Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters 15 (2005) 2197–2201
Bioorganic & Medicinal Chemistry Letters

Effects of linking 15-zinc finger domains on DNA binding specificity and multiple DNA binding modes

Tsuyoshi Hirata, Wataru Nomura, Miki Imanishi and Yukio Sugiura*

Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan

Received 27 January 2005; revised 25 February 2005; accepted 10 March 2005

Abstract—To assess the possibility of multi-connection of zinc finger domains for understanding of DNA binding mechanisms and gene regulation, the longest artificial zinc finger protein, Sp1ZF15, has been constructed. This zinc finger consists of 5 units of Sp1 zinc finger peptide connected by canonical short linker sequences (TGEKP). Recognition of the 50 base pairs of DNA and potential binding to shorter targets by Sp1ZF15 were determined. Sequence alterations of the GCG triplet to ATA at a target site clearly showed that Sp1ZF15 changes its DNA binding mode depending on the target sequences. Of special interest is the fact that Sp1ZF15 controls the number of finger domains active in DNA binding corresponding to the length and sequence of the target DNA. These results suggest that artificial transcription factors based upon these multi-zinc finger proteins have great potential for the regulation of a vast number of cellular processes.

1. Introduction

Zinc finger protein is the key DNA binding module for transcription factor complexes.[1,2] The Cys_2His_2-type zinc finger protein is universal in eukaryotes, and hence its structural and functional studies have been developed.[3–7] The numbers of zinc finger domains in natural multi-zinc fingers vary up to 37.[8] Among the zinc finger proteins maintaining more than nine-zinc finger domains, some finger domains are separated by long amino acid sequences. [9] In such a form of connection, zinc finger domains play roles other than DNA binding. In typical zinc finger proteins, the canonical short linkers connect the zinc finger domains and assist DNA binding. These facts suggest that the canonical linker is critical for DNA binding of zinc finger proteins by wrapping around the DNA major groove.[10,11]

Recognition of a long DNA sequence is a rare function in the natural DNA binding proteins. Restriction enzymes recognize target DNA sequences in the 3–8 bp range.[12] Although there are several proteins such as homing endonucleases[13] and the Transcription Factor IIIA (TFIIIA) zinc finger protein from *Xenopus laevis*[14,15] whose recognition sequence is relatively long, they recognize some parts of the target sequences. Our artificial zinc finger protein, Sp1ZF9, indicates that the zinc finger modules effectively serve as a binding module for successively longer DNA sequences.[16]

We have designed and created the longest artificial zinc finger protein, Sp1ZF15, consisting of 5 units of Sp1 zinc finger connected by the canonical linker (TGEKP) (Fig. 1A). Our results indicate that increasing the number of zinc finger units is a feasible strategy to recognize a long DNA target. Moreover we found that multi-zinc finger proteins show various DNA binding modes depending on the target sequence. There is much interest in building sequence-specific binding proteins that bind to longer regions of DNA and the present results make a contribution in that direction.

1.1. Creation of Sp1ZF15

The coding DNA of Sp1ZF15 was inserted into a pET42b plasmid (Novagen) and the protein was overexpressed as a soluble form in an *Escherichia coli* strain BL21(DE3) at 20 °C. The first purification of the protein was performed by using a HisTrap purification kit (Amersham Biosciences). In the second step, a GSTrap purification column (Amersham Biosciences) was utilized. Purification procedures were followed as protocols

Keywords: Zinc finger protein; DNA binding; Sp1; Gene regulation; Linker sequence; 15-Finger protein.
* Corresponding author. Tel.: +81 774 38 3210; fax: +81 774 32 3038; e-mail: sugiura@scl.kyoto-u.ac.jp

