# Interconversion between Serine and Aspartic Acid in the α Helix of the N-Terminal Zinc Finger of Sp1: Implication for General Recognition Code and for Design of Novel Zinc Finger Peptide Recognizing Complementary Strand†

Makoto Nagaoka, Yasuhisa Shiraishi, Yumiko Uno, Wataru Nomura, and Yukio Sugiura*

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

*Received February 6, 2002; Revised Manuscript Received May 26, 2002*

ABSTRACT: In the typical base recognition mode of the $C_2H_2$-type zinc finger, the amino acid residues at α-helical positions −1, 3, and 6 make a contact with the base in one strand (the primary strand), and the residue at position 2 interacts with the base in a complementary strand (the secondary strand). The N-terminal zinc finger of the three-zinc-finger domain of Sp1 has inherently a unique five-base-pair binding mode in which the guanine bases are recognized in both strands. To clarify the effect of the amino acid at position 2 on DNA binding affinity and base specificity, we have created a library of the mutants by the interconversion between serine and aspartic acid in the N-terminal zinc finger of Sp1 and recombinant variants of finger order. Gel mobility shift and methylation interference assays showed that the combination of arginine and serine at positions −1 and 2, respectively, provides a newly strong guanine contact in the secondary strand and a higher binding affinity than that of wild-type Sp1. Of special interest are the facts that the mutant with lysine and aspartic acid at positions −1 and 2 in the α helix predominantly recognizes the bases in the secondary strand and that its DNA binding affinity is higher than that of the wild-type. The aspartic acid or serine at position 2 independently contributes to the DNA binding affinity and base specificity. The present results provide useful information for the design of a novel zinc finger protein with priority for the bases in the secondary strand.

The regulation of gene expression at the transcription level is one of the most effective strategies for the creation of novel drugs and therapies, because most of the biological reactions are controlled by gene products. To accomplish this challenging project, it is essential that the artificial transcription factors with novel DNA binding specificities and regulatory activities are rationally designed on the basis of known superfamilies of DNA binding proteins. Therefore, the elucidation and modification of DNA binding specificities of proteins in such families and the following systematization of DNA recognition codes are extremely significant. One of the most general superfamilies of DNA binding proteins found in eukaryotes is the $C_2H_2$-type zinc finger (zf)[1], which has a tandemly repeated structure consisting of independent modules with the consensus sequence (Tyr, Phe)-X-Cys-$X_{2,4}$-Cys-$X_3$-Phe-$X_5$-Leu-$X_2$-His-$X_{3-5}$-His-$X_{2-6}$. Each finger domain forms a two-stranded antiparallel $\beta$ sheet and an α helix, which are held together tetrahedrally by coordination of a zinc ion with invariant cysteines and histidines. The zinc finger binds to the three-base-pair subsite with the amino acids at four key positions in the α helix in an antiparallel fashion (*1*−*3*). Therefore, it is expected that the characteristic DNA binding mode makes it possible to design $C_2H_2$-type zinc fingers with various sequence specificities intentionally. In practice, several design- and phage-display-based mutational analyses have been carried out and have shown modest success (*4*−*12*). However, the general principle for DNA recognition codes remains undetermined. The limitation appears to be provided by the complicated interface of the zinc finger−DNA complex (*13*). In addition, it is also assumed that the characteristic DNA binding mode of the zinc finger protein per se restricts the construction of the simple recognition code. In this mode, namely, the base recognition by the key amino acids at positions −1, 3, and 6 in the α helix, is predominant in one strand (the primary strand), and the only base interaction in the complementary strand (the secondary strand) is generally the recognition by the amino acid at position 2 in the α helix (*2*, *3*). The secondary strand recognition is influenced by the synergy between adjacent zinc fingers (*14*); hence, it is not so easy to control the recognition. Nevertheless, the selection of the novel zinc fingers has been carried out with the phage display method by considering the base recognition in the secondary strand (*15*).

Transcription factor Sp1 is a zinc finger protein isolated from human HeLa cell extracts and binds to the GC box by three contiguous repeats of $C_2H_2$-type zinc fingers at the C-terminal region (*16*, *17*). Of the three zinc fingers, only the N-terminal finger (finger 1) has a unique base recognition pattern; the finger binds not to the three- but to the five-base-pair subsite, 5′-GGGCC-3′ (*18*). To clarify the effect

† This study was supported in part by Grants-in-Aid for COE Project "Element Science" (12CE2005) and Scientific Research (12470505, 12793008, 13557210) from the Ministry of Education, Culture, Sports, Science, and Technology, Japan.

* To whom correspondence should be addressed. Phone: +81-774-38-3210. Fax: +81-774-32-3038. E-mail: sugiura@scl.kyoto-u.ac.jp.

[1] Abbreviations: Tris, tris(hydroxymethyl)aminomethane; TN, Tris-NaCl; CD, circular dichroism; zf, zinc finger.

