase activities when lysine 43 is substituted for leucine. The *NaeI*-DNA structure demonstrates that each of the Endo and Topo domains of *NaeI* binds one molecule of DNA duplex. DNA recognition induces dramatic rearrangements: narrowing the Topo domain 16 Å to grip DNA, widening the Endo domain 8 Å to bend DNA 45° for cleavage, and completely rebuilding the homodimer interface. The *NaeI*-DNA structure presents the first example of novel recognition of two copies of one DNA sequence by two different amino acid sequences and two different structural motifs in one polypeptide.

Protein-DNA recognition is central to biological processes from transcription to recombination. *NaeI*, a type IIe restriction endonuclease, cleaves the DNA sequence GCC-GGC and forms a covalent bond with the cleaved substrate^{1,2}. *NaeI* must bind two DNA recognition sequences to cleave one DNA sequence^{3,4} and demonstrates DNA-concentration-dependent catalysis with sigmoidal kinetics⁴. Changing Leu43 to lysine converts *NaeI* from endonuclease to topoisomerase/recombinase, suggesting that *NaeI* is a bridge between these protein families². To find out whether two DNA binding motifs recognize the same sequence of DNA, the crystal structure of *NaeI* complexed with a 17 basepair DNA substrate has been determined at 2.5 Å resolution.

Methods

NaeI (15 mg/ml) and the 17-basepair DNA substrate (TGC CAC GCC GGC GTG GC) were mixed (1:1.3) and co-crystallized in the space group P2₁2₁2₁ with a = 59.6, b = 65.0, c = 258.5 Å. The diffraction data were collected on beamline X12C of the National Synchrotron Light Source at Brookhaven National Laboratory. A total of 120725 measurements were reduced to 32365 unique reflections to 2.4 Å resolution, with R-merge of 0.098. The *NaeI*-17mer structure was refined by CNS⁵ to R-free of 0.293 at 2.4 Å resolution.

Two DNA duplexes per *NaeI* dimer

The structure of *NaeI* complexed with 17 base pair cognate DNA fragment (17mer) is a homodimer (Fig. 1). *NaeI* monomer contains two structural domains: an

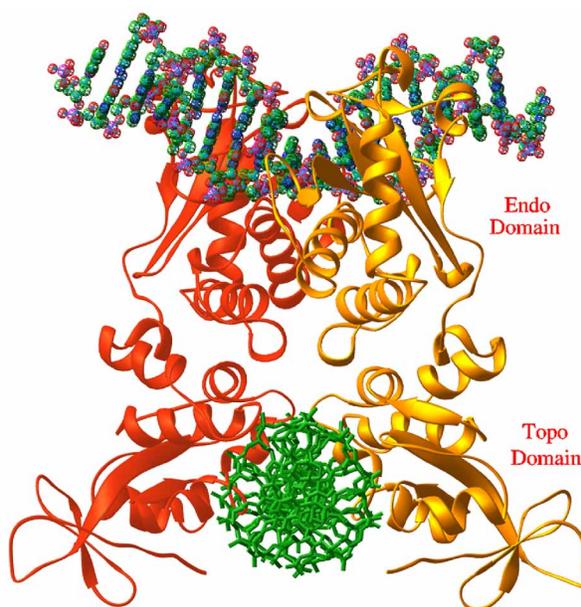


Figure 1. Ribbon presentation of dimeric *NaeI* (red and gold) bound to DNA (ball-stick).

N-terminal Endo domain (residues 1-162) and a C-terminal Topo domain (residues 172 to 313), linked by a relatively extended hinge loop (residues 163-171). The Endo domains of the *NaeI* dimer are jointly formed by two monomers, binds a double stranded DNA, and contains catalytic residues Glu70, Asp86, Asp95, and Lys97 that are conserved among type II endonucleases⁶⁻⁸. The Topo domains of the *NaeI* dimer contain the DNA binding motif of catabolite activator protein (CAP), observed in many DNA processing proteins⁹⁻¹¹, including DNA topoisomerases IA and II¹². The Topo domain cleft binds a second double-stranded DNA.