0960-894X/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2005.03.040

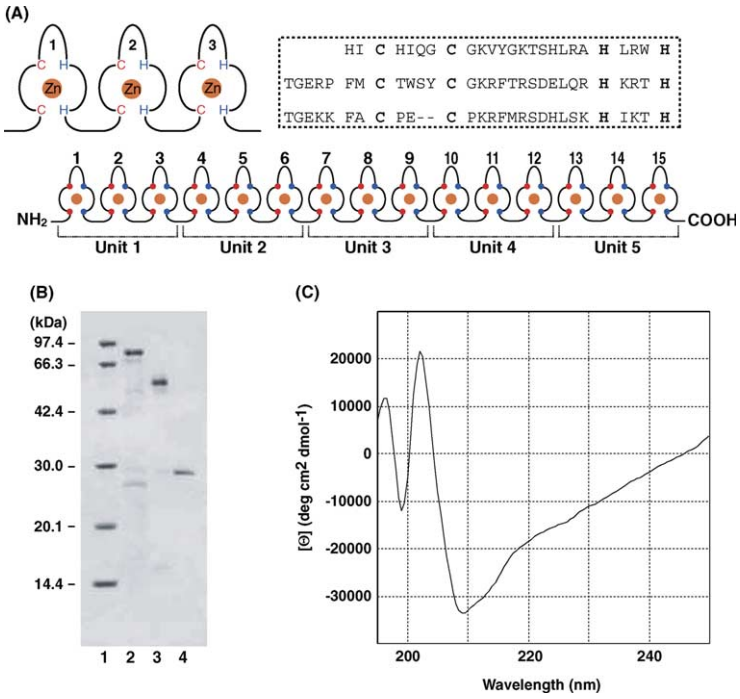

Figure 1. (A) A schematic representation of Sp1ZF15. The amino acid sequences are shown in the box of the figure. (B) Purified samples of Sp1ZF15 as GST fusion protein at lane 2, Sp1ZF15 cleaved by thrombin from the GST moiety at lane 3, and the GST moiety cleaved at lane 4. Protein marker is at lane 1 as the sizes are indicated at the left of the panel. Samples of Sp1ZF15 were analyzed by SDS-PAGE. Proteins were visualized by Coomassie blue staining. (C) The CD spectrum of Sp1ZF15. The spectrum was recorded on a Jasco J-720 spectropolarimeter in Hepes buffer (20 mM Hepes (pH 7.5), 50 mM NaCl, 1 mM dithiothreitol, and 10% glycerol) at 20 °C. The protein concentration was 3 μM.

of the manufacturer except for the buffer contents. The protein was cleaved from the GST moiety by 10 units of thrombin at 22 °C for 30 min. The fidelity of purified Sp1ZF15 was confirmed by SDS-PAGE (Fig. 1B) and time-of-flight mass spectroscopy (TOFMS), using a Voyager-DE STR system (Applied Biosystems): [MH^+] calcd. 570,97.6 and observed 569,70.2. The sample was further purified by HPLC on a COSMOSIL $5C_{18}$-AR-II (10 mm × 250 mm) column. The purified protein was lyophilized and dissolved in the buffer just before the experiment was conducted. The secondary structure of Sp1ZF15 was examined by circular dichroism (CD) spectroscopy (Fig. 1C). The value of $[\theta]_{222}$ was approximately $-20{,}000$ (deg cm^2 dmol^{-1}), corresponding to that of Sp1ZF15 consisting of 5 units of Sp1 three-zinc finger peptide.[17]

1.2. DNA binding of Sp1ZF15 to various lengths of DNA

The Sp1 zinc finger specifically recognizes and binds to GC box, 5′-GGG GCG GGGCC-3′.[19–22] DNA binding of Sp1ZF15 for GC-, 3GC-, and 5GC-boxes (Fig. 2A) were evaluated. The dissociation constants (K_d) were determined by gel mobility shift assays (Fig. 2B). In these results, the binding to GC-box showed single shifted band, and to 3GC- and 5GC-boxes showed several shifted bands on protein binding. To determine K_d values, quantity of shifted band was evaluated. The K_d values were 68 ± 0.7, 21 ± 1.1, and 7.0 ± 0.3 (nM) for GC-, 3GC-, and 5GC-boxes, respectively. Compared to the K_d value for GC-box, those for 3GC- and 5GC-boxes presented 3- and 10-fold increase in binding affinities, respectively. In general, the length of DNA binding sequence increases in proportion to the length of zinc finger motif.[16] In this study, we demonstrated that the affinity of a multi-zinc finger for DNA is increased as the length of the DNA is increased.

