NMR Structure of Transcription Factor Sp1 DNA Binding Domain[†,‡]

Shinichiro Oka,[§] Yasuhisa Shiraishi,[||] Takuya Yoshida,[§] Tadayasu Ohkubo,[§] Yukio Sugiura,*[,||] and Yuji Kobayashi*[,§]

Graduate School of Pharmaceutical Sciences, Osaka University, 1-6 Yamadaoka, Suita, Osaka 565-0871, Japan, and Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan

Received July 22, 2004; Revised Manuscript Received October 10, 2004

ABSTRACT: To understand the DNA recognition mechanism of zinc finger motifs of transcription factor Sp1, we have determined the solution structure of DNA-binding domain of the Sp1 by solution NMR techniques. The DNA-binding domain of Sp1 consists of three Cys_2His_2-type zinc finger motifs. They have typical $\beta\beta\alpha$ zinc finger folds and relatively random orientations. From DNA-binding analysis performed by NMR and comparison between structures determined here and previously reported structures of other zinc fingers, it was assumed that DNA recognition modes of fingers 2 and 3 would be similar to those of fingers of Zif268, in which each finger recognizes four base pairs strictly by using residues at positions −1, 2, 3, and 6 of the recognition helix. On the contrary, finger 1 can use only two residues for DNA recognition, Lys550 and His553 at positions −1 and 3 of the helix, and has more relaxed sequence and site specificity than other Cys_2His_2 zinc fingers. It is proposed that this relaxed property of finger 1 allows transcription factor Sp1 to bind various DNA sequences with high affinity.

Human transcription factor Sp1, which was originally isolated from HeLa cells, is well-known to be a ubiquitous factor to activate transcription by RNA polymerase II (*1*, *2*). Sp1 contains a DNA binding domain consisting of three C_2H_2-type zinc fingers at C-terminal region (Figure 1), which bind to GC-rich recognition elements (GC box) within a large number of cellular and viral promoters (*3*, *4*). It has been proposed that the consensus sequence of Sp1 binding is 5′-(G/T)GGGCGG(G/A)(G/A)(C/T)-3′. The crystal structure of Zif268, which contains three tandem repeat of this type of zinc finger, bound to the substrate DNA presented the simple DNA recognition mode, in which each zinc finger recognizes three base pairs in one strand via key amino acid residues at helical positions −1, 3, and 6, and one base pair in the complementary strand interacts with the amino acid at helical position 2 (*5*, *6*). This canonical recognition mode has been observed in many C_2H_2-type zinc fingers (Figure 2), and various DNA recognition modes of C_2H_2-type zinc fingers have been elucidated on the basis of it (*7*−*13*). From the previous NMR study about fingers 2 and 3 of Sp1, both zinc fingers have the typical structure of a C_2H_2-type zinc finger, and also their recognition modes are similar to those of Zif268 zinc fingers (*14*).

However, the structure of finger 1 has never been established at the present time, although several unique features in the DNA recognition mode of finger 1 have been reported in previous biochemical experiments, as follows: (1) In contrast to fingers 2 and 3, finger 1 has a less conserved subsite in the GC box (*3*, *4*, *15*, *16*). (2) The contribution of finger 1 to the DNA binding affinity of Sp1 is smaller than that of fingers 2 and 3, but the presence of finger 1 is still essential for the high DNA binding affinity (nanomolar order in K_d) of Sp1 (*17*). (3) Finger 1 showed unique five Gua bases recognition in both strands in the case of the subsite 5′-GGGCC-3′ (*17*). These unique features have never been detected in other zinc fingers. Therefore, the determination of the DNA recognition mode of finger 1 not only leads to an understanding of ubiquitous factor Sp1 but also helps to expand the DNA recognition mode of a C_2H_2-type zinc finger.

In the past, we tried unsuccessfully to determine the structure of synthesized peptide corresponding to finger 1 of Sp1 by classical homonuclear NMR experiments. In this study, we finally clarified the NMR structure of the total DNA binding domain of Sp1 including finger 1 expressed in *Escherichia coli* and evaluated the changes of chemical shifts of their NMR signals in binding of the GC box. On the basis of these structural data, the DNA recognition mode of finger 1 has been proposed.

