A Novel Zinc Finger-Based DNA Cutter: Biosynthetic Design and Highly Selective DNA Cleavage

Makoto Nagaoka, Masaki Hagihara, Jun Kuwahara,† and Yukio Sugiura*

Institute for Chemical Research
Kyoto University, Uji, Kyoto 611, Japan

Received November 2, 1993

Conversion of a DNA-binding protein to a DNA-cleaving molecule by attachment of a metal-chelating ligand is one of the most versatile methods for affinity cleaving. These chimeric proteins have largely utilized helix–turn–helix[1-6] or b-zip[7]-type motifs and interacted with DNA as a dimer. Therefore, their target sites are limited to palindromic base sequences with a dyad or a pseudodyad axis. On the other hand, DNA sequences recognized by Cys_2His_2-type zinc finger proteins are almost asymmetric because of their monomeric binding mode. We report here the design and function of a new DNA-cleaving metalloprotein consisting of the zinc finger motif.

Primary sequence of the zinc finger-based DNA cutter (designated Sp1GGH) comprising two functional domains is shown in Figure 1. The DNA-binding domain contains the C-terminal region (residues 529–696) of transcription factor Sp1 bearing three contiguous repeats of the Cys_2His_2-type zinc finger motif, which recognizes an asymmetric decanucleotide with consensus sequence 5′-(G/T)GGGCGG(G/A)(G/A)(C/T)-3′.[8,9] Each zinc finger domain coordinates a Zn(II) in a tetrahedral complex. In an effort to give DNA-cleaving activity to zinc finger protein, the tripeptide Gly-Gly-His (GGH) was attached to the N-terminus of the DNA-binding domain because of the availability of a genetic engineering method.[10] The GGH segment originally derived from the copper-binding domain of serum albumin is believed to bind Ni(II) in a 1:1 square-planar complex with coordination from an imidazole nitrogen, two deprotonated peptide nitrogens, and the terminal amino group, referring to the crystal structure of the Cu^{II}–GGH complex.[11] The Ni(II) complex

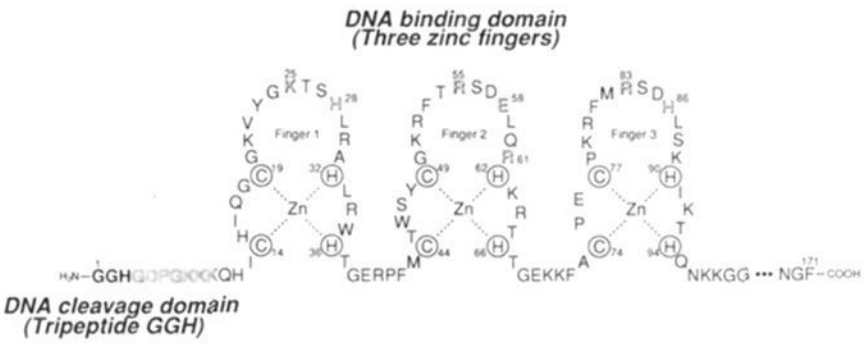

Figure 1. Schematic representation of Sp1GGH. Amino acid residues of Sp1GGH are indicated by one-letter abbreviations and numbered from the amino terminus. The tripeptide GGH, the linker region, and the amino acids relevant to specific base contact[16,22] are represented as bold, outlined, and shadowed types, respectively.

