ELSEVIER

Focused Review

Artificial zinc finger peptides: creation, DNA recognition, and gene regulation

Makoto Nagaoka, Yukio Sugiura*

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

Abstract

Proteins control most biological reactions and the disorder of their expression level causes many diseases. The advent of genomic sequencing and the availability of the complete sequences of several genomes provide new opportunities to study biology and to develop therapeutic strategies through specific modulation of the transcription of target genes. Therefore, regulation of the transcription level by ''artificial repressors'' is of special importance. Of the DNA-binding motifs that have been manipulated by design or selection, Cys_2-His_2 zinc finger proteins have demonstrated the greatest potential for manipulation into general and specific transcription factors. Of special interest is the feature that this family of proteins has modular structures and can recognize a diverse set of DNA sequences in a sequence-specific manner. Therefore, zinc finger motifs offer an attractive framework for the design of novel DNA binding proteins, and such a DNA binding protein would be expected to possess a unique binding sequence with high specificity and affinity. Principally, two approaches have been taken to the design of artificial zinc finger proteins. One is a selection strategy via phage display methods to alter the recognition sequence, and another is a structure-based linking strategy to extend the length of a DNA recognition sequence. Such novel zinc finger peptides (or proteins) offer great promise for genome-specific transcription switches in the near future. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Cys_2-His_2 zinc finger protein; Phage display strategy; Multiple zinc finger protein; DNA bending finger

1. Introduction

Proteins control most biological reactions and the disorder of their expression level causes many diseases. The advent of genomic sequencing and the availability of the complete sequences of several genomes provide new opportunities to study biology and to develop therapeutic strategies through specific modulation of the transcription of target genes. Therefore, regulation of the transcription level by ''artificial repressors'' is of special importance. As a first step to controlling artificial gene expression, artificial repressors are required to have new DNA binding specificities for rare binding sites associated only with the target gene, but probably distinct from the natural promoter. The size of a human genome amounts to more than three-billion base pairs. At present, the molecular design of DNA binding proteins with desired recognition sites within complex genomes is a challenging problem. Of the DNA-binding motifs that have been manipulated by design or selection, Cys_2-His_2 zinc finger proteins have demonstrated the greatest potential for manipulation into general and specific transcription factors.

The zinc finger motif of Cys_2-His_2 type is one of the most common DNA binding motifs found in eukaryotes. For example, more than 3% of the protein sequences inferred from the *Caenorhabitis elegans* genome contain sequence motifs characteristic of zinc-binding structural domains. Furthermore, approximately 0.7% of all proteins in *C. elegans* involve one or more Cys_2-His_2 zinc finger domains [1]. Each Cys_2-His_2 zinc finger domain consists of approximately 30 amino acids and typically binds three base pairs of double-stranded DNA sequences (Fig. 1) [2].

Of special interest is the feature that this family of proteins has modular structures and can recognize a diverse set of DNA sequences in a sequence-specific manner. Therefore, zinc finger motifs offer an attractive framework for the design of novel DNA binding proteins, and such a DNA binding protein would be expected to possess a unique binding sequence with high specificity and affinity. Principally, two approaches have been taken

*Corresponding author. Tel.: +81-774-38-3210; fax: +81-774-32-3038.