MATERIALS AND METHODS

Sample Preparation. A pEVSp1 vector that contains DNA fragment Sp1(530−623) was introduced into *E. coli* strain BL21(DE3)pLysS (*18*). Uniform $^{13}C/^{15}N$ labeling was achieved by use of CHL medium (Chlorella Industry). The cells were grown at 37 °C to $A_{660} = 0.6$, and protein expression was induced by addition of IPTG[1] and $ZnSO_4$ to final concentrations of 0.2 and 0.1 mM, respectively. After 12 h of induction, the cells were harvested by centrifugation. The

[†] This work is partially supported by the contracted research Protein 3000 Project, a Grant-in-Aid for COE Project Element Science (12CE2005), and a Grant-in-Aid for Scientific Research (13557210, 14370755) from the Ministry of Education, Culture, Sports, Science and Technology, Japan. Y.Shiraishi is a research fellow of the Japan Society for the Promotion.

[‡] The atomic coordinates for the 30 best conformers of each finger described in this paper have been deposited with the Brookhaven Protein Data Bank (accession codes 1VA1, 1VA2, and 1VA3).

* To whom correspondence should be addressed: (Y.K.) phone +81-6-6879-8220, fax +81-6-6879-8224, e-mail yujik@protein.osaka-u.ac.jp; (Y.Sugiura) phone +81-774-38-3210,; fax +81-774-32-3038, e-mail sugiura@scl.kyoto-u.ac.jp.

[§] Osaka University.

[||] Kyoto University.

10.1021/bi048438p CCC: $27.50
Published on Web 11/26/2004

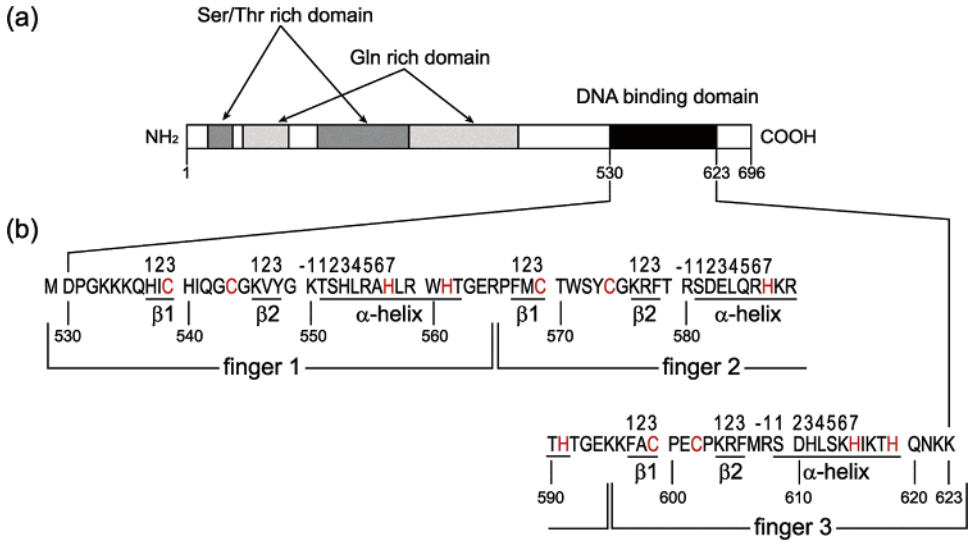

FIGURE 1: Summary of Sp1−696 structure. (a) Schematic representation of Sp1−696, which has the same transcriptional activity and DNA-binding activity as full-length Sp1(2). The Sp1−696 has two Ser/Thr-rich domains and two Gln-rich domains at the N-terminal region. These domains mediate transcriptional activity. The DNA-binding domain, called Sp1(530−623), is located at the C-terminal region. (b) Primary and secondary structure of Sp1(530−623), which was used in this study. Metal binding residues are shown in red. The N-terminal Met residue was added for expression.

