A direct comparison of the properties of natural and designed zinc-finger proteins

Yigong Shi and Jeremy M Berg*

Department of Biophysics and Biophysical Chemistry, The Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

Background: Zinc-finger proteins of the Cys_2His_2 type constitute an important family of DNA-binding proteins. Each zinc-finger domain has three residues that are thought to be important in determining DNA binding site specificity. Proteins have been designed previously by combining zinc-finger domains with a fixed sequence framework with different DNA-contacting residues.

Results: We compared the DNA-binding properties of the DNA-binding domain from the human transcription factor Sp1, which contains three zinc fingers, with designed proteins in which the sequences of the structural framework were greatly modified but the presumed DNA-contacting residues were retained. Frameworks based on a zinc-finger consensus sequence and on a minimalist sequence consisting largely of alanine residues were studied. The preference for binding to the target sequence, 5'-(G,T)GG G(C,A)G GG(G,T)-3', was retained in all cases tested. The consensus framework-based protein was found to be superior to the natural one in terms of overall DNA-binding affinity, the degree of sequence discrimination, and the resistance to inactivation by chelating agents.

Conclusions: Our observations provide direct evidence that the residues previously observed to interact with the DNA bases are indeed the most important residues for determining DNA-binding specificity. We have also shown that these domains can tolerate considerable sequence variation while retaining function as well as three-dimensional structure. Finally, they show that framework modification can be used to generate proteins that have normal or enhanced DNA-binding activity but have different metal-binding properties.

Chemistry & Biology February 1995, **2**:83–89

Key words: consensus framework, DNA-binding specificity, rational design, Sp1, zinc-finger proteins

Introduction

The human transcription factor Sp1 belongs to a subclass of the family of proteins that contain Cys_2His_2 zinc fingers, which recognize guanine-rich binding sites [1]. The crystal structure of the DNA-binding domain of one member of this class, Zif268, bound to its DNA-binding site has been determined [2]. The DNA-binding domain contains three zinc fingers, and the structure of the complex revealed that each zinc-finger domain interacts with a three-base-pair subsite. Most of the sequence-specific contacts are mediated by three amino-acid residues in each domain (hereafter referred to as the contact residues). The contact residues are also believed to be the most important factor in determining the DNA-binding site preference for other members of this protein family, a notion which is supported by experimental [3–5] and statistical [6] evidence. This has led to speculation that simple rules relating the identities of the contact residues and the sequences of the preferred DNA-binding sites may exist for this class of zinc-finger protein.

Comparison of the contact residues between Zif268 and Sp1 allows almost complete rationalization of the DNA-binding specificity of the three zinc-finger domains from Sp1 [1]. If the contact residues are, indeed, the only determinants of specificity, then it should be possible to alter the remaining residues within zinc-finger domains without significantly affecting DNA-binding specificity. We have previously reported the use of a consensus-sequence framework for the rational design of zinc-finger proteins with preselected DNA-binding characteristics [7]. Here, we report a direct comparison between a naturally-occurring DNA-binding domain and a consensus-sequence-based protein with the same contact residues. In addition, we explore the use of a minimalist zinc-finger [8] framework consisting of polyalanine with only the metal-binding, conserved hydrophobic, and contact residues preserved.

The consensus sequence zinc-finger peptide CP-1 was designed by selecting the amino acid that occurred most frequently at each position from a database of 131 zinc-finger domain sequences [9]. Experimental studies confirmed that CP-1 metal complexes adopt three-dimensional structures that are quite similar to those observed for peptides based on naturally-occurring zinc-finger sequences [9]. Additional studies indicated that CP-1 binds metal ions more tightly than other zinc-finger peptides that have been characterized [9]. As metal ion binding is required to induce folding for these peptides, the higher affinities for metal ion binding reflect more stable folding.

We have also designed and characterized a minimalist zinc-finger peptide [8]. The amino-acid sequence of this peptide, termed MZF, was LysTyrAla**Cys**AlaAla**Cys**AlaAlaAlaPheAlaAlaLysAlaAlaLeuAlaAla**His**AlaAlaAla**His**AlaLys. The metal-ligating residues (boldface) and the

*Corresponding author.

 ISSN 1074-5521

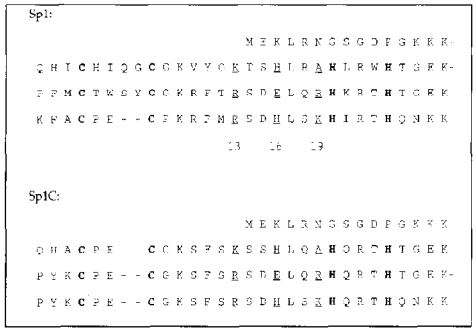

Fig. 1. The amino acid sequences of Sp1 and Sp1C, shown in single-letter code. Both proteins include the leader sequence MEKLRNGSGDPGKKK, as shown. The last nine of these amino acids occur in natural Sp1. The metal-chelating residues are in bold face. The presumed base-contacting residues (positions 13, 16, and 19) are underlined. In Sp1C, Asp is included in position 15 for domains 2 and 3 since Arg is present in position 13 in these domains. In domain 3, Ser is present in position 18. This deviation from the consensus sequence is believed to have no effect on DNA binding or other properties.

conserved hydrophobic residues (underlined) are included in this sequence, with the remainder of the residues being Ala except for three Lys residues which were included to increase water solubility. NMR and other studies revealed that this peptide binds metal ions with affinities in the same range as those for natural zinc-finger peptides and folds into the canonical zinc-finger structure [8].