10.1021/bi0256429 CCC: $22.00 
Published on Web 06/20/2002

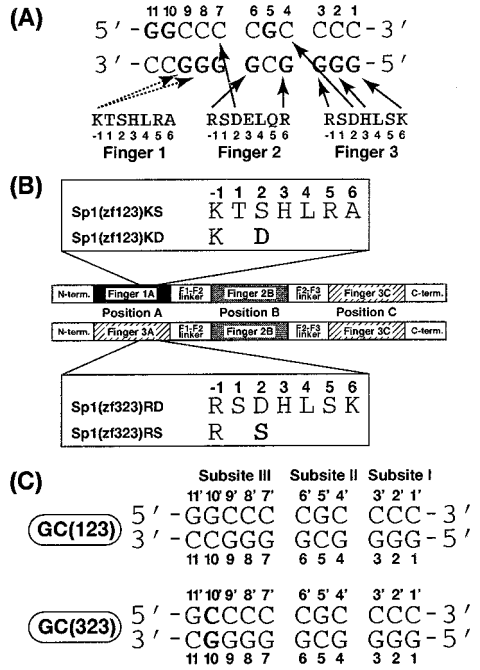


FIGURE 1: (A) Putative base recognition mode of three zinc fingers of Sp1. Amino acid residues at the N-terminus of the α helix in each finger are depicted by their one-letter codes with the number of the helical positions below. Solid arrows show the amino acid−base contacts assumed by the DNA binding mode of Zif268, and dotted arrows depict the contacts indicated by our previous report (*18*). The guanine bases, whose methylation interferes with the zinc finger binding, are in boldfaced print. (B) Primary structures of wild-type and mutant zinc finger peptides of Sp1. The designation of each zinc finger is shown by the original name (fingers 1−3) with an alphabetical letter indicating the absolute position (positions A−C). The amino acid residues of the α-helical positions −1 to 6 of each peptide are indicated by their one-letter codes, and the mutated residues are in boldfaced print. (C) Substrate DNA sequences used in this study. Each sequence is divided into subsites I−III. The substituted nucleotide in GC(323) is depicted in boldfaced print. The base numbers in the sequences are also shown.

of the amino acid at position 2 on DNA binding affinity and base specificity, we have created a library of the mutants by interconversion of the amino acid between serine and aspartic acid in the N-terminal zinc finger of Sp1 and recombinant variants of finger order. Indeed, two novel mutants with the ability to recognize the bases in the secondary strand were obtained. The results strongly indicate that novel zinc finger proteins can be created on the basis of the recognition rule for the base in the secondary strand.

## MATERIALS AND METHODS

*Chemicals.* The T4 polynucleotide kinase and restriction enzymes were purchased from New England Biolabs (Beverly, MA). Taq DNA polymerase and synthesized oligonucleotides for cloning of each mutant peptide were acquired from Qiagen (Valencia, CA) and Amersham Biosciences (Cleveland, OH), respectively. Labeled compound [$\gamma$-$^{32}$P] ATP was supplied by DuPont (Wilmington, DE). The plasmid pBS-Sp1-fl was kindly provided by Dr. R. Tjian. All other chemicals were of commercial reagent grade.

*Preparations of Zinc Finger Peptides from Sp1 and Substrate DNA Fragments.* Figure 1A shows the putative base recognition mode of Sp1 (*18*). On the basis of this recognition mode, we designed all of the zinc finger peptides used in this study, whose primary structures are summarized in Figure 1B. Each peptide is named by the combination of the one-letter codes for the amino acid residues at positions −1 and 2 in the α helix of the finger at position A. Sp1(zf123)KS which is the alias for Sp1(530−623) is coded on the plasmid pEVSp1(530−623), as previously described (*19*). The recombinant variant of finger order Sp1(zf323)RD is also identical to the previously prepared Sp1(zf323) (*20*). All other mutant peptides were created by standard polymerase chain reaction with the primer set of pEVSp1(530−623) as a template. Their sequences were confirmed by a GeneRapid DNA sequencer (Amersham Biosciences). These zinc finger peptides were overexpressed as a soluble form in the *Escherichia coli* strain BL21(DE3)pLysS at 20 °C and purified by the following procedure at 4 °C. *E. coli* cells were resuspended and lysed in TN buffer (10 mM Tris-HCl (pH 8.0), 50 mM NaCl, and 1 mM dithiothreitol). After centrifugation, the supernatant containing the soluble form of the Sp1 zinc finger peptide was purified by cation exchange chromatography using a 0.05−2.0 M NaCl gradient (Uno S, BioRad, Randolph, MA). Final purification was achieved by a gel filtration technique (Superdex 75, Amersham Biosciences) using TN buffer. For the substrate DNAs, the *Hin*d III−*Xba* I fragments of GC(123) (5′-GGG GCG GGG C C-3′) and GC(323) (5′-GGG GCG GGG G C-3′) were cut out and labeled at the 5′ end with $^{32}$P for the experiments as described previously (*10*) (Figure 1C).

*CD Measurements.* The CD spectra for all of the zinc finger peptides of Sp1 were recorded on a Jasco J-720 spectropolarimeter in 10 mM Tris-HCl (pH 8.0), 50 mM NaCl, 1 mM dithiothreitol, and 10 $\mu$M of the zinc finger peptide at 20 °C.

*Gel Mobility Shift Assays.* Gel mobility shift assays were carried out under the previous experimental conditions (*20*). Each reaction mixture contained 10 mM Tris-HCl (pH 8.0), 50 mM NaCl, 1 mM dithiothreitol, 10 $\mu$M $ZnCl_2$, 25 ng/$\mu$L poly(dI−dC), 0.05% Nonidet P-40, 5% glycerol, 40 mg/$\mu$L bovine serum albumin, the $^{32}$P-end-labeled substrate DNA fragment (∼50 pM), and 0−500 nM of the zinc finger peptide. After incubation at 20 °C for 30 min, the sample was run on an 8% polyacrylamide gel with 89 mM Tris-borate buffer at 20 °C. The bands were visualized by autoradiography and quantified with ImageMaster 1D Elite software (version 3.01). The dissociation constants ($K_d$) of the Sp1 peptide−DNA fragment complexes were estimated according to the previously reported procedure (*18*).