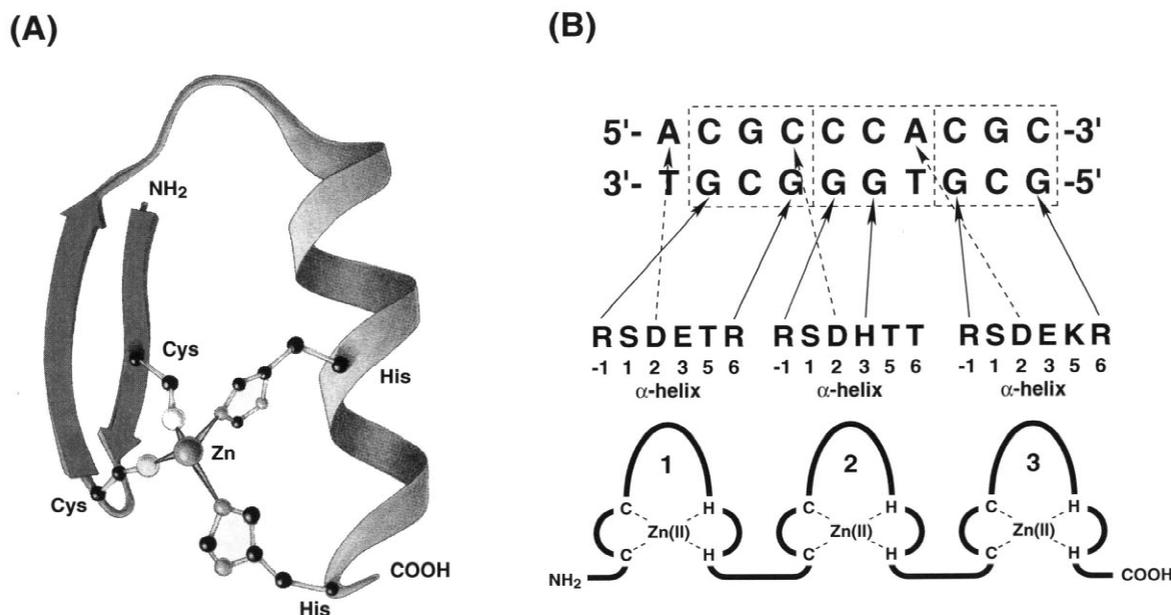*E-mail address:* sugiura@scl.kyoto-u.ac.jp (Y. Sugiura).

Fig. 1. Zinc finger structure (A) and DNA recognition mode of Cys_2-His_2 type zinc finger protein Zif268 (B).

to the design of artificial zinc finger proteins. One is a selection strategy via phage display methods starting from naturally occurring zinc finger proteins to alter the recognition sequence, and another is a structure-based linking strategy to extend the length of the DNA recognition sequence. Such novel zinc finger peptides (or proteins) offer great promise for genome-specific transcription switches in the near future.

2. Selection strategies

The zinc finger domain of the Cys_2-His_2 type belongs to a typical class of DNA binding protein and commonly contains a sequence of the type, (Tyr, Phe)-X-Cys-X_{2-4}-Cys-X_3-Phe-X_5-Leu-X_2-His-X_{3-5}-His, where X represents relatively non-conserved amino acids [3]. Each finger is approximately 30 amino acid residues long and consists of a simple $\beta\beta\alpha$-fold stabilized by hydrophobic interactions and also by chelation of a simple zinc ion with the conserved Cys_2-His_2 residues (Fig. 1(A)) [4–6].

The X-ray crystal structures of the Zif268- and the GLI–DNA complexes clearly revealed the characteristic DNA binding mode of Cys_2-His_2 zinc finger proteins as follows: (1) one finger typically recognizes contiguous three base pairs (bp) of a DNA sequence; (2) DNA recognition is mediated through base contacts with the side chains of specific amino acids located on the recognition helix; (3) multiple zinc finger domains are tandemly repeated by simple covalent linkage, namely connected by the consensus linker region; and (4) zinc finger proteins bind to the sequence of asymmetric base pairs unlike other nucleic acid recognition motifs such as a basic leucine-zipper and a helix-turn-helix (Fig. 1(B)) [7–11].

For designing proteins with desired sequence specificities, selection systems based on filamentous phage display have been used and developed [12]. In this system, DNA encoding the target proteins is fused to the gene for bacteriophage pIII protein, and the hybrid protein is expressed on the surface of the phage and also a library of variants is prepared by randomizing critical amino acids. This system has been applied to screening for the zinc fingers with novel binding specificities. The resulting "zinc finger phages" present very attractive characteristics for in vitro genetic selections. That is to say, they carry the zinc finger protein gene and exhibit the DNA binding properties of the zinc finger protein. Indeed, Pabo and co-workers tested this procedure in combination with sequential selection and successfully obtained new zinc finger peptides with TATA box, p53 binding site, and nuclear receptor elements (Fig. 2) [13,14]. The proteins obtained for these important regulatory sites bind to specific DNA sequences with nanomolar dissociation constants and discriminate effectively (greater than 20 000-fold) against non-specific DNA.

The sequential selection strategy provides a general and effective method for the design of new zinc finger proteins with novel DNA binding specificities. By using the phage display system, it may be possible to select zinc finger proteins for many important regulatory sequences.