Different DNA recognition in the Endo and Topo domains

The conformation of the 17mer DNA fragment is different when bound at the Endo versus Topo domains (Fig. 2). The DNA bound at the Endo domain active site is not cleaved but is bent about 45° around the central base pairs C9-G10. Its conformation is B-DNA but the pitch and basepairs per helical turn are increased to about 35 and 11 Å, respectively. The 17mer DNA bound in the Topo domain is neither cleaved nor bent and has B-DNA conformation with 10.3 base pairs and about 34 Å pitch per turn of helix. Surprisingly, the Endo and Topo domains of *NaeI* recognize GCCGGC using different amino acid compositions and different protein folds. The Endo domain recognizes GCCGGC

primarily through hydrogen bonds between b-turn Asn144-Lys 148 and the DNA bases in the major groove (Asn144-G7, Asp146-C8, Lys148-G10, and Lys 148-G11) and between a-helix H4 and the bases in the minor groove (Thr63-G10 and Lys59-C12). In the cleft formed by the two Topo domains, on the other hand, two CAP motifs, one from each monomer, sit next to each other along the major groove of the 17mer DNA (Fig. 2). Most amino acids involved in recognition differ between the two domains, except for Asp146 and Lys148 in the Endo domain and Asp 226 versus Lys 229 in the Topo domain, which recognize C8 and G11 of the DNA substrate, respectively.

Complete dimer rearrangement upon DNA binding

Structural comparison between unligated *NaeI* and the DNA complex reveals dramatic rearrangements of the *NaeI* dimer upon DNA binding. Superimposing monomer A of unligated *NaeI* with the corresponding monomer of *NaeI*-17mer shows that the Topo domain in monomer B of *NaeI*-17mer is rotated about 6° clockwise around the molecular 2-fold axis and translated about 16 Å closer to the other Topo domain (Fig. 3). This change brings the CAP recognition helices in contact with the DNA major groove. In contrast, the Endo domain of *NaeI*-17mer is rotated about 11° counterclockwise around the molecular 2-fold axis and trans-