	β1	β2	recognition helix	PDBID
	123	12 3	-11234567 . . .	
Sp1-f1	----MDPGKKKQHICHIQ-GCGKV---YGKTSHLRAHLR-WHTGE--			1VA1
Sp1-f2	----------RPFMCTWS-YCGKR---FTRSDELQRHKR-THTGE--			1VA2
Sp1-f3	----------KKFACPE---CPKR---FMRSDHLSKHIK-THQNKK-			1VA3
ZIF268-f1	--------MERPYACPVE-SCDRR---FSRSDELTRHIR-IHTGQ--			1AAY
ZIF268-f2	----------KPFQCRI---CMRN---FSRSDHLTTHIR-THTGE--			1AAY
ZIF268-f3	----------KPFACDI---CGRK---FARSDERKRHTK-IHLRQKD			1AAY
TFIIIA-f1	MGEKALPVVYKRYICSFA-DCGAA---YNKNWKLQAHLC-KHTGE--			1TF6
TFIIIA-f2	----------KPFPCKEE-GCEKG---FTSLHHLTRHSL-THTGE--			1TF6
TFIIIA-f3	----------KNFTCDSD-GCDLR---FTTKANMKKHFNRFHNIK--			1TF6
TFIIIA-f4	----------ICVYVCHFE-NCGKA---FKKHNQLKVHQF-SHTQQ--			1TF6
TFIIIA-f5	----------LPYECPHE-GCDKR---FSLPSRLKRHEK-VHAG---			1TF6
TFIIIA-f6	------------YPCKKDDSCSFV---GKTWTLYLKHVAECHQD---			1TF6
GLI1-f1	----------VYETDCRWD-GCSQE---FDSQEQLVHHINSEHIHG--			2GLI
GLI1-f2	--------ERKEFVCHWG-GCSRELRPFKAQYMLVVHMR-RHTGE--			2GLI
GLI1-f3	----------KPHKCTFE-GCRKS---YSRLENLKTHLR-SHTGE--			2GLI
GLI1-f4	----------KPYMCEHE-GCSKA---FSNASDRAKHQNRTHSNE--			2GLI
GLI1-f5	----------KPYVCKLP-GCTKR---YTDPSSLRKHVKTVHGPDA-			2GLI

FIGURE 2: Sequence alignments of structure-determined zinc fingers. Metal binding residues are shown in red. Well-conserved hydrophobic residues are outlined. The positions of the two β-strands and recognition helix are indicated with the position numbering within them.

harvested cells were resuspended in buffer A (8.0 mM Na_2HPO_4, 1.5 mM KH_2PO_4, 2.7 mM KCl, 137 mM NaCl, and 1 mM DTT) and disrupted by sonication. The lysate was centrifuged to remove insoluble debris. The supernatant was loaded on Econo-pac High S column (Bio-Rad Laboratories) equilibrated with buffer A and eluted by a gradient of NaCl. The eluted fractions containing Sp1(530−623) were dialyzed against buffer A and loaded on a Mono S HR 10/10 column (Amersham Bioscience) equilibrated with buffer A. The protein was eluted by a gradient of NaCl and further purified on a Superdex 75 column (Amersham Bioscience) equilibrated with buffer B (10 mM Tris-HCl pH 8.0, 50 mM NaCl, and 1 mM DTT). The eluted fractions containing Sp1(530−623) were concentrated by ultrafiltration (Centricon YM3, Millipore). Buffer exchange to NMR measurement buffer (10 mM Tris, pH 7.4, 50 mM NaCl, 1 mM DTT, and 10% D_2O) was achieved at the same time. Because we found that finger 1 peptide of Sp1 was not stable in acidic solution used in the previous NMR study on fingers 2 and 3 (*14*), we performed all NMR experiments at neutral pH. The homogeneity and the DNA-binding activity of the prepared sample were confirmed by SDS−PAGE and gel mobility-shift assay, respectively. Synthesized oligonucleotides 5′-AGGGGCGGGGCCT-3′ and its complementary sequences were annealed and used in binding studies.

Structure Determination. All NMR experiments were performed at 25 °C on a Varian INOVA 600 spectrometer equipped with triple-resonance *z*-gradient probe. ^{15}N-HSQC, HNCACB, C(CO)NH-TOCSY, and HNCO spectra were acquired for sequential backbone assignments. ^{13}C-ctHSQC, ^{15}N-separated TOCSY, H(CCO)NH-TOCSY, and HCCH-COSY spectra were acquired for assignments of aliphatic side chains. ^{13}C-ctHSQC, ^{13}C-HSQC-TOCSY, ^{13}C-separated NOESY, and ^{15}N-separated NOESY spectra for aromatic

[1] Abbreviations: G-strand, guanine-rich strand of GC box sequence; C-strand, cytosine-rich strand of GC box sequence; IPTG, isopropyl β-thiogalactoside; DTT, dithiothreitol; HSQC, heteronuclear single quantum coherence spectroscopy; COSY, correlated spectroscopy; TOCSY, total correlation spectroscopy; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser enhancement spectroscopy.