3. Structure-based linking strategies

On the basis of the DNA recognition mode and the structural features unique to the Cys_2-His_2 class of nucleic acid binding, approaches to link Cys_2-His_2 zinc fingers with other functional modules such as DNA binding
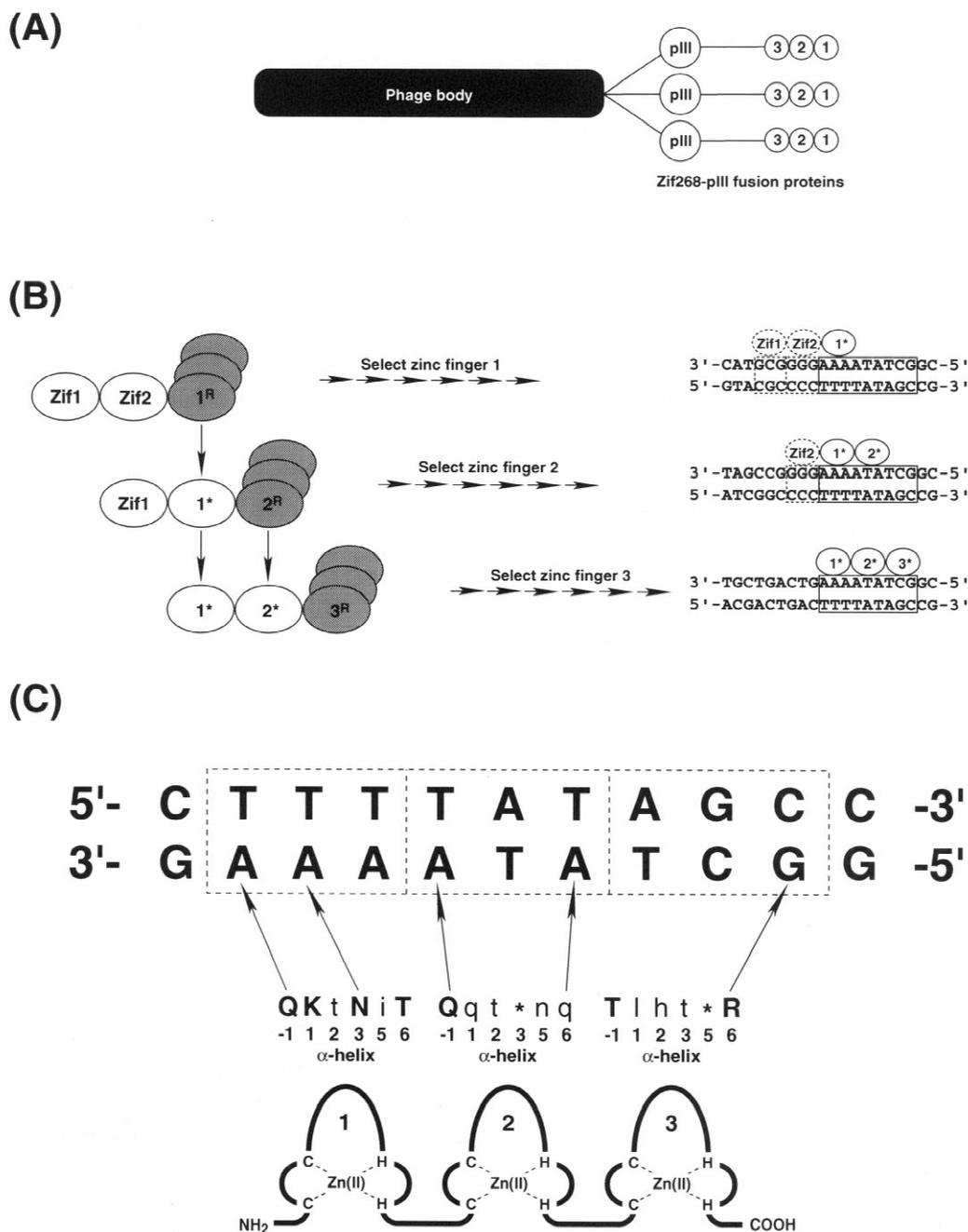
Fig. 2. (A) Sketch of the zinc finger phage. (B) Overview of protocol that successively selects fingers one, two and finger three to create a new zinc finger protein. Zif1 and Zif2, wild-type Zif268 fingers; R, a randomized finger library; and asterisk, a selected finger. Small horizontal arrows indicate the multiple cycles of selection and amplification used when selecting each finger by phage display. The right side shows the binding sites used in selections with the TATA site and indicates the overall binding mode for the selected fingers. (C) Amino acid sequences of the new zinc finger peptide selected by phage display that recognize the TATA box. Fully conserved, partially conserved, and non-conserved residues are shown in boldface, lower case letters, and an asterisk in the consensus sequence, respectively.