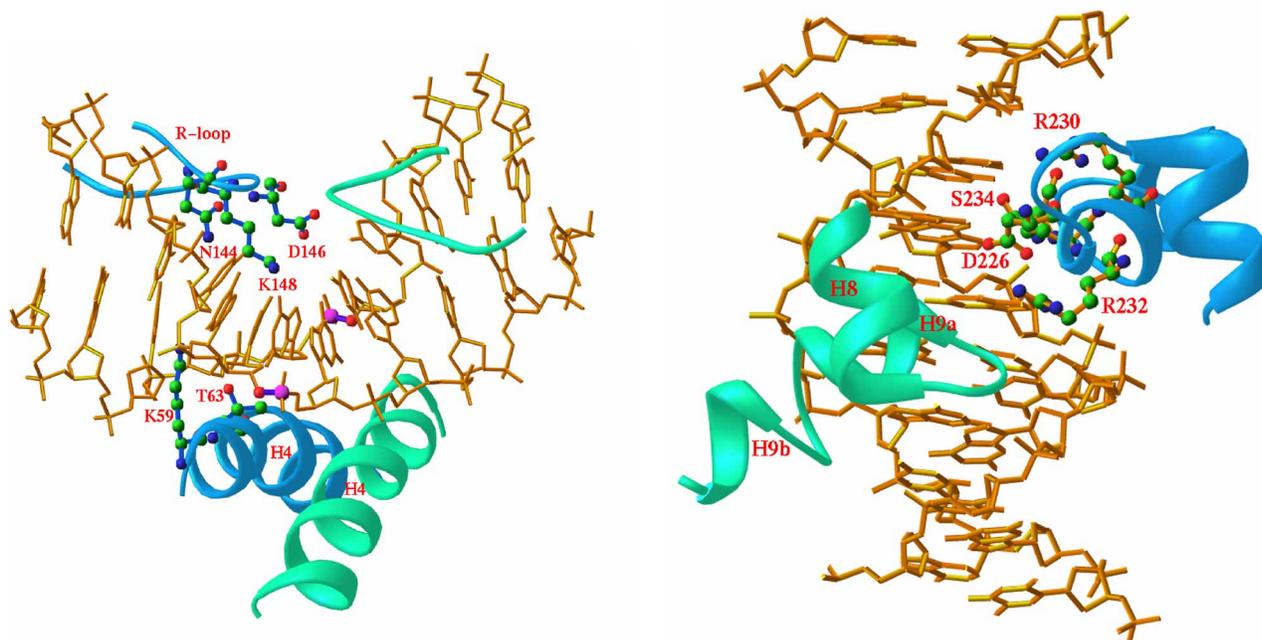


Figure 2. *left*, DNA binding to the Endo domain. Central 10 basepairs of the 17mer DNA are shown (gold sticks: ACGCC GGCGT, scissile bonds are ball and stick). H4 helices are shown as green and cyan ribbons. Asn144, Asp146 and Lys148 (ball-stick, shown in one monomer) bind in the major groove. Lys59 and Thr63 bind in the minor groove. *right*, DNA binding to the helix-turn-helix/CAP motifs of the Topo domains. H9a is a 3_{10} helix. The central 10 pairs of the DNA substrate are shown (gold sticks). The recognition residues Asp226, Lys229, Arg230, Arg232 and Ser234 are represented by ball-stick models.

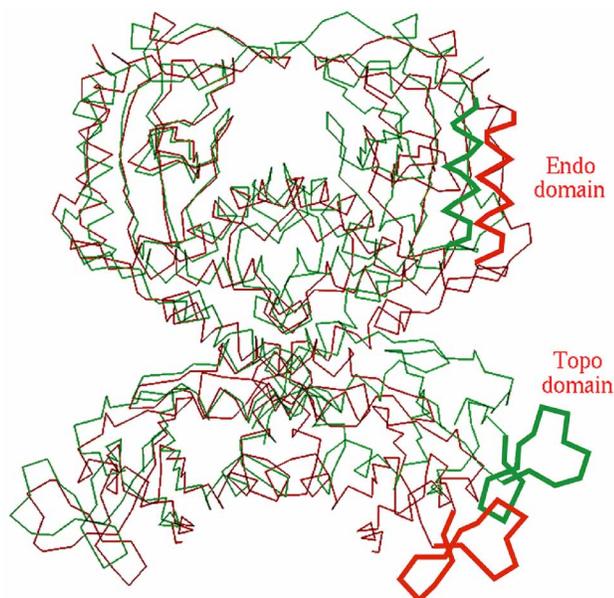


Figure 3. Superposition of the C α atoms of the structures of unligated (green) and DNA bound NaeI (red). The monomers on the left were best superimposed and the matrix from the superposition was applied to the monomers on the right.

lated about 8 Å away from the other Endo domain, thus widening the catalytic cleft for DNA binding. In addition, helix H4 (Lys59 to Phe77), an essential element for NaeI dimer formation, undergoes two significant changes. H4 is in two pieces in unligated NaeI, H4a and H4b with a non-helical backbone conformation of

His64 ($f = -113^\circ$ and $j = 17^\circ$). In the NaeI-17mer complex, however, it is a single united helix (His64: $f = -56^\circ$ and $j = -24^\circ$). The Ca atoms of Lys59, Thr60, and Glu61 of H4 migrated 2.5-3 Å, about three times the average deviation of all Ca atoms between the two structures. Concomitant with uniting H4, all ten hydrogen bonds at the interface of the unligated NaeI dimer are abolished in the NaeI-DNA structure and replaced with five new hydrogen bonds (Fig. 4). A majority of van der Waals interactions between the two NaeI monomers also undergo rearrangement. Finally, the homodimer of the NaeI-17mer complex is symmetric, in contrast to the asymmetric domain arrangement in the unligated NaeI dimer with a 3° rotational offset and a 10 Å translational shift⁶.

The dramatic rearrangement of the NaeI dimeric structure on DNA binding enables us to speculate about the allosteric cleavage of DNA by NaeI. The substrate-binding cleft of unligated NaeI, containing the active sites of the Endo domains presumably in their resting state, is not wide enough to accommodate DNA substrate. DNA binding to the CAP motifs in the Topo domains not only narrows the effector-binding cleft, but also opens the active site in the substrate-binding cleft for the second DNA binding. This is most likely accomplished through communication between domains using a network of amino acids to mediate information transfer. The complexity of the structural changes responsible for the allosteric response is underscored by the complete rearrangement of the NaeI dimeric interface followed by reshaping of the active site and bending of the DNA recognition sequence, thus facilitating the DNA cleavage in the Endo domains.