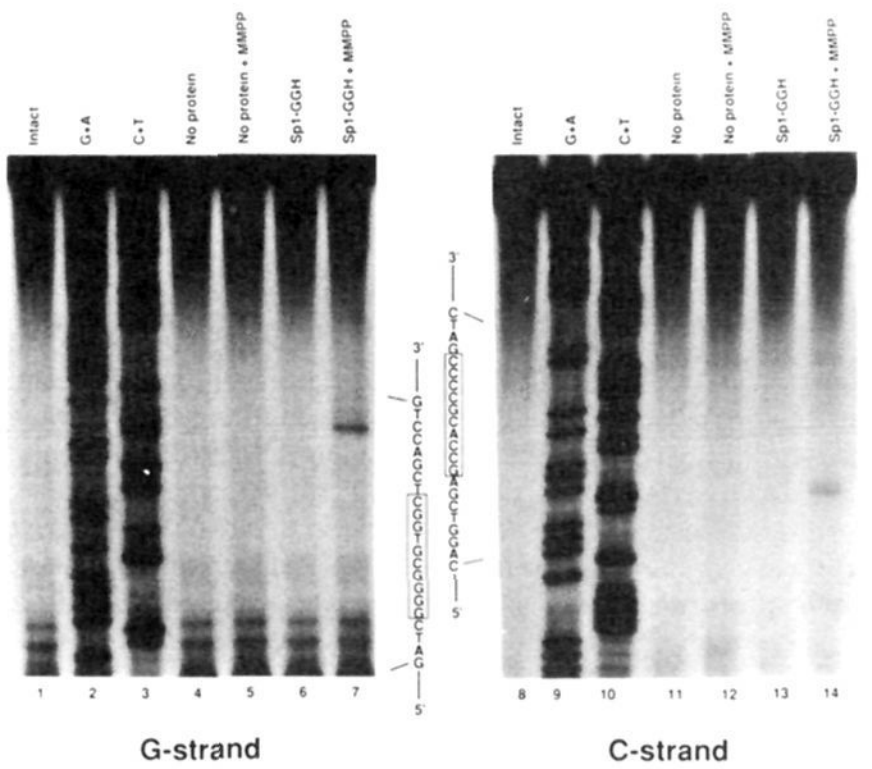

Figure 2. Autoradiograms of 10% denaturing gel for the Ni-mediated Sp1GGH cleavage of a 5′-^{32}P-labeled restriction fragment containing single GC box DNA. Reaction samples contained 7.5 mM Tris-HCl (pH 7.5), 35 mM NaCl, 0.4 μg of sonicated calf thymus DNA (av 400 bp), and ~30 000 cpm end-labeled restriction fragment (*EcoRI*/*HindIII*) from pIBI7T[17] in a total volume of 20 μL. Sp1GGH was dialyzed in the presence of $ZnCl_2$ and $Ni(NO_3)_2$. After preincubation for the equilibration of Sp1GGH with DNA at 20 °C for 30 min, the cleavage reactions were initiated by the addition of magnesium monoperoxyphthalate (100 μM), run at 20 °C for 30 min, and then terminated by ethanol precipitation. Cleavage products were analyzed on a 10% polyacrylamide/7 M urea sequencing gel. Lanes 1–7 and 8–14 contain the guanine-rich strand (G-strand) and the cytosine-rich strand (C-strand), respectively. Lanes 1 and 8 are intact DNA lanes. Lanes 2 and 9 show the Maxam–Gilbert sequencing reactions for G + A, as do lanes 3 and 10 for C + T.[23] Lanes 4 and 11 involve $ZnCl_2$ (3.5 μM) and $Ni(NO_3)_2$ (3.5 μM). Lanes 5 and 12 contain $ZnCl_2$ (3.5 μM) and $Ni(NO_3)_2$ (3.5 μM), followed by the addition of magnesium monoperoxyphthalate. Lanes 6 and 13 involve $ZnCl_2$ (3.5 μM), $Ni(NO_3)_2$ (3.5 μM), and Sp1GGH (4 μM). Lanes 7 and 14 contain $ZnCl_2$ (3.5 μM), $Ni(NO_3)_2$ (3.5 μM), and Sp1GGH (4 μM), followed by the addition of magnesium monoperoxyphthalate.

of the peptide ligand attached to the DNA-binding protein can successfully cut DNA in the presence of peracid.[12,13] The amino terminus of the GGH segment must be a primary amine, although all bacterially expressed proteins have methionine residues on their N-termini. This problem can be overcome by incorporation of the recognition sequence IEGR of blood coagulation factor (factor Xa) preceding the GGH segment.[14] Repeated dialysis of

* Author to whom correspondence should be addressed.

† Present address: Faculty of Pharmaceutical Sciences, Tokushima University, Tokushima 770, Japan.