*Methylation Interference Analyses.* The recognition of guanines in the primary and secondary strands of GC(123) and GC(323) (the G and the C strands, respectively) by each peptide was investigated by methylation interference assays as described previously (*21*). The binding reaction mixture contained 40 mM Tris-acetate, 50 mM NaCl, 1 mM dithiothreitol, 40 ng/$\mu$L sonicated calf thymus DNA, 5% glycerol, the $^{32}$P-end-labeled methylated DNA fragment (approximately 500 Kcpm), and 20−300 nM of the zinc finger peptide. After incubation at 20 °C for 30 min, the peptide-bound and free DNAs were separated on a 10% nondenaturing polyacrylamide gel and eluted from the gel with a standard elution buffer. To examine both the strong and weak base contacts, we selected the experimental conditions under which the peptide/DNA molar ratio in the binding reaction is about 10−20% bound. The recovered

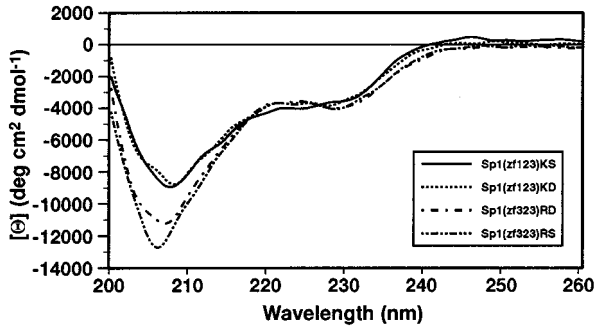


FIGURE 2: CD spectra of wild-type and mutant zinc finger peptides of Sp1 at 20 °C.

methylated DNA was reacted in 100 $\mu$L of 1 M piperidine at 90 °C for 30 min. The lyophilized cleavage products were analyzed on a 15% polyacrylamide/7 M urea sequencing gel. The bands were visualized by autoradiography and quantified with ImageMaster 1D Elite software (version 3.01). The extent of interference was estimated for each base by calculating the ratio of the cutting probabilities in the free and bound lanes.

## RESULTS

*Folding Property of Wild-Type and Mutant Zinc Finger Peptides of Sp1.* In respect to the $\beta\beta\alpha$ structure of the zinc finger, the folding property of the peptides was evaluated by measurements of the CD spectra. Figure 2 shows the CD spectral results for the wild-type and mutant peptides at 20 °C. The spectrum for wild-type Sp1(zf123)KS was similar to those of the single- and three-finger-peptides of Sp1 previously described (*22*−*24*). Negative Cotton effects in the far-UV region with a minimum at 206 nm and a shoulder around 222 nm suggest that Sp1(zf123)KS has an ordered secondary structure. The spectrum for the Sp1(zf123)KD remarkably resembled to that of Sp1(zf123)KS. On the other hand, Sp1(zf323)RD and Sp1(zf323)RS exhibited the spectra somewhat different from that of Sp1(zf123)KS. As for the ellipticities at 206 nm, the value of Sp1(zf123)KS ($[\theta]_{206} = -8454$) was smaller than those of Sp1(zf323)RD and Sp1(zf323)RS ($[\theta]_{206} = -11056$ and $-12751$, respectively). The increase in negative Cotton effects of Sp1(zf323) derivatives was also observed in our previous report (*20*). These results indicate that the conformation of the zinc finger domain in each peptide is not identical but is comparable.

*DNA Binding Affinity of Each Peptide.* From the database analysis, the amino acid residue at position 2 in the $\alpha$ helix is conserved as a serine in approximately 50% of natural zinc finger proteins (*25*). However, the serine is generally suggested to have no significant effect on DNA base recognition or specificity (*26*, *27*). As observed in Zif268−DNA complex, on the other hand, the aspartic acid residue at position 2 plays the following two roles in DNA binding: (1) formation of a hydrogen bond−salt bridge interaction with an arginine residue at position −1 and subsequent stabilization of the guanine−arginine bidentate interaction, and (2) formation of cross-strand interaction with the 3′ base, cytosine or adenine, of the subsite for the preceding finger (*2*, *3*) (Figure 3). To examine whether these interactions affect the DNA binding affinity of the present peptides, we prepared two types of substrate DNA GC(123) and GC(323) and determined the apparent dissociation constants ($K_d$) of each peptide for the DNAs using gel mobility shift assays (Table 1). Uniquely, GC(323) has the C10′ for the afore-described cytosine recognition by aspartic acid. Sp1(zf123)-KS bound to GC(123) and GC(323) with 13.96 and 28.60 nM dissociation constants, respectively, whereas the $K_d$ values for the Sp1(zf323)RD−GC(123) and −GC(323) complexes were 26.07 and 7.67 nM, respectively. These values are comparable with those in the previous report (*20*). Taking into account the fact that Sp1(zf323)RD contains an aspartic acid at position 2 in the $\alpha$ helix of finger 3A, the interactions by aspartic acid described here exist in the Sp1-(zf323)RD−GC(323) complex. On the contrary, the binding affinity of Sp1(zf323)RS for GC(123) (8.95 nM) was 3.5-fold higher than that for GC(323) (31.57 nM), and the order of affinity was quite in contradiction to that of Sp1(zf323)-RD. The results demonstrate that the mutation from aspartic acid to serine at position 2 in the finger 3A followed by the loss of such interactions has a significant effect on the DNA binding affinity of Sp1(zf323) derivatives. In analogy with the result for Sp1(zf323)RD, Sp1(zf123)KD had 2.7-fold higher affinity for GC(323) than did Sp1(zf123)KS ($K_d$ = 10.45 and 28.60 nM, respectively), indicating that the aspartic acid−C10′ interaction also exists in the Sp1(zf123)KD−GC-(323) complex. However, Sp1(zf123)KD bound to GC(123) with almost the same affinity as that for GC(323) ($K_d$ = 12.56 and 10.45 nM, respectively). The existence of the two roles of aspartic acid in each peptide is summarized in Table 2.