We selected Sp1 as a model system to compare directly the properties of a natural zinc-finger protein with those of designed proteins. This choice was based on the extensive work on the DNA-binding properties of this protein and its DNA-binding domain [10,11] and our familiarity with this system [4,12].

Results and discussion

Design of consensus framework-based Sp1 and binding-site selection results

The sequence of the DNA-binding domain of Sp1 is shown in Fig. 1, together with that of a designed protein termed Sp1C. In Sp1C, the Sp1 framework has been replaced by sequences derived from a slightly modified consensus peptide beginning at the first metal-coordinating Cys residue on the first zinc-finger domain and ending at the last metal-coordinating His residue of the third zinc-finger domain. The contact residues in positions 13, 16, and 19 of each domain from the Sp1 sequence were retained. In addition, for the second and third domains an Asp residue in position 15 was retained, since an Arg residue

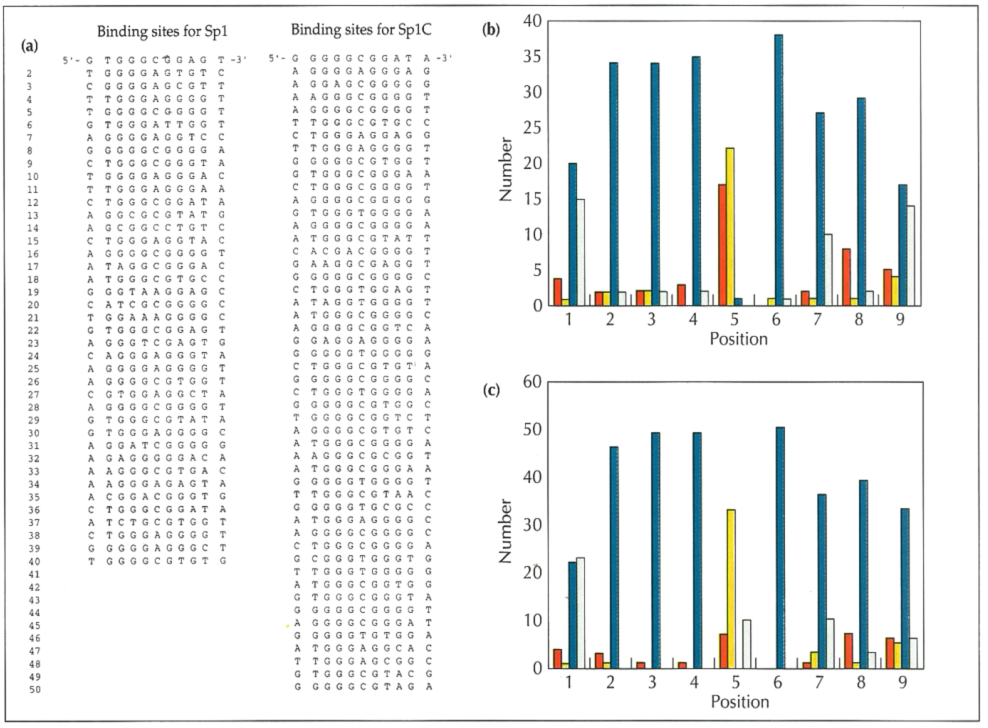

Fig. 2. Binding-site selection results for Sp1 and Sp1C. **(a)** Aligned DNA oligonucleotide sequences after the fifth round of selection. Only the presumed binding site (9 bp) and its flanking two bases are shown. **(b)** Binding-site histograms for Sp1 and **(c)** Sp1C. Red, A; yellow, C; blue, G; green, T.

is present in position 13 in these domains and Arg13 and Asp15 have been found to interact based on statistical [12] and structural observations [2]. The switch from the Sp1 framework to the consensus framework reflects a total of 25 point mutations and 4 residue deletions from Sp1.