(1) Chen, C.-H. B.; Sigman, D. S. *Science* **1987**, *237*, 1197–1201.
(2) Bruice, T. W.; Wise, J.; Rosser, D. S. E.; Sigman, D. S. *J. Am. Chem. Soc.* **1991**, *113*, 5446–5447.
(3) Sluka, J. P.; Horvath, S. J.; Bruist, M. F.; Simon, M. I.; Dervan, P. B. *Science* **1987**, *238*, 1129–1132.
(4) Mack, D. P.; Iverson, B. L.; Dervan, P. B. *J. Am. Chem. Soc.* **1988**, *110*, 7572–7574.
(5) Mazzarelli, J. M.; Ermácora, M. R.; Fox, R. O.; Grindley, N. D. F. *Biochemistry* **1993**, *32*, 2979–2986.
(6) Ebright, R. H.; Ebright, Y. W.; Pendergrast, P. S.; Gunasekera, A. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 2882–2886.
(7) Oakley, M. G.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1992**, *31*, 10969–10975.
(8) Kadonaga, J. T.; Jones, K. A.; Tjian, R. *Trends Biochem. Sci.* **1986**, *11*, 20–23.
(9) Gidoni, D.; Dynan, W. S.; Tjian, R. *Nature* **1984**, *312*, 409–413.

(10) The Sp1GGH was prepared as described below. Two complementary synthetic oligonucleotides, 5′-CT AGC CCC GGG ATC GAA GGT CGT GGT GGA CAT GGT-3′ (35-mer) and 5′-GATC ACC ATG TCC ACC ACG ACC TTC GAT CCC GGG G-3′ (35-mer) were annealed and inserted into *Nhe*I/*Bam*HI site of the derivative of pJK4,[16] pJK42, where *Sma*I, *Cla*I, and *Hin*dIII sites were deleted. The pJK4GGH overexpresses the polypeptide JK4GGH. The mature form of Sp1GGH was obtained by proteolysis of JK4GGH with factor Xa. Digestion with factor Xa was carried out according to the method described by Nagai and Thøgersen.[14] The JK4GGH in 5 M urea/10 mM Tris-HCl (pH 7.5)/0.2 M NaCl was dialyzed against factor Xa buffer (50 mM Tris-HCl (pH 8.0)/0.1 M NaCl/1 mM $CaCl_2$/0.1 mM 2-mercaptoethanol) 10 times and then digested with factor Xa at an enzyme-to-substrate weight ratio of 1:100 at 25 °C for 1 h. The resulting Sp1GGH was purified by reverse-phase C_8 HPLC.

(11) Camerman, N.; Camerman, A.; Sarkar, B. *Can. J. Chem.* **1976**, *54*, 1309–1316.
(12) Mack, D. P.; Dervan, P. B. *J. Am. Chem. Soc.* **1990**, *112*, 4604–4606.
(13) Mack, D. P.; Dervan, P. B. *Biochemistry* **1992**, *31*, 9399–9405.
(14) Nagai, K.; Thøgersen, H. C. *Nature* **1984**, *309*, 810–812.

0002-7863/94/1516-4085$04.50/0

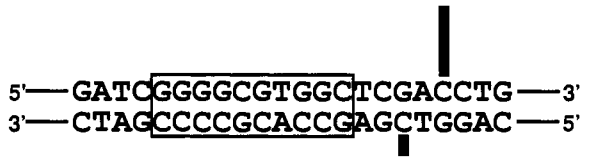

Figure 3. Histograms of cleavage sites in GC box DNA sequence based on densitometric analysis of the gel autoradiograms in Figure 2. Top and bottom sequences show G- and C-strands, respectively (Figure 2, lanes 7 and 14). The box indicates the GC box, and the length of bars represents the extent of cleavage. Relative extent of cleavage was estimated by subtracting the intensity of the bands in lanes 1 and 8 from that for the corresponding bands in lanes 7 and 14, respectively.

Sp1GGH against buffers containing different concentrations of metals enables selective binding of Zn(II) in the finger region and of Ni(II) in the GGH segment, respectively, reflecting inherent metal preference of two functional domains.[15] It has been confirmed that the three-finger region of Sp1 binds three Zn(II) ions.[16] Incorporation of the GGH segment onto the N-terminal region of Sp1 presumably does not affect the binding geometry of the three-finger domain, because extension of 12 amino acid residues from the N-terminus of Sp1(167*) has shown the exact same methylation interference pattern as that for the shorter form, Sp1(167*).[17]