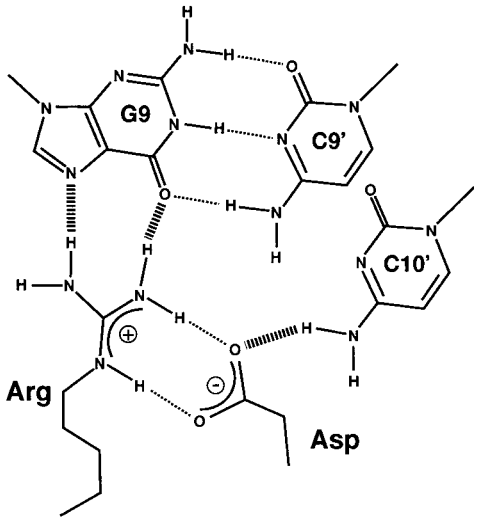


FIGURE 3: Possible hydrogen bonding network in the base recognition by the arginine and aspartic acid at positions −1 and 2, respectively. This structure deduced from the Zif268−DNA complex shows the base recognition by finger 3A of Sp1(zf323)-RD binding to GC(323) (*3*).

Table 1: Apparent Dissociation Constants ($K_d$) for Sp1(zf123)KS, Sp1(zf123)KD, Sp1(zf323)RD, and Sp1(zf323)RS Bindings to GC(123) and GC(323)

| binding site[b] | $K_d$ (nM)[a] | | | |
|---|---|---|---|---|
| | Spl(zf123)KS | Spl(zf123)KD | Spl(zf323)RD | Spl(zf323)RS |
| GC(123) | 13.96 ± 1.59 | 12.56 ± 1.36 | 26.07 ± 3.22 | 8.95 ± 1.00 |
| GC(323) | 28.60 ± 2.02 | 10.45 ± 1.76 | 7.67 ± 1.84 | 31.57 ± 2.48 |

[a] Apparent dissociation constants were determined by titration using a gel mobility shift assay as described in Materials and Methods. Values are averages of three or more independent determinations with standard deviations. [b] The nomenclature is described in the text (see Figure 1).

Table 2: Classification of the Roles of Aspartic Acid at Position 2 in the α Helix of the Position A Finger, and the Base Recognition Mode of the Finger in the Zinc Finger Peptides Used in This Study

| | Spl(zf123)KS | Spl(zf123)KD | Spl(zf323)RD | Spl(zf323)RS |
|---|---|---|---|---|
| Asp-C interaction in GC(323) | − | + | + | − |
| stabilization of Arg-G | − | − | + | − |
| | | Base Recognition Mode[a] | | |
| GC(123) | both strand | both strand | G strand | **both strand** |
| GC(323) | G strand | **C strand** | G strand | G strand |

[a] The base recognition mode is classified into three types. See the text for details.