region and (Hβ)Cβ(CγCδ)Hδ spectrum were acquired for assignment of aromatic side chains. Experimental details of NMR spectroscopy are provided in the review (*19*). The processing and analysis of the collected data were carried out with the NMRPipe (*20*) and NMRView (Merck Research Laboratories) software packages. By using a series of 2D and 3D heteronuclear NMR experiments, almost complete ^{1}H, ^{13}C, and ^{15}N resonance assignments were achieved. A total of 811 NOE restraints collected from ^{13}C-edited NOESY and ^{15}N-edited NOESY and 87 dihedral angle restraints calculated from chemical shifts of HN, N, Hα, Cα, Cβ, and C′, by use of program TALOS (*21*), were used for structural calculations. NOEs were classified as strong, medium, weak, or very weak, corresponding to distance restraints of 1.8−2.8, 1.8−3.4, 1.8−4.2, or 1.8−5.0 Å, respectively. Because no interfinger NOE was observed (Figure 4), the orientations between fingers could not be determined (data not shown). Hence the structure calculation of each finger was performed individually. All structure calculations were performed by use of CNS solve 1.1 (*22*). Initial structure calculations were performed without zinc coordination restraints. Zinc-coordinating residues with their tetrahedral geometry were clearly identified from these structures. Thereafter, zinc coordination restraints were incorporated in structural calculations (*14*, *23*). After the final minimization step, 30 structures with lowest energy were used for further analysis.

NMR Study of DNA Complex. The protein was complexed with synthesized oligonucleotide in the above-mentioned NMR experiment conditions. NMR measurements were performed under the same conditions as well. Because Sp1(530−623)−DNA complex was not sufficiently stable for multidimensional experiments that require much time, for example, 3D X-filtered NOESY, only three experiments, ^{1}H−^{15}N HSQC, HNCA, and HN(CO)CA, were acquired with good S/N ratio. The backbone H−N resonances in the ^{1}H−^{15}N HSQC spectrum were 90% assigned by use of these spectra.

RESULTS

As shown in Figure 3, observed NOEs were concentrated inside each finger and no NOE between fingers was detected. Therefore, the relative orientations of fingers and conformations of linker regions between fingers were not converged in the structure calculation of the entire DNA-binding domain. This result indicates that the orientation of any linker residue is not defined in DNA-free state. Meanwhile, the ensembles of structures converged well individually in each finger (Figure 4). The resulting structure of each finger shows a typical globular folding consisting of an α helix and an antiparallel β sheet, which is commonly observed in classical Cys_2His_2 zinc fingers. These secondary structure elements pack with each other through coordination of zinc and hydrophobic interactions. In each finger, zinc-coordinating residues were determined experimentally by NOE connectivities, which were consistent with those predicted from the amino acid sequence. The structure statistics are summarized in Table 1.

Structure of Finger 1. A short antiparallel β sheet (His537−Cys539 and Lys546−Tyr548) is formed through two hydrogen bonds, His537(H_N)−Tyr548(O) and Tyr548(H_N)−His537(O). An α helix corresponding to Thr551−

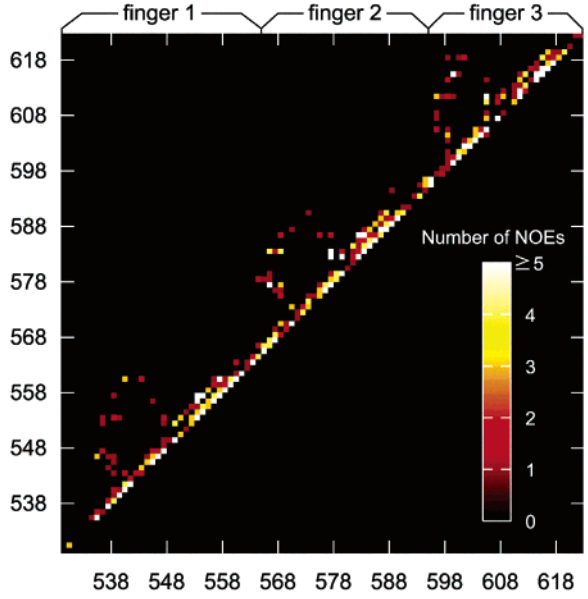

FIGURE 3: Diagonal plot of the NOE constraints used in the structure calculations.

His561 is stabilized in its N-terminus by two hydrogen bonds, His553(H_N)−Lys550(O) and Leu554(H_N)−Lys550(O), in addition to regular hydrogen bonds of α helix. A turn structure is observed in the residues Leu541−Cys544, which is maintained through two hydrogen bonds, Gly543(H_N)−Leu541(O) and Cys544(H_N)−Ile541(O). van der Waals interactions between the aliphatic side chain of Ile541 and an imidazole ring of His561 stabilize a loop structure. A hydrophobic core is formed by side chains of Tyr548, Leu554, and the β-methylene group of His537, of which the imidazole ring is fixed by a hydrogen bond between His537(Hδ1) and Gln536(O). The residues from N-terminus to Gln536, where only intraresidue and sequential NOEs are observed, may be disordered in solution.