domains to generate artificial chimeric peptides with long binding sites [15,16], and DNA-cleavage modules to produce novel sequence specific nucleases [17,18], have been performed. Artificial chimeric DNA binding peptides display sequence specificity for extended, chimeric DNA binding sites. In other words, unlike most of the DNA binding proteins such as many restriction enzymes, these fusion peptides can recognize the sequences of asymmetric base pairs. In addition, specific delivery of a DNA binding protein to a single site within a genome as complex as that found in humans, three-billion base pairs, requires an address of at least 16 or 17 base pairs. Statistically, assuming random base distribution, a unique 16- or 17-base pair sequence will occur only once in 4.3 billion or 17

billion nucleotides, roughly the same or a bigger size than the human genome (3.0×10^9 base pairs). Although natural proteins containing long polydactyl arrays of zinc finger domains have been inferred from the sequence, no zinc finger proteins have been demonstrated to bind such a long, contiguous DNA sequence. Therefore, it is of special interest that a six- or nine-fingered peptide has been created and demonstrated to bind 18 or 27 contiguous base pairs of DNA in a sequence-specific fashion. Such poly-dactyl zinc finger peptides should be broadly applicable as genome-specific transcription switches in gene therapy strategies and the development of novel transgenic plants and animals. To date special zinc fingers have been fused to a homeodomain [19,20], to the TATA-binding protein [21], and to other zinc fingers [15,16,22,23]. The fusion proteins are useful as prototypes for exploring how zinc fingers can target particular promoters and thus provide site-specific regulation of gene expression.

By using a structure-based design strategy, Kim et al. prepared a new fusion protein in which the three zinc fingers of Zif268 were linked to the COOH terminus of the yeast TATA box-binding protein [21]. This protein forms an extraordinarily stable complex (half-life, 410 h) with a DNA sequence (5′-ATGCGTGGGCGTCCGTATATAAAGCGAC-3′). In vitro transcription experiments and transient cotransfection assays revealed that the fusion protein acts as a site-specific repressor.