The favored binding sites for Sp1 and Sp1C were determined using selection from pools of oligonucleotides containing a potential binding site of 11 randomized positions. After five rounds of binding-site selection, the shifted signal appeared to be saturated both for Sp1 and for Sp1C. More than 40 sequences from each cloned pool were determined, as shown in Fig. 2. The DNA-binding site preferences for both Sp1 and Sp1C were readily observed to be 5'-(T,G)GG G(C,A)G GG(G,T)-3'. This agrees well with reported Sp1 binding sites. No significant preferences were found for the flanking sequences, indicating that essentially all of the base-specific contacts were directed at these nine base pairs. The binding-site profiles of Sp1 and Sp1C resemble each other closely, implying the same set of interactions between DNA bases and side chains of the contact residues. The strong preference of both proteins for G in the seventh position of the binding site is striking in that the 'contact' residue for this base is position 19 in the first finger domain, which is an Ala residue in both cases. Many zinc-finger domains of this subclass have Arg in this position, which is known to contact this G directly [2]. In natural Sp1, the previous residue in the sequence is Arg, which might have been taken to indicate that an alternative Arg G contact can form. In Sp1C, however, this Arg is replaced with Gln and yet no significant difference in the binding site preference is observed, ruling out this explanation for the specificity for G at position 7.

Comparison of DNA-binding thermodynamics for Sp1 and Sp1C

To examine the thermodynamic basis of binding-site selection, we used individual binding assays to determine the dissociation constants for the two proteins bound to different DNA sites. As shown in Table 1, Sp1 and Sp1C bind to the sequence 5'-GGG GCG GGG-3' optimally with apparent affinities of 25 and 4 nM, respectively. Apparent dissociation constants were determined for a series of binding sites with single base changes from the optimal sequence (Fig. 3). The overall trends in discrimination are quite similar for Sp1 and Sp1C. However, Sp1C discriminates more strongly against the alternative binding sites than does natural Sp1. For the sites examined, the average improvement in discrimination is ~2-fold. Thus, natural Sp1 is not as selective as possible for its consensus binding site. Natural Sp1 acts as a general transcription factor and must therefore interact with a large number of different sites *in vivo* [10], perhaps explaining the lack of optimized binding specificity.

Table 1. Dissociation constants (K_d) of Sp1 and Sp1C DNA complexes.

Binding site	K_d for Sp1 (nM)	K_d for Sp1C (nM)
GGG GCG GGG	25 (1)	4 (0.3)
T GG GCG GGG	23 (1)	4 (0.4)
GAG GCG GGG	205 (31)	47 (5)
GGA GCG GGG	350 (18)	98 (6)
GGG ACG GGG	367 (65)	102 (17)
GGG GAG GGG	70 (3)	25 (2)
GGG GCA GGG	408 (29)	121 (8)
GGG GCG AGG	42 (6)	11 (1)
GGG GCG GTG	191 (21)	31 (6)
GGG GCG GGT	45 (2)	40 (3)

Estimated standard deviations are shown in parentheses.

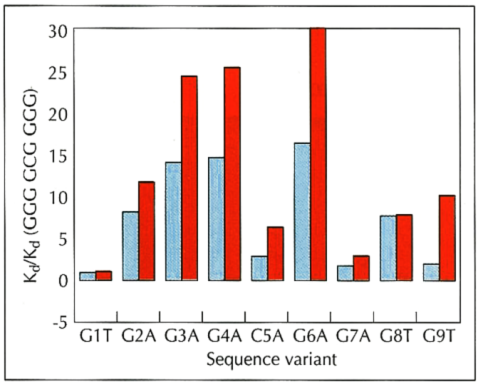

Fig. 3. Binding-site discrimination by Sp1 and Sp1C. The ratios of the dissociation constants for a series of binding sites with single base changes relative to the optimal site, 5'-GGG GCG GGG-3', are plotted for both proteins. Sp1, blue bars; Sp1C, red bars. These trends for the two proteins are quite similar, with Sp1C showing better discrimination at all positions.

Binding to the optimal site 5'-GGG GCG GGG-3' was also examined by DNaseI footprinting and methylation interference assays (Fig. 4). The results with the two proteins were quite similar. In both cases, methylation of any of the guanines in the binding sequence was found to inhibit binding. Interestingly, methylation of G in position 7 was found to have a slightly smaller inhibitory effect on binding than methylation at the other sites. This is consistent with the presence of Ala in the corresponding contact position as discussed above and with the relatively low (~2–3 fold) discrimination between G and A in this position (Table 1). The structural basis for this preference for G remains to be determined.