1.3. DNA binding modes of Sp1ZF15 to 3GC- and 5GC-box DNA

The DNA binding modes of Sp1ZF15 for 5GC- and 3GC-boxes was assessed by DNase I footprinting analyses. The footprint was clearly observed at 5GC-box site (Fig. 3A). This result strongly indicates that Sp1ZF15 specifically binds to 5GC-box site. However, the footprint was weak around three base pairs at the both ends of the site. The enhancement of cleavage by DNase I was observed at the 3′-external region of 5GC-box. This cleavage suggests that Sp1ZF15 recognizes the 3′-end of the site as exhibited in the binding of the Sp1 zinc finger to GC-box.[19,20] Evidently, the footprint on the C-strand was also exhibited on the entire 5GC-box site (data not shown). For binding to 3GC-box, the clear footprint similarly occurred at the binding site, but the enhancement of cleavage was not observed (Fig. 3B). In addition, as the same with binding to 5GC-box DNA, the weak footprint was observed at the both ends of the site. In this DNA binding, Sp1ZF15 is presumed to use about nine-zinc finger domains for recognition on 3GC-box site. Although it is not clear which finger domains of Sp1ZF15 are employed for this recognition, this 15-zinc finger protein distinguishes the length of the binding sites by employing the finger domains of corresponding numbers in DNA binding. As suggested by

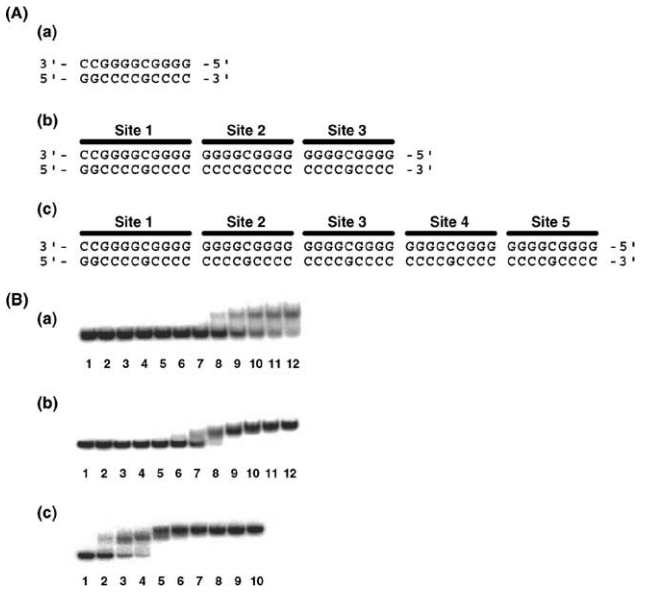

Figure 2. (A) A DNA sequence alignment of target DNA, GC- (a), 3GC- (b), and 5GC- (c) boxes. The sequences displaced by units are represented. The upper and lower sequences are designated as the G- and C-strands, respectively. (B) Gel mobility shift assays of Sp1ZF15 for binding to GC- (a), 3GC- (b), and 5GC- (c) boxes. The lanes 1–12 in the panels (a) and (b) represent the protein concentrations 0, 0.244, 0.488, 0.977, 1.95, 3.91, 7.81, 15.6, 31.3, 62.5, 125, and 250 (nM), respectively. The lanes 1–10 in the panel (c) represent the protein concentrations 0, 1.95, 3.91, 7.81, 15.6, 31.3, 62.5, 125, 250, and 500 (nM), respectively. The reaction time was 24 h for GC- and 3GC-boxes, and 72 h for 5GC-box. Each reaction mixture contained 20 mM Hepes (pH 7.5), 50 mM NaCl, 1 mM dithiothreitol, 10 μM $ZnCl_2$, 25 ng/μL poly(dI-dC), 0.05% Nonidet P-40, 5% glycerol, 40 ng/μL bovine serum albumin, the ^{32}P 5′-end-labeled substrate DNA fragment (~50 pM), and protein. The binding reactions were performed at 20 °C. The dissociation constants (K_d) are averages of three independent determinations with standard deviations.