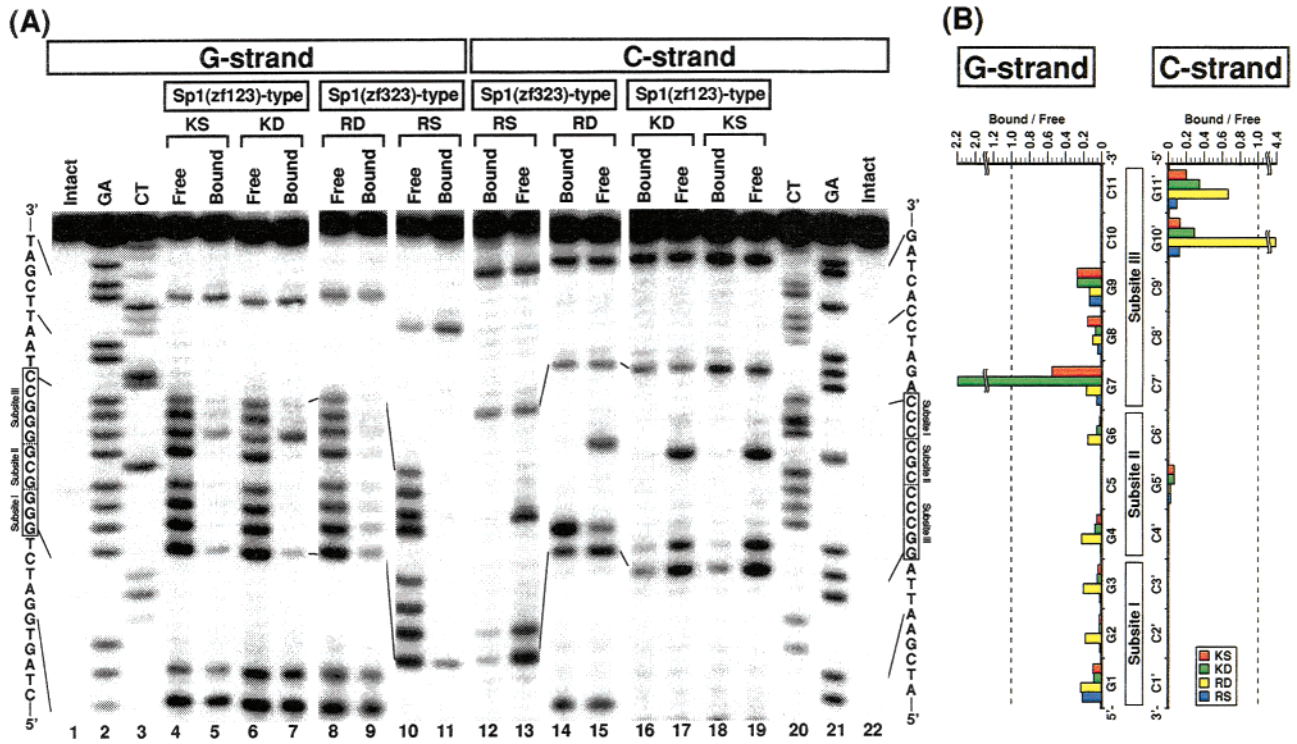


FIGURE 4: Methylation interference analyses for the binding of each zinc finger peptide of Sp1 to GC(123). (A) The results of autoradiograms from electrophoresis. The left (lanes 1−11) and right (lanes 12−22) panels show the results for the G and C strands, respectively. (Lanes 1 and 22) intact DNA; (lanes 2 and 21) G+A (Maxam−Gilbert reaction products); (lanes 3 and 20) C+T (Maxam−Gilbert reaction products); (lanes 4, 6, 8, 10, 12, 13, 15, 17, and 19) free DNA samples; (lanes 5, 7, 9, 11, 12, 14, 16, and 18) peptide-bound DNA samples. (B) A histogram showing the extent of methylation interference, which was calculated as the ratio of the cutting probabilities for the two bands (bound/free).

*Detection of Specific Base Recognition by Methylation Interference Assays.* Figure 4A shows the methylation interference pattern of each peptide for GC(123). The extent of the interference based on a densitometric analysis is presented by histograms (Figure 4B). As for the Sp1(zf123)-KS−GC(123) complex, all of the guanine bases were recognized, although the recognition level was weak only at G7, consistent with our previous results (*18*−*20*) (Figure 4A, lanes 4, 5, 18, and 19). By considering the guanine recognition at subsite III, the recognition mode is defined as the both-strand-type. The interference patterns of Sp1-(zf123)KD and Sp1(zf323)RS were almost the same as that of Sp1(zf123)KS except for the extent of recognition of G7; Sp1(zf323)RS recognized G7 strongly, whereas no recognition of G7 was observed in the Sp1(zf123)KD−GC(123) complex (Figure 4A, lanes 6, 7, 10−13, 16, and 17). In the case of Sp1(zf323)RD, all of the guanines were uniformly recognized in the G strand. However, the recognition levels for G10′ and G11′ were zero and moderate, respectively, suggesting that the characteristic of the base recognition mode is the G-strand-type, because the guanine bases at subsite III are mainly recognized in the G strand. (Figure 4A, lanes 8, 9, 14, and 15). Therefore, the base recognition modes of the present peptides at subsite III in the binding to GC(123) can be classified into two categories: both-strand-type including Sp1(zf123)KS, Sp1(zf123)KD, and Sp1-(zf323)RS, and G-strand-type including Sp1(zf323)RD (Table 2).

Figure 5 demonstrates the results of the methylation interference assay for the binding of each peptide to GC-(323), which has the guanine at position 10 in the G strand instead of the cytosine in GC(123). Sp1(zf123)KS made contact with all of the guanines at positions 1−11, despite the modest recognition of G10, parallel to that in the binding to GC(123) except for the recognition levels of guanines at positions 7 and 10 (Figure 5A, lanes 4, 5, 18, and 19). In the binding of Sp1(zf323)RD, the guanines at positions 2−9 were strongly recognized, whereas medium recognition was observed at G1 and G11′ (*20*) (Figure 5A, lanes 8, 9, 14, and 15). The peptide made no contact with G10. The pattern for Sp1(zf323)RS revealed the same tendency as that for Sp1-(zf323)RD, although the recognition of the guanines at