Structure of Finger 2. A short antiparallel β sheet (Phe567−Cys569 and Lys576−Phe578) is formed through two hydrogen bonds, Phe578(H_N)−Phe567(O) and Cys569(H_N)−Lys576(O). An α helix corresponding to Ser581−His591 is stabilized in its N-terminus by two hydrogen bonds, Glu583(H_N)−Arg580(O) and Leu584(H_N)−Arg580(O), in addition to regular hydrogen bonds of α helix. In the turn region of finger 2, Thr570−Cys574, the backbone structure converged into several clusters that show different hydrogen-bonding patterns. In the ^{15}N-HSQC spectrum of Sp1, backbone amide resonances of Thr570 and Ser572 are not observed. These results indicate that the structure in this turn region is exchanging among several conformations in solution. The van der Waals contact between an indole ring of Trp571 and an imidazole ring of His591 are observed in all structures and stabilize the turn structure. A hydrophobic core is formed by side chains of Phe578 and Leu584 and the β-methylene groups of Phe567 and Cys569. The superposition of the resulting structure of the well-conserved region, Phe567−His591, with that of the reported structure of finger 2 (*14*) (PDB code 1SP2), shows the rmsd of 1.07 ± 0.13 Å in backbone atoms. Both structures are nearly identical in α helix, β strands and hydrophobic core. The turn structure of 1SP2 is well-defined and distinct from our structures in hydrogen-bonding pattern. It is suggested that different measurement conditions (1SP2, pH 5.9 at 8 °C), may affect