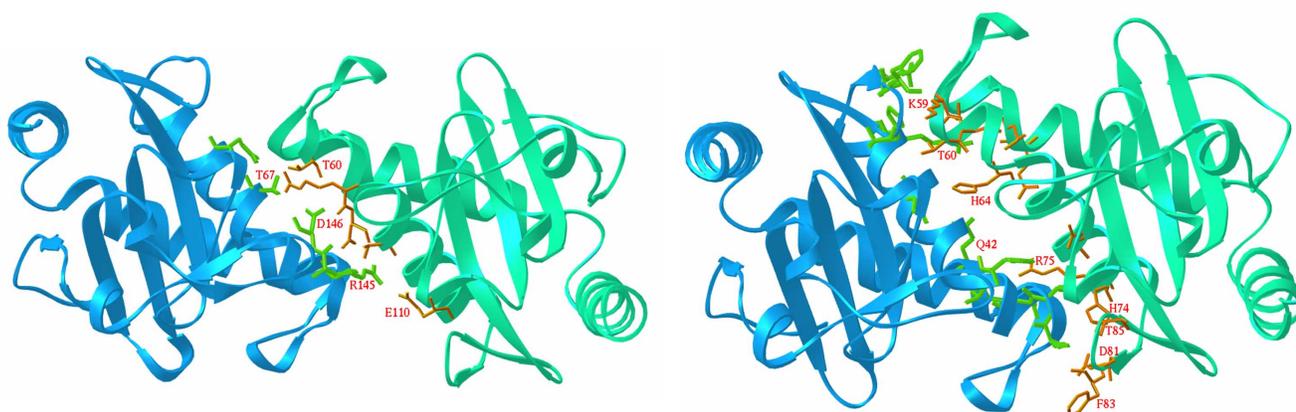


Figure 4. Dimer interface of unligated (right) and DNA-bound NaeI (left).

Conclusion

The *NaeI*-DNA structure demonstrates unusual characteristics for a monofunctional DNA processing enzyme. *NaeI* shows unique DNA recognition: two different structural folds and two different sets of amino acids bind two copies of the same DNA sequence. It also shows organized communication between two DNA binding domains via dramatic allosteric conformation changes. The structural features of *NaeI*-DNA mimic the hallmark of two-site DNA binding and dramatic domain rearrangements of the topoisomerases¹³ and recombinases¹⁴, providing further evidence that *NaeI* bridges DNA enzyme families.

Acknowledgements

We thank Dr. R. Sweet for help with data collection at beamline X12C at the National Synchrotron Light Source. This research is supported by NIH grant GM52123. Research carried out at the National Synchrotron Light Source, Brookhaven National Laboratory, which is supported by the U.S. Department of Energy, Division of Materials Sciences and Division of Chemical Sciences, under Contract No. DE-AC02-98CH10886.

References

1. Holtz, J.K. & Topal, M.D. *J. Biol. Chem.* **269**, 27286-27290 (1994).
2. Jo, K. & Topal, M.D. *Science* **267**, 1817-1820 (1995).
3. Conrad, M. & Topal, M.D. *Proc. Natl. Acad. Sci. USA* **86**, 9707-9711 (1989).
4. Yang, C.C. & Topal, M.D. *Biochemistry* **31**, 9657-9664 (1992).
5. Brünger, A.T. *et al. Acta Cryst.* **D54**, 905-921 (1998).
6. Huai, Q. *et al. EMBO J.* **19**, 3110-3118 (2000).
7. Colandene, J.D. & Topal, M.D. *Biochemistry* **39**, 13703-13707 (2000).
8. Kovall, R.A. & Matthews, B.W. *Curr. Opin. Chem. Biol.* **3**, 578-583 (1999).
9. Harrison, S.C. *Nature*, **353**, 715-719 (1991).
10. Pabo, C.O. & Sauer, R.T. *Annu. Rev. Biochem.* **61**, 1053-1095 (1992).
11. Nelson, H.C.M. *Curr. Opin. Genet. Dev.* **5**, 180-189 (1995).
12. Berger, J.M., Fass, D., Wang, J.C., & Harrison, S.C. *Proc. Natl. Acad. Sci. USA* **95**, 7876-7881 (1998).
13. Keck, J.L. & Berger, J.M. *Nature Struct. Biol.* **6**, 900-902 (1999).
14. Gopaul, D.N. & Van Duyne, G.D. *Curr. Opin. Struct. Biol.* **9**, 14-20 (1999).