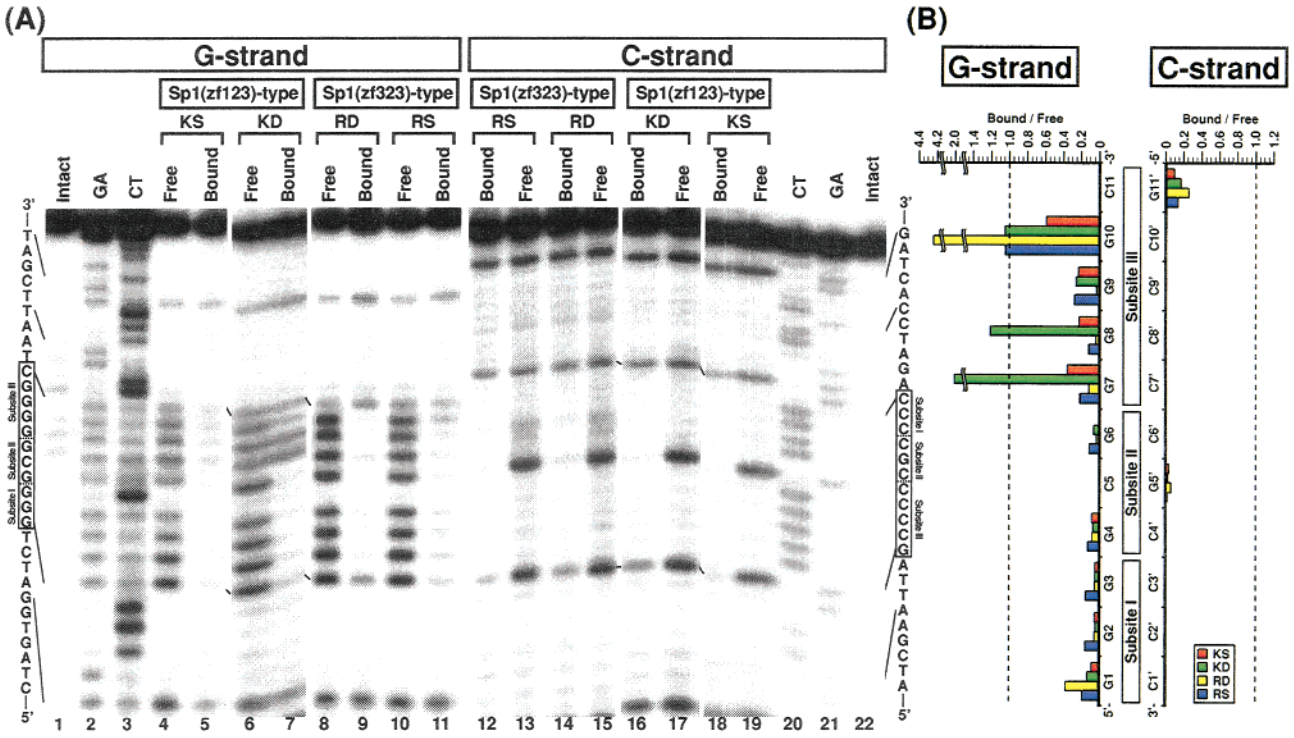


FIGURE 5: Methylation interference analyses for the binding of each zinc finger peptide of Sp1 to GC(323). (A) The results of autoradiograms from electrophoresis. The left (lanes 1−11) and right (lanes 12−22) panels show the results for the G and C strands, respectively. (Lanes 1 and 22) intact DNA; (lanes 2 and 21) G+A (Maxam−Gilbert reaction products); (lanes 3 and 28) C+T (Maxam−Gilbert reaction products); (lanes 4, 6, 8, 10, 12, 13, 15, 17, and 19) free DNA samples; (lanes 5, 7, 9, 11, 12, 14, 16, and 18) peptide-bound DNA samples. (B) A histogram showing the extent of methylation interference, which was calculated as the ratio of the cutting probabilities for the two bands (bound/free).

positions 2−8 for Sp1(zf323)RS was weakened (Figure 5A, lanes 10−13). The event of base recognition at subsite III of these peptides occurred predominantly in the G strand of GC(323). Accordingly, their base recognition modes are characterized as the G-strand-type. Of special interest is the result in which Sp1(zf123)KD showed an unexpected interference pattern at subsite III (Figure 5A, lanes 6, 7, 16, and 17). In this pattern, G9 and G11′ were evidently recognized, whereas no recognition was detected at G7, G8, and G10, indicating that the finger 1A of Sp1(zf123)KD has a unique base recognition mode without precedent. Together with the results of the estimation of $K_d$ values showing the presence of an aspartic acid−C10′ interaction in the Sp1-(zf123)KD−GC(323) complex, Sp1(zf123)KD primarily interacts with the bases in the C strand. This characteristic base recognition mode was not observed in the binding of the Sp1(zf123)KD derivative with lysine, aspartic acid, and arginine at α-helical positions −1, 2, and 6 in the finger 1A, respectively (data not shown). Consequently, the base recognition modes of the peptide binding to GC(323) were divided into two classes, G-strand- and C-strand-types (Table 2).

## DISCUSSION

From the previous mutational analysis, the finger 1A of Sp1(zf123)KS has a base recognition mode distinct from the typical base recognition pattern as observed in the Zif268−DNA complex (*18*). In such a recognition system, only the lysine residue at position −1 recognizes two guanines, G8 and G9, in the G strand (Figure 1A). On the other hand, the recent X-ray structural analysis of a variant zinc finger−

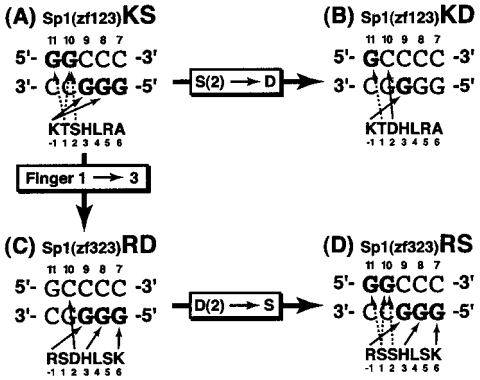


FIGURE 6: Proposed base recognition modes of the finger at position A of each peptide used in this study. The substrate bound by each peptide with high affinity is indicated. Arrows show the amino acid−base contacts assumed by the previous analyses of the DNA binding of Zif268 and Sp1 (*1*, *3*, *18*). Dotted arrows depict the putative amino acid−base interaction speculated by the structural analyses of the designed zinc finger−DNA complexes (*27*, *28*). The guanine bases, whose methylation interferes with the zinc finger binding, are in boldfaced print.