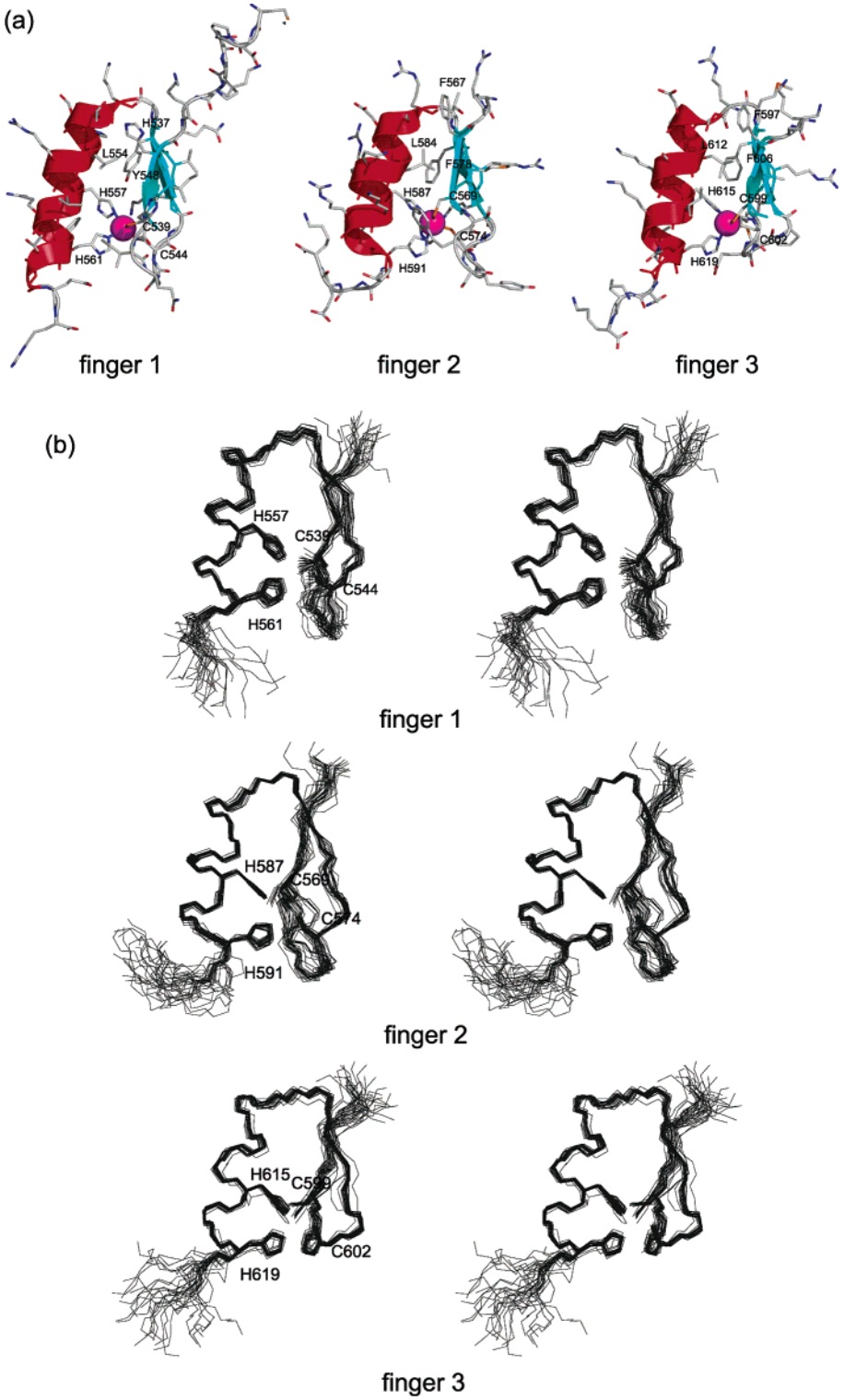

FIGURE 4: Calculated structures of the zinc fingers from Sp1(530−623). (a) Average refined structures of each finger. The positions of the β-strands, α-helix (in ribbon), and zinc ion (in sphere) are shown. Metal binding residues and well-conserved hydrophobic residues are annotated with residue numbers. (b) Stereoview of the 30 energy-minimized structures of each finger. The structures are best-fit superpositioned by use of backbone atoms of highly conserved region (finger 1, 537−561; finger 2, 567−591; finger 3, 597−619). The zinc-coordinating residues are also shown.

the turn-stabilizing interaction between Trp571 and His591 and cause such a difference. In the N-terminal region of finger 2, a hydrogen bond between Phe567(H_N) and ARG565-(O), which is not observed in 1SP2, makes a bent structure as generally seen in the DNA complex of Cys_2His_2 zinc finger protein. A proceeding finger moiety is probably necessary for such a conformation.

Structure of Finger 3. A short antiparallel β sheet, (Phe597−Cys599 and Lys604−Phe606), is formed through two hydrogen bonds, Phe606(H_N)−Phe597(O) and Cys599-(H_N)−Lys604(O). An α helix corresponding to Ser609−His619 is stabilized in its N-terminus by a hydrogen bonds, Leu612(H_N)−Arg608(O), in addition to regular hydrogen bonds of α helix. A turn structure is observed in the residues

Table 1: NMR Structure Statistics

	finger 1	finger 2	finger 3
NOE upper distance restraints	291	310	248
no. of violations > 0.2 Å	0	0	0
RMSD from upper bounds (Å)	0.0063 ± 0.0010	0.0168 ± 0.0013	0.0066 ± 0.0008
dihedral angle restraints (φ and ψ) (deg)	35	33	17
no. of violations > 2 deg	0	0	0
RMSD from upper bounds (deg)	0.0628 ± 0.0163	0.3281 ± 0.0684	0.0605 ± 0.0307
	Energies (kcal mol^{-1})		
F_{total}	24.15 ± 1.02	50.54 ± 4.80	25.11 ± 0.57
F_{NOE}	0.96 ± 0.31	7.16 ± 1.07	0.89 ± 0.20
F_{tor}	0.01 ± 0.01	0.25 ± 0.11	0.01 ± 0.01
F_{vdw}	3.88 ± 0.67	14.62 ± 2.82	4.15 ± 0.41
	RMSD from Idealized Geometry		
bonds (Å)	0.0009 ± 0.000 05	0.0019 ± 0.000 12	0.0013 ± 0.000 04
angles (deg)	0.3308 ± 0.0027	0.4123 ± 0.0092	0.3651 ± 0.0017
impropers (deg)	0.0957 ± 0.0066	0.1991 ± 0.0146	0.0994 ± 0.0057
conserved region	537−561	567−591	597−619
main-chain atoms rmsd (Å)	0.50 ± 0.11	0.49 ± 0.17	0.37 ± 0.14
all heavy atoms rmsd (Å)	1.21 ± 0.17	1.28 ± 0.29	1.03 ± 0.17