DNA-binding properties of proteins containing minimalist zinc-finger domains

To explore the minimum requirements for the zinc-finger framework necessary for maintaining specific DNA binding, we generated variants of Sp1C that had single domains replaced with domains based on a minimalist zinc finger. As shown in Fig. 5, a protein termed Sp1C-2M was prepared in which the middle zinc-finger domain of Sp1C was replaced with a minimalist one. This domain has 12 of the residues between the first Cys and the second His replaced with Ala; the only

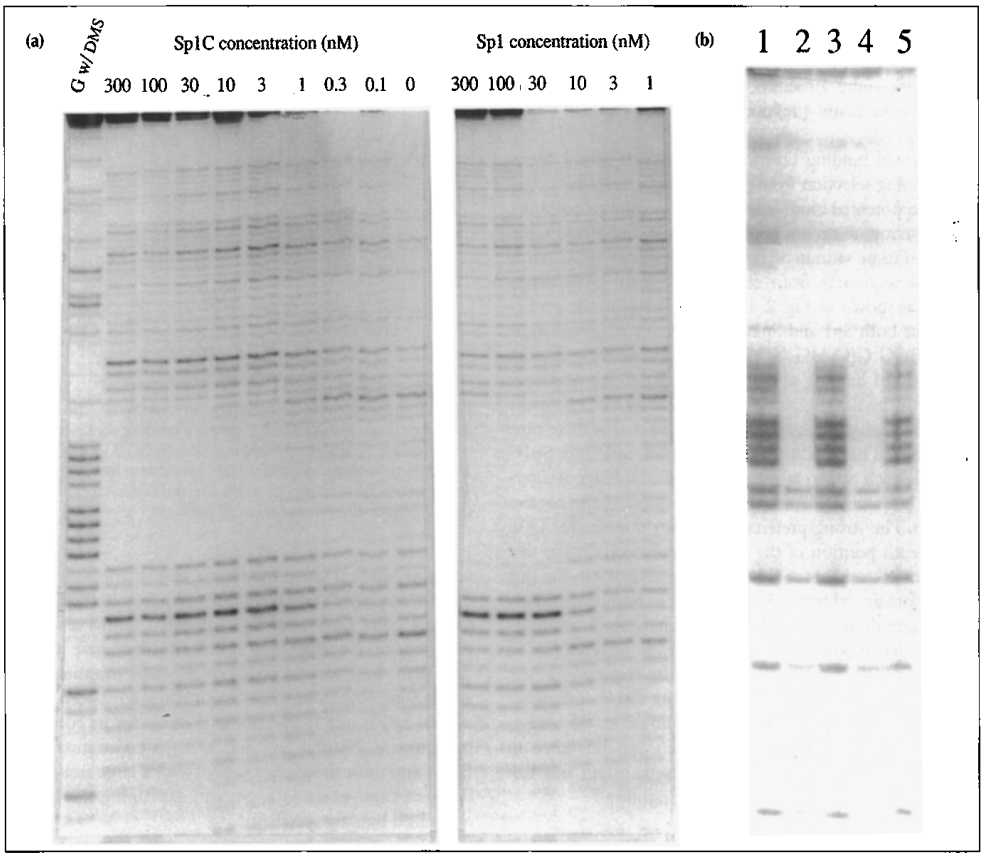

Fig. 4. Sequence-specific DNA binding properties of Sp1C and Sp1. **(a)** DNaseI footprint titrations. The first lane shows the guanine positions derived from reaction with dimethyl sulfate (DMS). **(b)** Methylation interference assay. Lane 1, DNA probe reacted with only DMS (control lane); lane 2, shifted band, 30 nM Sp1C; lane 3, unshifted band, 30 nM Sp1C; lane 4, shifted band, 150 nM Sp1; lane 5, unshifted band, 150 nM Sp1.

residues left fixed are the four metal-binding residues, the conserved hydrophobic residues (Phe, Leu), and the three presumed DNA-contacting residues (Arg, Glu, Arg). In addition, a second protein, Sp1C-2M(K9, D15), was created that replaced the Ala residue in position 9 with Lys, a relatively conserved residue [6] that makes a DNA backbone contact [2], and the Ala residue in position 15 with Asp, a residue that interacts with Arg in position 13 [2]. Finally, a protein termed Sp1C-3M, which has the carboxy-terminal zinc-finger domain replaced with a minimalist one, was also prepared.

Partial-binding-site selections, in which only the three bases presumed to interact with the minimalist finger domains were randomized, were performed. An oligonucleotide pool containing sequences of the form 5'-GGG NNN GGG-3' was employed for selection on Sp1C-2M and Sp1C-2M(K9, D15) whereas a pool containing sites of the form 5'-NNN GCG GGG-3' was used with Sp1C-3M. The results of the binding-site selection experiments are shown in

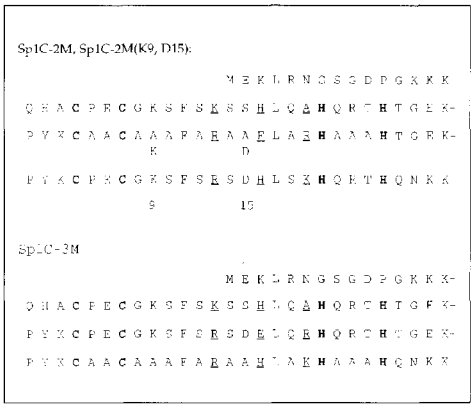

Fig. 5. The amino-acid sequences of proteins containing minimalist zinc-finger domains. Bold, metal-chelating residues; underlining denotes presumed base-contacting residues.