TATA box complex demonstrates that the amino acid at position 1 in the α helix recognizes the middle and 3′-side bases in the subsite for the preceding finger in the secondary strand (*28*). In addition, the 3′-side base can also be recognized by the serine at position 2 in the α helix of the designed zinc finger (*27*). On the basis of this evidence, the putative base recognition mode in Sp1(zf123)KS−GC(123) complex is proposed in Figure 6A.

*Effect of Aspartic Acid at Position 2 in Finger 1A on Base Recognition Mode and DNA Binding Affinity of Sp1(zf123)-KD.* The comparison of $K_d$ values revealed that an aspartic

acid−C10′ interaction is evidently formed in the Sp1(zf123)-KD−GC(323) complex. In addition, the decrease in the guanine recognition at subsite III was detected only in the binding to GC(323). However, the dissociation constant of the complex is comparable to those of Sp1(zf123)KD− and Sp1(zf123)KS−GC(123) complexes. Together with these results, it is suggested that the aspartic acid−C10′ interaction compensates for the decrease in the DNA binding affinity derived from the loss of the guanine contacts in Sp1(zf123)-KD−GC(323) complex. On the basis of the information on the base recognition of Sp1(zf123)KS, the putative base recognition mode of the finger 1A of Sp1(zf123)KD is proposed in Figure 6B. In this model, effective base recognition probably occurs exclusively by lysine, threonine, and aspartic acid at positions −1, 1, and 2, respectively. No zinc fingers with the base recognition mode of this type have been selected from the zinc finger library displayed on the phage surface. Previously, Choo and Klug suggested that, in the absence of arginine at position −1, the aspartic acid at position 2 is insufficient to contribute to the formation of the stable protein−DNA complex with any other contacts, because aspartic acid rarely occupies position 2 without arginine at position −1 in almost all of the selected zinc fingers (*5*). This proposition is inconsistent with our result in which the Sp1(zf123)KD−GC(323) complex is more stable than the Sp1(zf123)KS−GC(123) complex. Moreover, Elrod-Erickson and Pabo examined the DNA binding affinity and specificity of various mutants of Zif268 (*29*), and they reported that the aspartic acid residue at position 2 does not significantly contribute to the DNA binding affinity by comparing the $K_d$ value of the mutant peptide D20A with that of the wild-type. This is also not in agreement with our results. Presumably, in their experimental system, the DNA binding affinity of the mutant peptide is so high that the additional base contact by aspartic acid does not cause a further increase in the DNA binding affinity. Therefore, our study clearly indicates that an aspartic acid−cytosine interaction independently has a significant effect on both the DNA binding affinity and the base recognition mode, even if there is no arginine at position −1.

*Effects of Serine at Position 2 in Finger 3A on Base Recognition Mode and DNA Binding Affinity of Sp1(zf323)-RS.* Sp1(zf323) derivatives are unable to form aspartic acid (position 2)−C10′ interaction in the binding to GC(123), because GC(123) contains no C10′. Therefore, it is expected that no difference is detected in the guanine recognition mode or DNA binding affinity between the bindings of these peptides to GC(123). However, we obtained unpredictable results. It is an interesting fact that Sp1(zf323)RS binds to GC(123) in a manner similar to that of Sp1(zf123)KS with stronger guanine recognition at subsite III and higher DNA binding affinity than those of Sp1(zf123)KS. The amino acid at position 2 in the α helix of finger 3A is serine in Sp1-(zf323)RS, suggesting that G10′ and G11′ can be recognized by the serine at position 1 as described previously. In Sp1-(zf323)RS, the arginine−aspartic acid interaction is lost, and the G9 recognition by arginine is not stabilized by serine. Probably, the disappearance of such interaction may also enable the arginine to recognize G9 with some flexibility. Because of the flexibility, the G9 recognition by the long side chain of arginine and the G10′ and G11′ contacts with the short side chain of serine might be compatible. On the basis of the base recognition of Sp1(zf123)KS, the base recognition pattern of Sp1(zf323)RS binding to GC(123) is postulated in Figure 6D. Kim and Berg assumed that the base recognition mediated by the serine at position 2 is unlikely to contribute significantly to specificity (*27*). Nevertheless, the present results clearly demonstrate that not only aspartic acid but also serine evidently contributes to the base specificity.