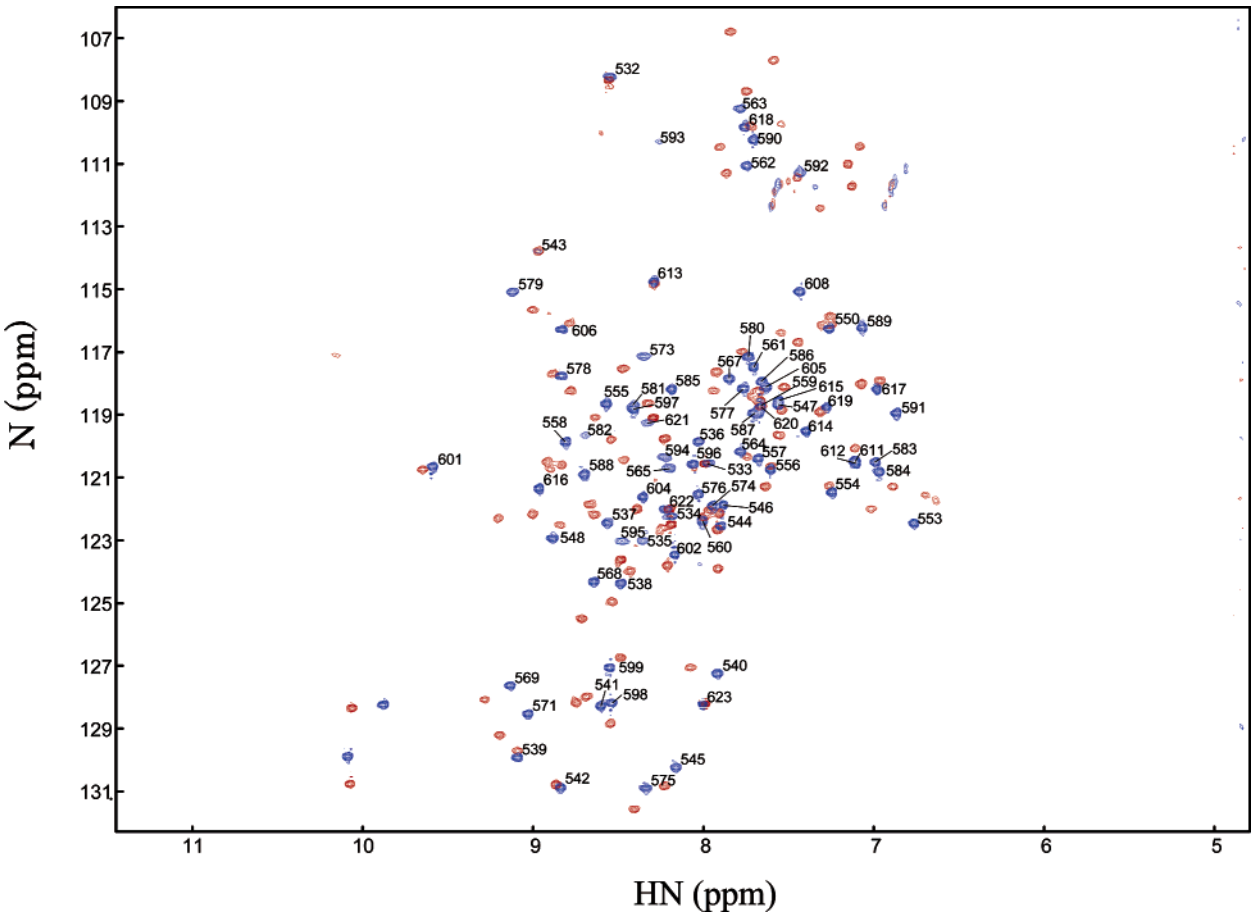

FIGURE 5: ^{15}N-HSQC spectra of Sp1(530−623) and Sp1(530−623)−DNA complex, shown in blue and red, respectively. The spectrum of Sp1(530−623) is shown with assignments.

Cys599−Cys602, which is maintained through a hydrogen bond between Cys602(H_N) and Cys599(O). A hydrophobic core is formed by side chains of Phe606, Leu612, and the β-methylene group of Phe597. The superposition of the resulting structure of the well-conserved region Phe597−His619 with that of the reported structure of finger 3 (*14*) (PDB code 1SP1) shows an rmsd of 0.75 ± 0.10 Å in backbone atoms.