We have created novel multiple zinc finger peptides, Sp1ZF6 and Sp1ZF9, from the three zinc finger motif of transcription factor Sp1 (Fig. 3) [15]. Sp1ZF6 and Sp1ZF9 were constructed by linking two and three Sp1's zinc finger domains with the Krüppel-type linker (TGEKP) peptide. DNA binding of the nine zinc finger Sp1ZF9 was compared with the native three zinc finger Sp1 (530–623) and the artificial six zinc finger Sp1ZF6 peptide. Transcription factor Sp1 encompasses three Cys_2-His_2 type zinc finger motifs as the DNA binding domain and binds specifically to the GC (5′-GGG-GCG-GGG-3′) sequence. DNaseI footprinting assays and methylation interference analyses demonstrated that Sp1ZF6 and Sp1ZF9 efficiently formed specific binding complexes with the expected 18 and 27 bp DNA sequences, respectively, by using all fingers. These artificial multiple zinc finger peptides show extended sequence specificity and their favorable se-quences depend on both the number of motifs and the character of the Sp1's three-zinc-finger DNA binding domain. The DNA binding of the native nine-zinc-finger protein TFIIIA is well known to be dominated by inter-action of a select few fingers. Therefore, it is of special interest to create a new multiple zinc finger peptide that can bind to a DNA sequence over an extended region of
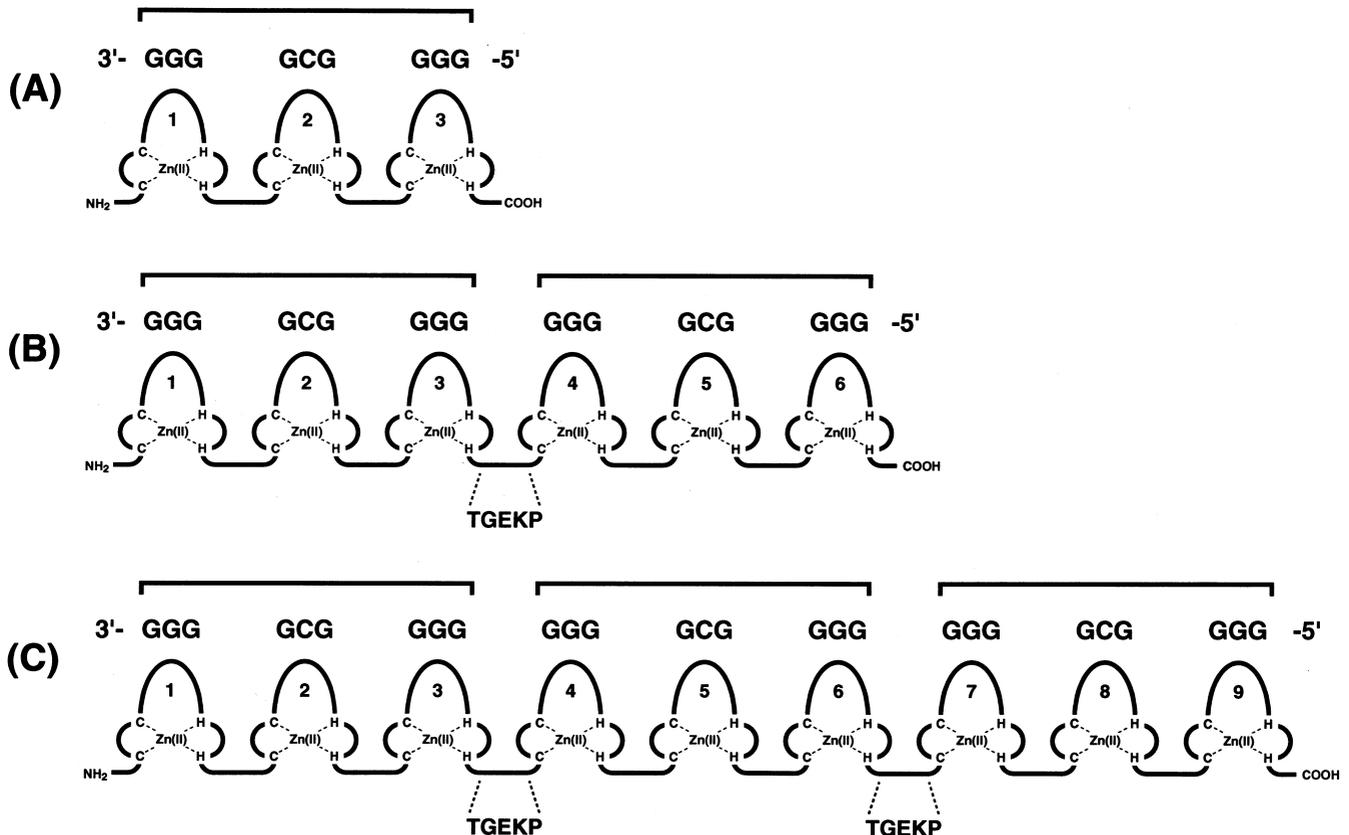

Fig. 3. Schematic models of Sp1 (530–623), Sp1ZF6, and Sp1ZF9 and their target sequences.

20–30 base pairs. Such a result would provide useful information for the design of new DNA binding proteins to recognize long DNA sequences. To specify a single site within the three billion base pairs of the human genome, DNA binding ligands which specifically recognize approximately 20 base pairs are necessary. For this reason, recognition of a long sequence represents a milestone in the development of chemical and biochemical approaches to DNA recognition.