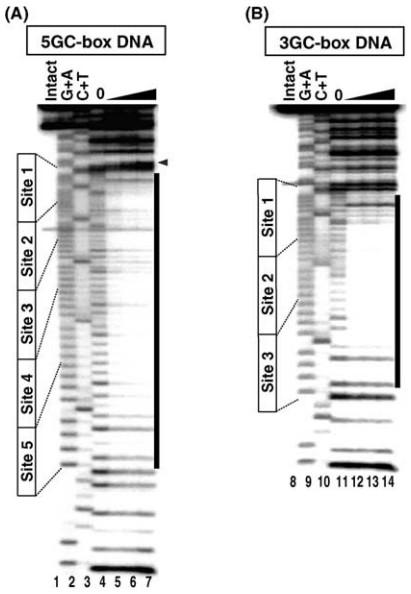

Figure 3. DNase I footprinting analyses of Sp1ZF15 for binding to 5GC- (A) and 3GC- (B) boxes. DNase I footprinting experiments were performed according to the method of Brenowitz et al.[18] Sites indicated in the panels correspond to the DNA sequence in Figure 1. The panels are the reactions for the G-strand. Lanes 1 and 8, intact DNA; lanes 2 and 9, G + A (Maxam–Gilbert reaction product); lanes 3 and 10, C + T (Maxam–Gilbert reaction product); lanes 4–7 and 11–14 are DNase I reaction products at the protein concentrations 0, 0.25, 0.50, and 1.0 (μM), respectively. The triangle indicates the enhancement site of cleavage by DNase I. The thick bars at the side of the panels exhibit the region of the footprint.

the results of DNase I footprinting, the differences in binding affinities depending on the length of target sites reflect the altered finger numbers active in DNA binding.

1.4. Alteration of DNA binding of Sp1ZF15 to ATA permutated DNA

To elucidate the binding mode of Sp1ZF15 in the sequence alteration from G·C to A·T base pairs, 5GCatI- and 5GCatII-boxes, which have the ATA permutated targets, were prepared (Fig. 4A). The dissociation constants (K_d) of Sp1ZF15 were determined to be 8.5 ± 0.1 and 11 ± 0.1 (nM) for 5GCatI- and 5GCatII-boxes, respectively (Fig. 4B). Sp1ZF15 exhibited a slight decrease in binding affinities to these sequences compared with that to 5GC-box. Surprisingly, the DNA footprint pattern was changed depending on the positions of the ATA triplet (Fig. 4C). In both sequences, a distinct footprint was detected at the 5′-side residue of the ATA triplet. In addition, a weak footprint was observed at the 3′-side. In both fragments, cleavage enhancement occurred at the 5′-side adenine of the ATA triplet. Presumably, this result indicates a local distortion of the DNA helix at different positions on protein binding. Interestingly, the observed enhancement sites of cleavage by DNase I were changed depending on the positions of the ATA triplet. The binding to the ATA permutated fragments reveals that Sp1ZF15 binds to the entire sequence, but no finger domains of Sp1ZF15 were used for binding to the ATA site. These results show that Sp1ZF15 recognizes the ATA permutated site in 5GC-box, and that the permutation to the ATA triplet and its positions affects the binding modes and affinities of Sp1ZF15. The versatile recognition on altered DNA sequences is characteristic of Sp1ZF15, which is remarkably unique among natural DNA binding modules.