In the presence of magnesium monoperoxyphthalate (100 μM), quite specific cleavage of DNA occurred predominantly at two cytosine bases on both strands with single-base specificity (Figure 2). Cutting at the cytosine on the guanine-rich strand (G-strand) was much stronger than that for the other cytosine base on the opposite cytosine-rich strand (C-strand) (Figure 3). Termini at the cleavage sites appear to be 3'-phosphate groups. The remarkably restricted range of cleavage at the single-base position suggests that a nondiffusible oxidant might be generated by the Ni–GGH complex in the minor groove. The facts show that orientation of the peptide backbone in Sp1 is antiparallel to the primarily interacting strand (G-strand). The cleavage center for the above two bases was 4 bp apart from the 3' end of GC box on the G-strand.

The 3'-staggered cutting pattern clearly demonstrates that the cleavage event occurs in the minor groove of DNA, and thus the Ni–GGH domain of Sp1GGH appears to be situated in the minor groove. In contrast, previous studies[9,18] have shown that the three-finger domain of Sp1 contacts with guanine bases in the major groove. The Ni–GGH domain attached to the N-terminal arm of Hin recombinase (139–190) also shows the 3'-staggered cleavage pattern.[12,13] Indeed, the homeodomains[19,20] and Hin recombinase[3,19] make base contacts in both grooves, with the helix–turn–helix region in the major groove of DNA and the N-terminal arm in the adjacent minor groove. Two functional domains of Sp1GGH are spanned by seven amino acid residues (GGHGDPGKKKQHIC). Given that the N-terminal linker region adopts an extended, flexible conformation in a manner similar to the conformation of the N-terminal arm of the homeodomains, the linker region can bridge across the sugar–phosphate backbone of DNA and permits the Ni–GGH domain to locate at the cleavage center in the minor groove. The finding that the N-terminal finger 1 of Sp1 makes loose contacts with DNA might reinforce our explanation.[18]

Tandemly repeated structures of multiple finger modules and asymmetry of the recognition base sequences are notable features of the Cys_2His_2-type zinc finger proteins. The modular structure revealed by the crystal structure of the three-finger protein Zif268-DNA complex,[21] in which the single finger domain strictly recognizes three base pairs, sheds light on the great versatility of the Cys_2His_2-type zinc finger proteins in terms of designing DNA-binding protein with novel specificities. It might be possible to recognize any desired sequence by combination of different finger motifs. Our work is pertinent to the design of the novel artificial restriction enzyme based on the zinc finger motif applicable to chromosome mapping and sequencing.

Acknowledgment. This work was supported in part by a Grant-in-Aid for Scientific Research to J.K. and a Grant-in-Aid for Scientific Research on Priority Area to Y.S. from the Ministry of Education, Science, and Culture, Japan.

(15) The Sp1GGH was dialyzed against a buffer (50 mM Tris-HCl (pH 7.5)/10 mM NaCl/5% glycerol/50 μM $ZnCl_2$) 7 times and further against a buffer (50 mM Tris-HCl (pH 7.5)/10 mM NaCl/5 μM $ZnCl_2$/5 μM $Ni(NO_3)_2$) 3 times.

(16) Kuwahara, J.; Coleman, J. E. *Biochemistry* **1990**, *29*, 8627–8631.
(17) Nagaoka, M.; Kuwahara, J.; Sugiura, Y. *Biochem. Biophys. Res. Commun.* **1993**, *194*, 1515–1520.
(18) Kuwahara, J.; Yonezawa, A.; Futamura, M.; Sugiura, Y. *Biochemistry* **1993**, *32*, 5994–6001.
(19) Affolter, M.; Percival-Smith, A.; Müller, M.; Billeter, M.; Qian, Y. Q.; Otting, G.; Wüthrich, K.; Gehring, W. J. *Cell* **1991**, *64*, 879–880.
(20) Wolberger, C.; Vershon, A. K.; Liu, B.; Johnson, A. D.; Pabo, C. O. *Cell* **1991**, *67*, 517–528.
(21) Pavletich, N. P.; Pabo, C. O. *Science* **1991**, *252*, 809–817.
(22) Kriwacki, R. W.; Schultz, S. C.; Steitz, T. A.; Caradonna, J. P. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 9759–9763.
(23) Maxam, A. M.; Gilbert, W. *Methods Enzymol.* **1980**, *65*, 499–560.