*DNA Recognition Code by Zinc Finger Derived from Combination of Amino Acid Residues at Positions −1 and 2.* The base recognition patterns of the position A finger of the present peptides are summarized in Figure 6, in which the substrate bound by each peptide with higher affinity is shown. The incorporation of finger 3 into position A of Sp1-(zf123)KS changes the base recognition mode of the finger at position A from both-strand-type to G-strand-type (Figure 6C). In such a mutant, the extra mutation of the aspartic acid to serine at position 2 results in the restoration of the base recognition mode to both-strand-type (Figure 6D). In the Sp1-(zf123)KS, interestingly, single point mutation from serine to aspartic acid induces the drastic change in the base recognition mode. Sp1(zf123)KD, which has an aspartic acid at position 2 and lacks arginine at key positions, predominantly makes contact with the bases at subsite III in the C strand of GC(323) (Figure 6B). Two rules in the base recognition of zinc finger are proposed: (1) the zinc finger with an arginine and a serine at positions −1 and 2, respectively, recognizes the bases in the manner characterized as both-strand-type, and (2) in the absence of arginine at position −1 or 6, the zinc finger with an aspartic acid at position 2 exhibits a novel base recognition mode, namely, a predominant base contact in the C strand (C-strand-type). The second rule is most remarkable, because the phage display method selects no zinc fingers that mainly recognize the bases in the secondary strand. Therefore, the combination of this type leads to the design of a new zinc finger recognizing the base in the secondary strand. Hitherto, the creation of zinc fingers that recognize the cytosine-rich sequence in the primary strand was unsuccessful by any strategy. By the incorporation of an amino acid residue bearing a favorable contact with the guanine base into position 2 of zinc fingers classified into recognition code (*2*), therefore, this is of great promise for the creation of such a novel zinc finger.

## CONCLUSIONS

We demonstrate that both aspartic acid and serine at position 2 in the α helix play a significant role in DNA binding affinity and base specificity. In particular, the base specificity is closely related to the combination of the amino acid at key positions in the α helix. Therefore, a comprehensive survey and understanding of the combination are essential for the design of a new zinc finger. Thus far, phage display and structure-based methods have generally been employed for the creation of the zinc fingers with novel base specificity. However, sequences which cannot be recognized by the zinc finger have remained. For example, the zinc fingers for 5′-CCC-3′ and for the sequence with alternate AT- and GC-rich subsites have never been created. To overcome this problem, the laborious but steady strategy presented here is significant. It is also important to connect rationally various zinc fingers by considering the effect of

the adjacent zinc finger on DNA recognition. In this report, we successfully created zinc fingers with a novel base recognition mode. Sp1(zf123)KD−GC(323) and Sp1(zf323)-RS−GC(123) complexes are especially attractive, because their base recognitions in the secondary strand contribute significantly to both the DNA binding affinity and the base specificity. Consequently, the present results would provide an expansion of the DNA recognition code by the zinc finger which is useful for the design of a zinc finger for any given DNA sequences.

## REFERENCES


1. Pavletich, N. P., and Pabo, C. O. (1991) *Science 252*, 809−817.
2. Fairall, L., Schwabe, J. W. R., Chapman, L., Finch, J. T., and Rhodes, D. (1993) *Nature 366*, 483−487.
3. Elrod-Erickson, M., Rould, M. A., Nekludova, L., and Pabo, C. O. (1996) *Structure 4*, 1171−1180.
4. Reber, E. J., and Pabo, C. O. (1994) *Science 263*, 671−673.
5. Choo, Y., and Klug, A. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 11163−11167.
6. Choo, Y., and Klug, A. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 11168−11172.
7. Jamieson, A. C., Kim, S.-H., and Wells, J. A. (1994) *Biochemistry 33*, 5689−5695.
8. Desjarlais, J. R., and Berg, J. M. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 7345−7349.
9. Desjarlais, J. R., and Berg, J. M. (1992) *Proteins 12*, 101−104.
10. Desjarlais, J. R., and Berg, J. M. (1992) *Proteins 13*, 272.
11. Desjarlais, J. R., and Berg, J. M. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 2256−2260.
12. Shi, Y., and Berg, J. M. (1995) *Chem. Biol. 2*, 83−89.
13. Pabo, C. O., Peisach, E., and Grant, R. A. (2001) *Annu. Rev. Biochem. 70*, 313−340.
14. Isalan, M., Choo, Y., and Klug, A. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 5617−5621.
15. Isalan, M., Klug, A., and Choo, Y. (1998) *Biochemistry 37*, 12026−12033.
16. Dynan, W. S., and Tjian, R. (1983) *Cell 32*, 669−680.
17. Kadonaga, J. T., Jones, K. A., and Tjian, R. (1986) *Trends Biochem. Sci. 11*, 20−23.
18. Yokono, M., Saegusa, N., Matsushita, K., and Sugiura, Y. (1998) *Biochemistry 37*, 6824−6832.
19. Nagaoka, M., and Sugiura, Y. (1996) *Biochemistry 35*, 8761−8768.
20. Uno, Y., Matsushita, K., Nagaoka, M., and Sugiura, Y. (2001) *Biochemistry 40*, 1787−1795.
21. Wissmann, A., and Hillen, W. (1991) *Methods Enzymol. 208*, 365−379.
22. Kuwahara, J., and Coleman, J. E. (1990) *Biochemistry 29*, 8627−8631.
23. Frankel, A. D., Berg, J. M., and Pabo, C. O. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4841−4845.
24. Hori, Y., Suzuki, K., Okuno, Y., Nagaoka, M., Futaki, S., and Sugiura, Y. (2000) *J. Am. Chem. Soc. 122*, 7648−7653.
25. Jacobs, G. H. (1992) *EMBO J. 11*, 4507−4517.
26. Kim, C. A., and Berg, J. M. (1995) *J. Mol. Biol. 252*, 1−5.
27. Kim, C. A., and Berg, J. M. (1996) *Nat. Struct. Biol. 3*, 940−945.
28. Wolfe, S. A., Grant, R. A., Elrod-Erickson, M., and Pabo, C. O. (2001) *Structure 9*, 717−723.
29. Elrod-Erickson, M., and Pabo, C. O. (1999) *J. Biol. Chem. 274*, 19281−19285.


BI0256429