Engineered zinc finger domains have the potential to mediate sequence-specific protein–DNA interaction for all possible codon combinations or unique sites from the human genome.[7] The results presented here suggest that Sp1ZF15 can bind to various lengths of DNA with considerable affinity and can also recognize changes in the DNA sequence. These properties may allow us to engineer Sp1ZF15 type zinc finger proteins to recognize several sets of target sites or to bind in different modes. The zinc finger module is able to locate as the effector domains on adequate DNA sites for the gene activation (or repression) and DNA cleavage.[23,24] The various binding modes of 15-zinc finger proteins would offer new functional control of such domains.

In summary, the longest artificial zinc finger protein, Sp1ZF15, has control of the number of finger domains active in DNA binding corresponding to the length and sequence of the target DNA. In our opinion, artifi-

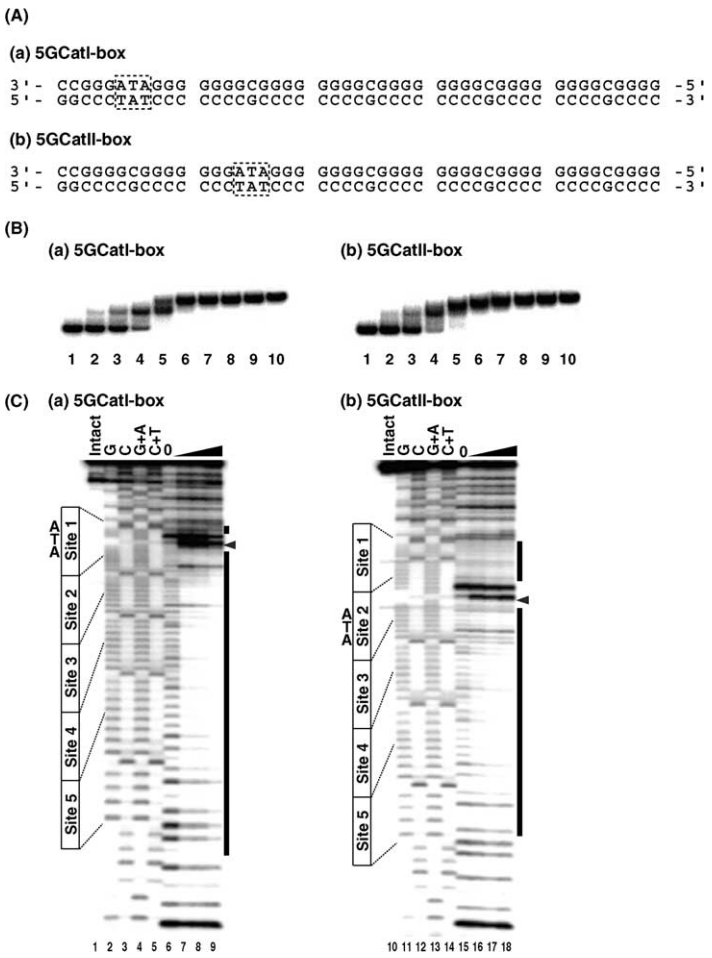

Figure 4. (A) Base sequences of 5GCatI- (a) and 5GCatII- (b) boxes. Designations of each strand are exhibited in the figure. (B) Gel mobility shift assays of Sp1ZF15 for binding to 5GCatI- (a) and 5GCatII- (b) boxes. The reaction time was 72 h. The lanes 1–10 in panels represent the protein concentrations 0, 1.95, 3.91, 7.81, 15.6, 31.3, 62.5, 125, 250, and 500 (nM), respectively. (C) DNase I footprinting analyses of Sp1ZF15 for binding to 5GCatI- (a) and 5GCatII- (b) boxes. The panels are the reactions for the G-strands. Lanes 1 and 10, intact DNA; lanes 2 and 11, G + A (Maxam–Gilbert reaction product); lanes 3 and 12, C + T (Maxam–Gilbert reaction product); lanes 4 and 13, G (Maxam–Gilbert reaction product); lanes 5 and 14, C (Maxam–Gilbert reaction product). DNase I reaction products at the protein concentrations 0, 0.25, 0.50, and 1.0 (μM) are shown in lanes 6–9 and 15–18. The triangles indicate the enhancement site of cleavage by DNase I. The thick bars at the side of the panels exhibit the region of the footprint. The ATA permutated sites are shown at the side of the boxes for sequence indication.

cial transcription factors based upon these multi-zinc finger proteins have great potential for the regulation of a vast number of cellular processes.