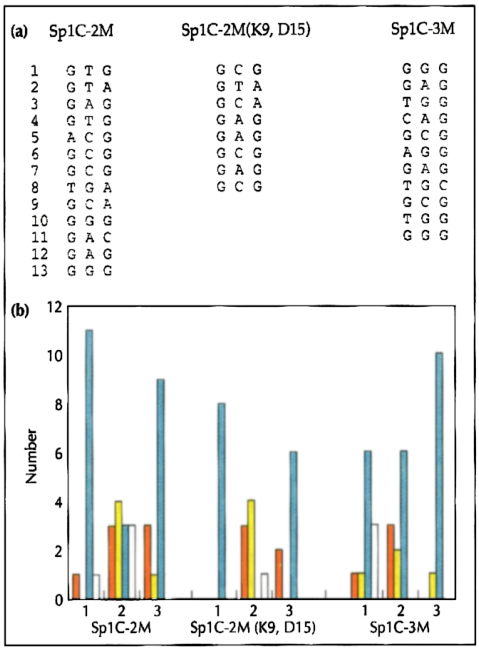

Fig. 6. Binding-site preferences for proteins containing minimalist domains. **(a)** Binding site tabulations and **(b)** histograms for Sp1C-2M and Sp1C-2M(K9, D15) and Sp1C-3M. Results for Sp1C-2M and Sp1C-2M(K9, D15) are on binding sites of the form 5'-CCA GGG NNN GGG TGG-3'. Those for Sp1C-3M are on binding sites of the form 5 - CCA NNN GCG GCG TGG-3 . Red, A; yellow, C; blue, G; green, T.

Fig. 6. For Sp1C-2M the subsite 5'-GNG-3' was found to be preferred. The G bases are presumably being contacted by two Arg residues in positions 13 and 19. However, there was no apparent preference for the central position which would be recognized via the Glu residue in position 16. Interestingly, Glu residues in corresponding positions in the Zif268 co-crystal structure were not observed to make specific hydrogen-bonding interactions with the DNA [2]. The inclusion of the two additional non-alanine residues in Sp1C-2M(K9, D15) led to selection of the subsite 5'-G(C,A)G-3'. This more closely matches the subsite recognized by Sp1 and Sp1C. Thus, these additional residues give increased sequence discrimination, presumably by increasing the intimacy of the interaction between the protein and the DNA. Binding site selections revealed that the third zinc-finger domain of Sp1C-3M prefers the subsite 5'-GGG-3', as do the corresponding domains in Sp1 and Sp1C (Fig. 6). All three minimalist finger-containing proteins were found to bind DNA in a sequence-dependent manner, as judged by the DNAseI footprinting results shown in Fig. 7.

Despite the maintenance of the same optimal binding sequences, the minimalist constructs display decreased affinities towards their optimal binding sites. Based on gel-shift titrations, Sp1C-2M binds to 5'-GGG GCG GGG 3' with 2.8 μM affinity, about 400-fold weaker than Sp1C. Sp1-2M(K9, D15) showed an affinity of 600 nM, only a three-fold increase over Sp1-2M. The affinity of Sp1C-3M for its optimal site was found to be 2 μM. Thus, other features of the natural and consensus framework must be responsible for their increased affinity for DNA compared with the minimalist domain-based proteins.

Stability of Sp1 and Sp1C to inactivation by chelating agents

Many zinc-finger proteins are known to be inactivated by treatment with zinc-chelating agents. The increased affinity of the single-domain consensus zinc finger peptide for metal ions compared with natural-sequence zinc-finger peptides [9] suggested that Sp1C would be less