On the basis of the multiple zinc finger peptide described above, we designed a DNA bending finger to regulate gene expression (Fig. 4) [24]. DNA structural changes such as bending play an important role in various biological reactions. Not only is protein binding to specific DNA sequences indispensable for unique gene expression but so also is DNA bending induced by the protein. Therefore, an artificial protein that induces a DNA conformational change is interesting as a transcriptional regulator of a specific gene. We created six zinc finger proteins, Sp1ZF6(Gly)n ($n=4$, 7, 10), by connecting two DNA binding domains of transcription factor Sp1 through flexible polyglycine peptide linkers. Gel mobility shift and footprint analyses revealed that Sp1ZF6(Gly)7 and Sp1ZF6(Gly)10 bind to two distal GC boxes and result in DNA bending. The phasing assays strongly suggested that the induced DNA bending was directed toward the major groove and that Sp1ZF6(Gly)7 caused the most drastic directional change in DNA bending. Of special interest are the facts that newly designed six-finger peptides

Sp1ZF6(Gly)7 and Sp1ZF6(Gly)10 can induce DNA bending at the intervening region of the two distal binding sites and that the linker length between two and three zinc finger motifs has a crucial effect on the entire DNA bending direction. Such DNA-bending fingers may be feasible for use as a gene expression regulator based on structural change in DNA in the future.

Artificial restriction endonucleases with novel sequence specificities have also been engineered by combining zinc finger proteins (Sp1) with a glycylglycylhistidine site [17] or by linking two different zinc finger proteins (Zif268 and Sp1) to the cleavage domain of *Fok* I endonuclease [18] (Fig. 5). These designed proteins are active and indeed cleave double-stranded DNA in a sequence-specific manner under optimal conditions. The results are of special interest because of the potential versatility of zinc finger proteins in recognizing different DNA sequences and as a guide to de novo design of the novel artificial restriction enzyme based on the zinc finger motif applicable to chromosome mapping and sequencing.

4. Perspective

Cys$_2$-His$_2$ zinc finger proteins provide an attractive framework for the design and selection of proteins with designed site specificity. Based on selection strategies by phage display techniques, novel zinc finger proteins with diverse DNA target sites have been generated. On the other
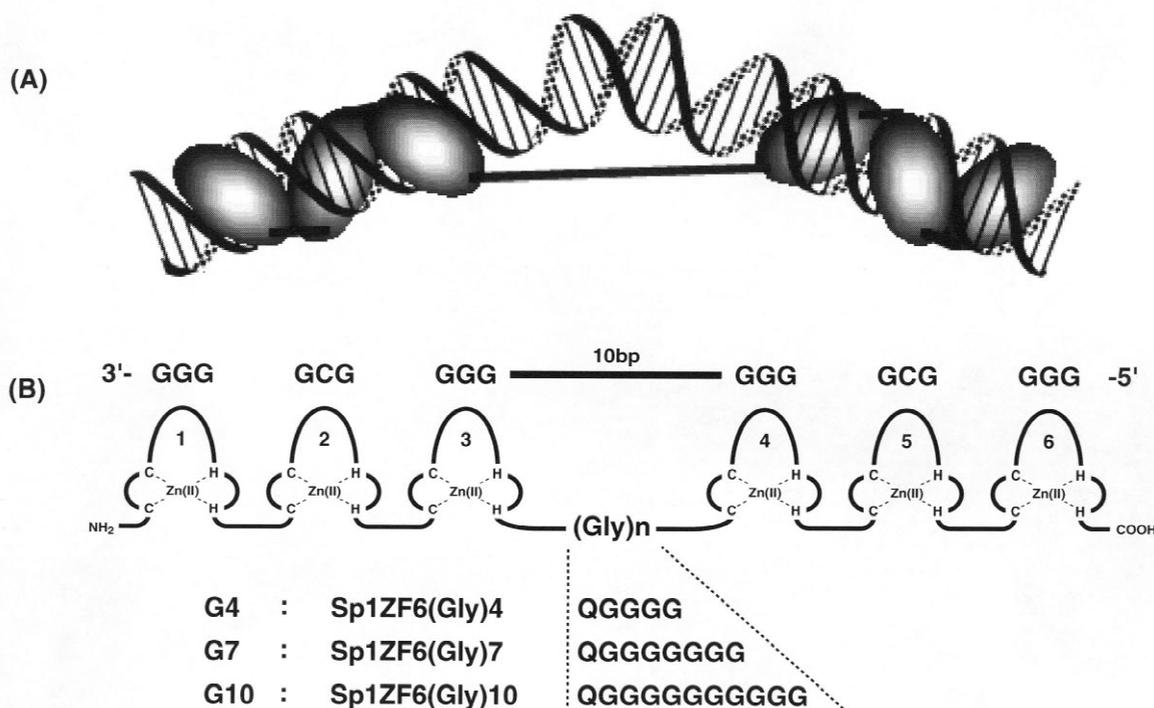