Acknowledgements

This study was supported in part by a Grant-in-Aid for COE Project 'Element Science' (12COE2005) and a Grant-in-Aid for Scientific Research (14370755• 16659028) from the Ministry of Education, Culture, Sports, Science, and Technology, Japan. W.N. is a research fellow of the Japan Society for the Promotion of Science (JSPS). The authors also thank Dr. Shiro Futaki and Toshihide Takeuchi for advice on part of the experiments.

References and notes

1. Dynan, W. S.; Tjian, R. *Cell* **1983**, *35*, 79–87.
2. Dynan, W. S.; Tjian, R. *Cell* **1983**, *32*, 669–680.
3. Klug, A. *J. Mol. Biol.* **1999**, *293*, 215–218.
4. Choo, Y.; Isalan, M. *Curr. Opin. Struct. Biol.* **2000**, *10*, 411–416.
5. Wolfe, S. A.; Nekludova, L.; Pabo, C. O. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 183–212.
6. Pabo, C. O.; Peisach, E.; Grant, R. A. *Annu. Rev. Biochem.* **2001**, *70*, 313–340.
7. Segal, D. J.; Barbas, C. F., III. *Curr. Opin. Biotechnol.* **2001**, *12*, 632–638.
8. Ruiz i Altaba, A.; Perry-O'Keefe, H.; Melton, D. A. *EMBO J.* **1987**, *6*, 3065–3070.
9. Matthews, J. M.; Kowalski, K.; Liew, C. K.; Sharpe, B. K.; Fox, A. H.; Crossley, M.; Mackay, J. P. *Eur. J. Biochem.* **2000**, *267*, 1030–1038.
10. Laity, J. H.; Dyson, H. J.; Wright, P. E. *J. Mol. Biol.* **2000**, *295*, 719–727.
11. Peisach, E.; Pabo, C. O. *J. Mol. Biol.* **2003**, *330*, 1–7.
12. Roberts, R. J.; Macelis, D. *Nucleic Acids Res.* **1997**, *25*, 248–262.
13. Lambowitz, A. M.; Belfort, M. *Annu. Rev. Biochem.* **1993**, *62*, 587–622.
14. Clemens, K. R.; Liao, X.; Wolf, V.; Wright, P. E.; Gottesfeld, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 10822–10826.

15. Nolte, R. T.; Conlin, R. M.; Harrison, S. C.; Brown, R. S. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 2938–2943.
16. Kamiuchi, T.; Abe, E.; Imanishi, M.; Kaji, T.; Nagaoka, M.; Sugiura, Y. *Biochemistry* **1998**, *37*, 13827–13834.
17. Nagaoka, M.; Shiraishi, Y.; Uno, Y.; Nomura, W.; Sugiura, Y. *Biochemistry* **2002**, *42*, 8819–8825.
18. Brenowitz, M.; Senear, D. F.; Shea, M. A.; Acker, G. K. *Methods Enzymol.* **1986**, *130*, 132–181.
19. Kriwacki, R. W.; Schultz, S. C.; Steitz, T. A.; Caradonna, J. P. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 9759–9763.
20. Kuwahara, J.; Yonezawa, A.; Futamura, M.; Sugiura, Y. *Biochemistry* **1993**, *32*, 5994–6001.
21. Nagaoka, M.; Sugiura, Y. *Biochemistry* **1996**, *35*, 8761–8768.
22. Yokono, M.; Saegusa, N.; Matsushita, K.; Sugiura, Y. *Biochemistry* **1998**, *37*, 6824–6832.
23. Segal, D. J.; Barbas, C. F., III. *Curr. Opin. Chem. Biol.* **2000**, *4*, 34–39.
24. Bibikova, M.; Beumer, K.; Trautman, J. K.; Carroll, D. *Science* **2003**, *300*, 764.