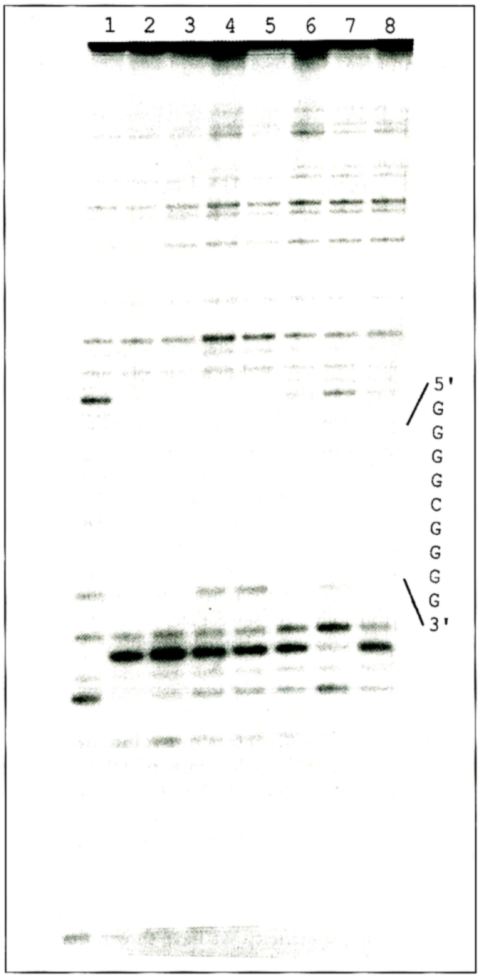

Fig. 7. DNaseI footprinting results of the Sp1 variants. Lane 1, no protein; lane 2, 500 nM Sp1; lane 3, 250 nM Sp1; lane 4, 100 nM Sp1C; lane 5, 50 nM Sp1C; lane 6, 5 μM Sp1C-3M; lane 7, 4 μM Sp1C-2M; lane 8, 2 μM Sp1C-2M(K9, D15).

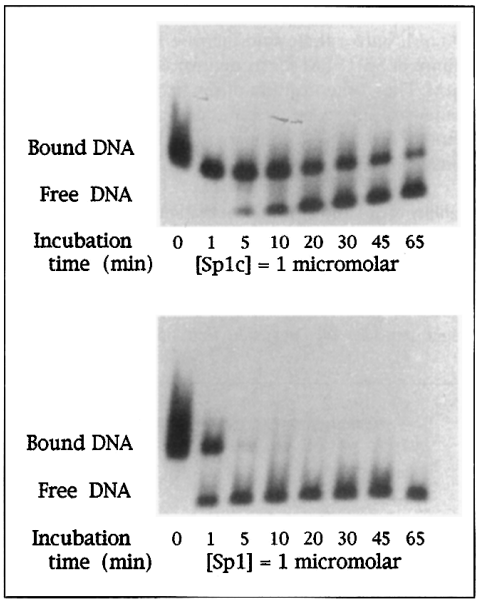

Fig. 8. Time-course of EDTA inactivation assays of Sp1 and Sp1C. Sp1C and Sp1 were treated with 50 mM EDTA for varying periods of time before incubation with DNA. DNA binding was analyzed by gel mobility shift assays.

susceptible to inactivation than its natural counterpart. As shown in Fig. 8, treatment of 1 μM Sp1 with 50 mM EDTA rapidly and completely abolished its DNA-binding activity. Similar treatment of Sp1C for up to 65 min still led to readily detectable protein–DNA complex formation as determined by gel mobility shift assays. Similar results were obtained when the EDTA treatment was performed after incubating the proteins and DNA together (data not shown). These results indicate that the consensus-sequence-based protein is both kinetically and thermodynamically less susceptible to inactivation by chelating agents than is the natural protein.

Significance

Many transcription factors use zinc fingers to recognize DNA. The selectivity of DNA binding by the zinc finger determines the specificity of the transcription factor for its binding site, and thus determines its biological effects. Previous mutagenesis and structural studies have suggested that three amino-acid residues in each domain contact DNA and are important in determining DNA-binding specificity. We have shown that these contact residues can be transferred from a natural protein, Sp1, onto completely different zinc-finger structural frameworks with retention of DNA-binding specificities. These results provide the most direct evidence to date that the contact residues determine the specificity of binding.

We also studied proteins containing three different minimal zinc-finger domains, in which essentially all of the amino-acid residues except for the metal-binding, conserved hydrophobic, and contact residues are replaced with Ala. Although these minimal domains are functional in that they bind to zinc and to DNA, they show decreased specificity and affinity for DNA binding. Thus, although the identity of the contact residues is the most important single factor in determining the DNA-binding characteristics of zinc-finger proteins, other elements in the framework regions also have some effect.

The designed protein Sp1C, which contains the consensus zinc-finger framework residues, binds more tightly and more specifically to a consensus target site than does Sp1, and also binds more tightly to zinc. Thus, the DNA-binding and metal ion binding characteristics of natural Sp1 are not optimized. This may be important for the ability of Sp1 to bind to a large number of different promoters. It should be possible in the future to alter both the metal-ion affinity of a protein and its DNA-binding properties in a rational manner.

Materials and methods

Strains, plasmids, and cloning

The bacterial strains used include *Escherichia coli* K12 strain 71-18 (New England Biolabs), B strain BL21 (DE3) pLys S [13], and *Epicurian coli* SURE cells (Stratagene). The bacterial plasmids used include pKK223-3 (Pharmacia), pEMBL [14] and pG5 [15]. The expression vectors pKK223-3 and pG5 contain a Tac promoter and T7 RNA polymerase promoter, respectively, both of which are inducible with isopropylthiogalactoside (IPTG). All synthetic DNA oligonucleotides used in the polymerase chain reaction (PCR) were synthesized on an Applied Biosystems Model 392 synthesizer, and purified from oligonucleotide purification cartridges (Applied Biosystems). Standard methods were used throughout this work for the isolation and manipulation of plasmids and DNA fragments. Designed proteins were constructed using established PCR site-directed mutagenesis techniques [16] and verified through plasmid sequencing with Sequenase Version 2.0 (United States Biochemical). All restriction enzymes and buffers were purchased from New England Biolabs.