Fig. 4. Model of DNA bending induced by zinc finger peptide (A) and artificial DNA bending fingers Sp1ZF6(Gly)4, Sp1ZF6(Gly)7, and Sp1ZF6(Gly)10 (B).
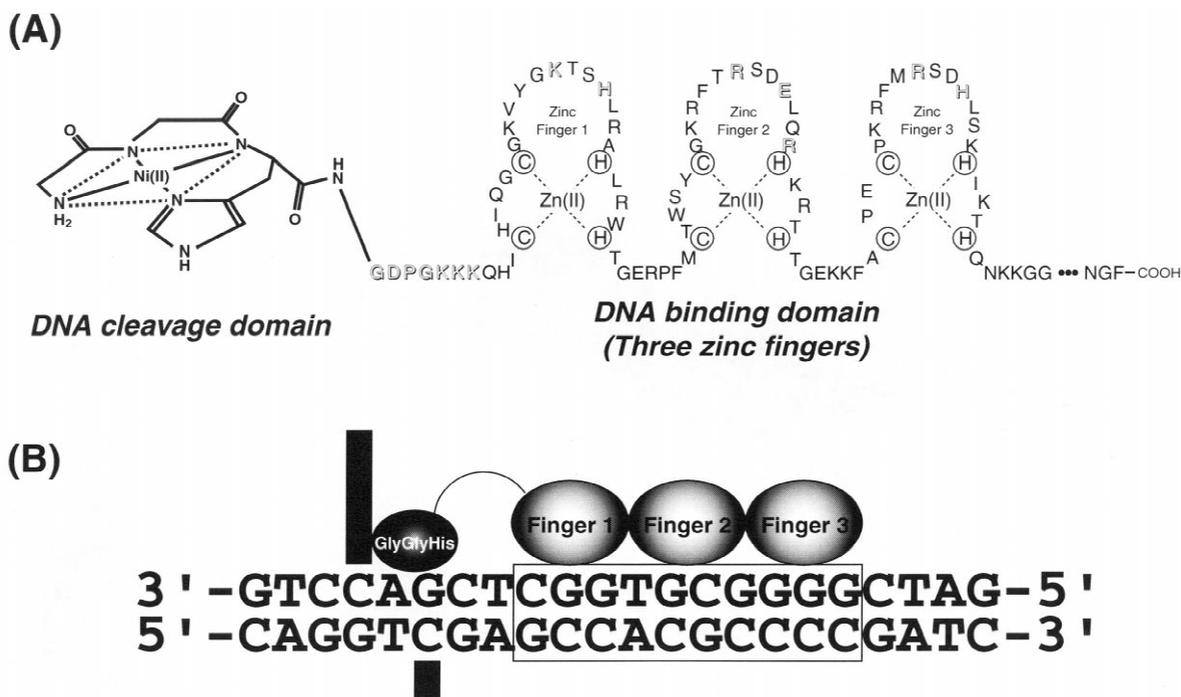
Fig. 5. Artificial zinc finger peptide Sp1GGH as DNA cutter (A) and its DNA cleavage pattern (B). The box indicates the GC box, and the length of the bars represents the extent of cleavage.

hand, artificial multiple zinc finger proteins with long binding sites have also been created on the basis of structure-based linking strategy. Such designed zinc finger proteins bind with nanomolar dissociation constants and discriminate effectively against non-specific DNA. In addition, these proteins can function in living cells as transcriptional activators or repressors. Moreover, six zinc finger proteins with a long linker sequence have also been designed to induce DNA bending. The DNA bending finger may be one of the prototypes for novel gene expression regulators. In general, the combination strategy of phage display and structure-based design is possible to specifically recognize almost any desired promoter and thus regulate endogeneous gene expression by targeting zinc fingers to the flanking bases and constructing appropriate variants.

Retroviral nucleocapsid proteins are known to contain one or two CCHC-type zinc knuckle domains. Recently, the structure of the HIV-1 nucleocapsid protein bound to the RNA recognition element determined by NMR spectroscopy reveals that certain residues within the CCHC arrays significantly contribute to the RNA recognition specificity [25]. Therefore, this information may provide insight into the creation of new RNA-binding zinc finger proteins that serve for the development of inhibitors designed to interfere with genome encapsidation.