Construction and purification of Sp1 variants

A synthetic 87 base pair oligonucleotide harboring *NdeI*, *MluI*, *AgeI*, *XmaI* and *SalI* restriction sites in tandem was cloned into *NdeI*/*SalI* sites in pG5 to create a suitable polylinker. The DNA sequence between *NdeI* and *MluI* sites (5'-CATATGGAAAAA-CTGCGGAACGGATCCGGGGACCCTGGCAAAAAGAA-ACAGCACGCGT-3') encodes an 18 amino-acid leader sequence before the first zinc-coordinating Cys residue. The oligonucleotides coding for the first and third zinc fingers were ligated into *MluI*/*AgeI* sites and *XmaI*/*SalI* sites, respectively. The second finger was cloned into *AgeI*/*XmaI* sites to complete a three-finger construct. The genes encoding all Sp1 variants were created in this fashion. To minimize homologous recombination for the minimalist variants, *Epicurian coli* SURE cells were employed. The

Sp1-encoding fragment was expressed from pKK223-3. All of the proteins were expressed and purified to >95 % homogeneity as described previously [7]. The concentration of Sp1 was estimated from the absorbance at 280 nm based on a molar extinction coefficient of 22 400 M^{-1} cm^{-1}. The concentrations of the designed proteins were determined from the absorbance at 275 nm based on a molar extinction coefficient of 2840 M^{-1} cm^{-1}.

Binding-site selection assays

Binding sites were determined via gel-shift selection/PCR amplification procedures. For Sp1 and Sp1C, a 46-base DNA oligonucleotide containing *Eco*RI and *Bam*HI sites at either end, 5'-CTCGGTACCGAATTCCANNNNNNNNNNTAAG-CTTGGATCCTCTAG-3', was used. The ^{32}P-labeled DNA fragment was suspended in distilled H_2O and incubated with target protein in a 20 µl reaction volume with 25 mM Tris·Cl, pH 8, 10 % glycerol, 100 mM KCl, 100 µM $ZnCl_2$, 2 mM dithiothreitol (DTT), 2 µg ml^{-1} dI–dC and 50 µg ml^{-1} bovine serum albumin (BSA) for 30 min at 25 °C. The protein concentrations used for the first round of selection varied from 0.2 to 5 µM. Typically, a series of five-fold dilutions of protein was employed to ensure that any retarded bands were observed. The gel mobility shift assay was carried out in a 1.8 % Sea-Plaque low-melting agarose (FMC BioProducts) gel with Tris-glycine buffer (50 mM Tris-Cl, 400 mM glycine, pH 8.5) at 4 °C. The shifted band was excised and melted in 50 µl H_2O at 65 °C. 1 µl of the melted gel was used for PCR amplification. After five rounds of selection, the PCR product was gel-purified, digested by *Bam*HI/*Eco*RI and cloned into pEMBL. A sufficient number of clones were sequenced to deduce a well-defined consensus binding site. For Sp1C-2M, Sp1C-2M(K9, D15), and Sp1C-3M, the center 11-base randomized sequence was replaced with either 5'-GGGNNNGGG-3' or 5'-NNNGCGGGG-3' to confine the randomized site to the region contacted by the minimalist zinc-finger domain. Three rounds of selection were sufficient to determine the binding-site consensus.

Measurement of dissociation constants

Dissociation constants (K_d) for zinc-finger protein–DNA interaction were measured via gel mobility shift assays. All binding sites used were synthesized and subcloned into *Eco*RI/*Bam*HI sites in pEMBL. All binding sites used were ~80 base-pairs long. Plasmids containing binding sites were digested with *Eco*RI, ^{32}P-labeled, and further digested with *Bam*HI. The resulting DNA probes were purified with 3 % Nusieve agarose gels (FMC BioProducts). The concentrations of the DNA probes were estimated to be less than 100 pM. Pilot experiments were performed to estimate the approximate dissociation constant, then five concentrations of protein, centered around the K_d, were used for gel mobility shift assays. The radioactive signal was analyzed using a Phosphorimager (Molecular Dynamics) and the data were fit using Kaleidagraph (Abelbeck Software). For binding sites of the form 5'-(G,T)GG G(C,A)G GG(G,T), each tabulated K_d represents the average of five gel-shift experiments. For single point mutations of the site 5'-(T, G)GG GCG GGG-3', each tabulated K_d represents the average of two experiments.