Transcriptional control at the DNA level presents distinct advantages over antisenses or ribozymes, because only a simple site needs to be occupied compared with targeting the many copies of mRNA. Moreover, a structur-

al change of DNA in the promoter region influences the transcription level of the gene. Accordingly, artificial zinc finger proteins of the types described here could find broad application in future gene therapy strategies. Highly specific control of gene expression in vivo can be achieved using artificial transcription factors containing the novel designed zinc finger DNA binding domains, provided delivery is carefully controlled.

5. Definition list

bp   base pair

References

[1] N.D. Clarke, J.M. Berg, Science 282 (1998) 2018–2022.
[2] M. Imanishi, Y. Hori, M. Nagaoka, Y. Sugiura, Eur. J. Pharm. Sci. (in press).
[3] J. Miller, A.D. McLachlan, A. Klug, EMBO J. 4 (1985) 1609–1614.
[4] M.S. Lee, G.P. Gippert, K.Y. Soman, D.A. Case, P.E. Wright, Science 245 (1989) 635–637.
[5] R.E. Klevit, J.R. Herriot, S.J. Horvath, Proteins 7 (1990) 215–226.
[6] J.G. Omichinski, G.M. Clore, E. Apella, K. Sakaguchi, A.M. Gronenborn, Biochemistry 29 (1990) 9324–9334.
[7] N.P. Pavletich, C.O. Pabo, Science 252 (1991) 809–817.
[8] N.P. Pavletich, C.O. Pabo, Science 261 (1993) 1701–1707.
[9] J.M. Berg, Y. Shi, Science 271 (1996) 1081–1085.
[10] M. Elrod-Erickson, M.A. Rould, L. Nekludova, C.O. Pabo, Structure 4 (1996) 1171–1180.

[11] M. Elrod-Erickson, T.E. Benson, C.O. Pabo, Structure 6 (1998) 451–464.

[12] J.K. Scott, G.P. Smith, Science 249 (1990) 386–390.

[13] H.A. Greisman, C.O. Pabo, Science 275 (1997) 657–661.

[14] J.S. Kim, J. Kim, K.L. Cepec, P.A. Sharp, C.O. Pabo, Proc. Natl. Acad. Sci. USA 94 (1997) 3616–3620.

[15] T. Kamiuchi, E. Abe, M. Imanishi, T. Kaji, M. Nagaoka, Y. Sugiura, Biochemistry 37 (1998) 13827–13834.

[16] Q. Liu, D.J. Segal, J.B. Ghiara, C.F. Barbas III, Proc. Natl. Acad. Sci. USA 94 (1997) 5525–5530.

[17] M. Nagaoka, M. Hagihara, J. Kuwahara, Y. Sugiura, J. Am. Chem. Soc. 116 (1994) 4085–4086.

[18] Y.G. Kim, J. Cha, S. Chandrasegaran, Proc. Natl. Acad. Sci. USA 93 (1996) 1156–1160.

[19] J.L. Pomerantz, P.A. Sharp, C.O. Pabo, Science 267 (1995) 93–96.

[20] J.L. Pomerantz, C.O. Pabo, P.A. Sharp, Proc. Natl. Acad. Sci. USA 92 (1995) 9752–9756.

[21] J.S. Kim, J. Kim, K.L. Cepec, P.A. Sharp, C.O. Pabo, Proc. Natl. Acad. Sci. USA 94 (1997) 3616–3620.

[22] R.R. Beerli, D.J. Segal, B. Dreier, C.F. Barbas III, Proc. Natl. Acad. Sci. USA 95 (1998) 14628–14633.

[23] J.S. Kim, C.O. Pabo, Proc. Natl. Acad. Sci. USA 95 (1998) 2812–2817.

[24] M. Imanishi, Y. Hori, M. Nagaoka, Y. Sugiura, Biochemistry 39 (2000) 4383–4390.

[25] R.N. De Guzman, Z.R. Wu, C.C. Stalling, L. Pappalardo, P.N. Borer, M.F. Summers, Science 279 (1998) 384–388.