DNaseI footprinting, methylation interference, and EDTA inactivation assays

DNaseI footprinting experiments were carried out as described previously [7]. Experiments were performed in 25 mM Tris-Cl, pH 8, 100 mM KCl, 2 mM DTT, 1 mM $MgCl_2$, 5 mM $CaCl_2$, 100 µM $ZnCl_2$ with 2 µg ml^{-1} dI–dC and 50 µg ml^{-1} BSA. For methylation interference experiments, 100 000 cpm aliquots of DNA probe were incubated with 1 µl of dimethylsulfate (DMS) in 200 µl of reaction buffer (50 mM sodium cacodylate, 1 mM EDTA, pH 8) at 25 °C for 3 min. The reactions were stopped by addition of 50 µl DMS stop buffer (1.5 M sodium acetate, 1 M β-mercaptoethanol, pH 7). Using these probes, the methylation interference experiments were performed as described [17]. For the EDTA inactivation assays, 1 µM solutions of proteins were incubated with 50 mM EDTA for various amounts of time under gel mobility shift buffer conditions. The ^{32}P-labeled DNA probe was then added and the mixture was loaded into a running gel 30 s later.

Acknowledgements: We thank the National Institutes of Health for support of this work.

References

1. Berg, J.M. (1992). Sp1 and the subfamily of zinc finger proteins with guanine-rich binding sites. *Proc. Natl. Acad. Sci. USA* **89**, 11109–11110.
2. Pavletich, N.P. & Pabo, C.O. (1991). Zinc finger–DNA recognition: crystal structure of a Zif268–DNA complex at 2.1 Å. *Science* **252**, 809–817.
3. Nardelli, J., Gibson, T.J., Vesque, C. & Charnay, P. (1991). Base sequence discrimination by zinc-finger DNA-binding domains. *Nature* **349**, 175–178.
4. Desjarlais, J.R. & Berg, J.M. (1992). Towards rules relating zinc finger protein sequences and DNA binding site preferences. *Proc. Natl. Acad. Sci. USA* **89**, 7345–7349.
5. Fairall, L., Schwabe, J.W.R., Chapman, L., Finch, J.T. & Rhodes, D. (1993). The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. *Nature* **366**, 483–487.
6. Jacobs, G.H. (1992). Determination of the base recognition position of zinc fingers from sequence analysis. *EMBO J.* **11**, 4507–4517.
7. Desjarlais, J.R. & Berg, J.M. (1993). Use of a zinc-finger consensus sequence framework and specificity rules to design specific DNA binding proteins. *Proc. Natl. Acad. Sci. USA* **90**, 2256–2260.
8. Michael, S.F., Kilfoil, V.J., Schmidt, M.H., Amann, B.T. & Berg, J.M. (1992). Metal binding and folding properties of a minimalist Cys_2His_2 zinc finger peptide. *Proc. Natl. Acad. Sci. USA* **89**, 4796–4800.
9. Krizek, B.A., Amann, B.T., Kilfoil, V.J., Merkle, D.L. & Berg, J.M. (1991). A consensus zinc finger peptide: design, high-affinity metal binding, a pH-dependent structure, and a His to Cys sequence variant. *J. Amer. Chem. Soc.* **113**, 4518–4523.
10. Kadonaga, J.T., Jones, K. & Tjian, R. (1986). Promotor-specific activation of RNA polymerase II transcription by Sp1. *Trends Biochem. Sci.* **11**, 20–23.
11. Kriwacki, R.W., Schultz, S.C., Steitz, T.A. & Caradonna, J.P. (1992). Sequence-specific recognition of DNA by zinc-finger peptides derived from the transcription factor Sp1. *Proc. Natl. Acad. Sci. USA* **89**, 9759–9763.
12. Desjarlais, J.R. & Berg, J.M. (1992). Redesigning the DNA-binding specificity of a zinc finger protein: a data base guided approach. *Proteins* **12**, 101–104.
13. Studier, F.W., Rosenberg, A.H., Dunn, J.J. & Dubendorff, J.W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**, 60–89.
14. Dente, L., Cesareni, G. & Cortese, R. (1983). pEMBL: a new family of single stranded plasmids. *Nucleic Acids Res.* **11**, 1645–1655.
15. Alexander, P., Fahnestock, S., Lee, T., Orban, J. & Bryan, P. (1992). Thermodynamic analysis of the folding of streptococcal protein G IgG-binding domains B1 and B2: why small proteins tend to have high denaturation temperatures. *Biochemistry* **31**, 3597–3603.
16. Nelson, R. & Long, G.L. (1989). A general method of site-specific mutagenesis using amplification of the *Thermus aquaticus* polymerase chain reaction. *Anal. Biochem.* **180**, 147–151.
17. Baldwin, A.S., Jr. (1989). Methylation interference assay for analysis of DNA–protein interactions. In *Current Protocols in Molecular Biology*. (Ausubel, F.M., *et al.*, eds), pp. 12.3.1–12.3.6, John Wiley and Sons, New York.

Received: **10 Jan 1995**; revisions requested: **26 Jan 1995**; revisions received: **27 Jan 1995**. Accepted: **27 Jan 1995**.