DNA-Binding Analysis. We found that many resonances of ^{15}N-HSQC spectrum of Sp1 change their chemical shifts when complexed with DNA (Figures 5 and 6). In particular, linker regions between fingers show large changes. This result indicates that the backbone conformations of linker regions alter upon DNA binding. Such chemical shift perturbations in linker regions have been reported for other multifinger proteins (*24*, *25*). Generally, the interactions between adjacent fingers and major grooves of target DNA hold the linker regions. Although it is necessary to perform extensive analysis of heteronuclear relaxation in order to clarify the role of the flexibility in the binding, it was suggested that the linker regions, which are flexible in the free state, would become more rigid and ordered. The large

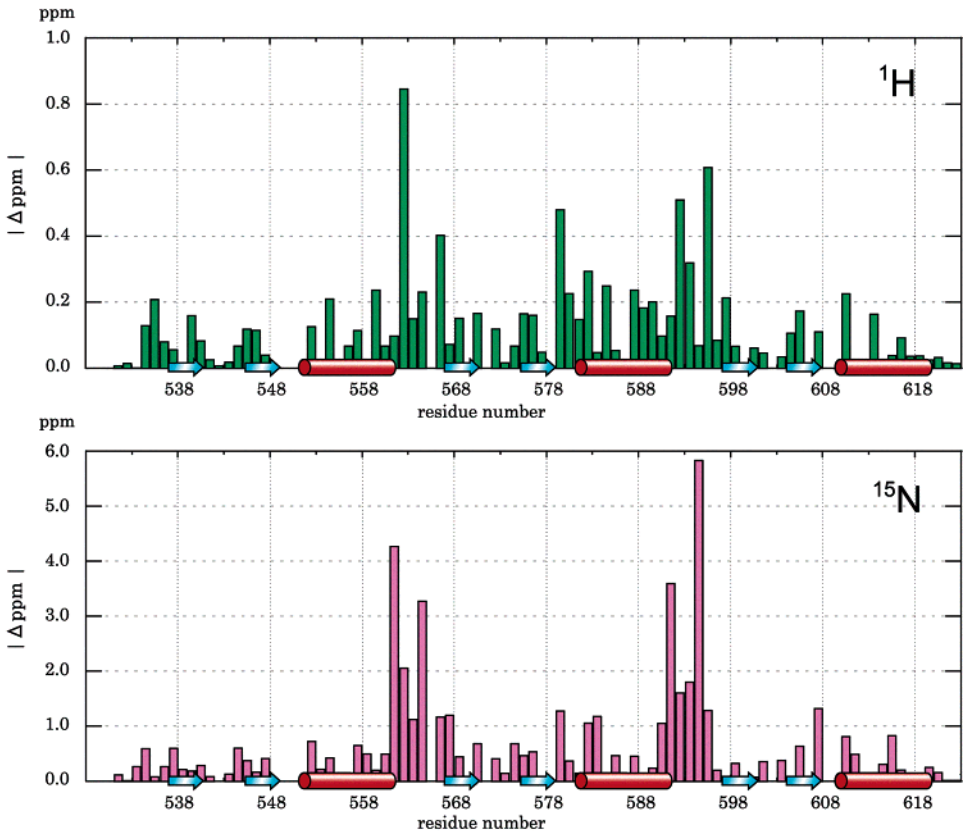

FIGURE 6: Chemical shift perturbation of backbone H−N atoms upon DNA binding. The secondary structure is represented schematically.

changes in chemical shifts in two linker regions of Sp1 indicate that the binding mode of three fingers, as a whole, is similar to that of Zif268 or TFIIIA (*5*, *6*, *26*, *27*).

DISCUSSION

From structural comparison of zinc fingers of Sp1 with other fingers, many interactions between Sp1(530−623) and bases or phosphate backbones of target DNA were predicted (Figure 7). As discussed in the following section, most of the molecular biological character of Sp1(530−623) derived from mutant analysis, methylation interference analysis, ethylation interference analysis, and so on, can be explained by these interactions.

Interactions between Sp1 and Phosphate Backbone. In canonical fingers, three well-conserved basic residues that interact with the phosphate backbone of G-strand through direct or indirect hydrogen bonds have been reported as follows: (1) His at position 7, (2) Lys or Arg at the head of the second β strand, and (3) Lys or Arg at the fourth positions in the linker region (TGEKP linker) (*28*, *29*). In the resulting structure of Sp1, side chains of such residues corresponding to 1 and 2 in three fingers (1, His557, His587, and His615; 2, Lys546, Lys576, and Lys604) show the same orientations as observed in the reported DNA complex structure of other zinc finger proteins. Therefore, these residues are likely to form hydrogen bonds to the phosphate backbone. However, instead of His557, Tyr548 at third position of the second β strand in finger 1 can donate a hydrogen bond from its hydroxyl group to the phosphate group as observed in finger 1 of TFIIIA (*26*, *27*) and fingers 3 and 5 of GLI (*30*). On the other hand, the orientations of residues corresponding to (3), Arg565 and Lys595, are not determined. However,

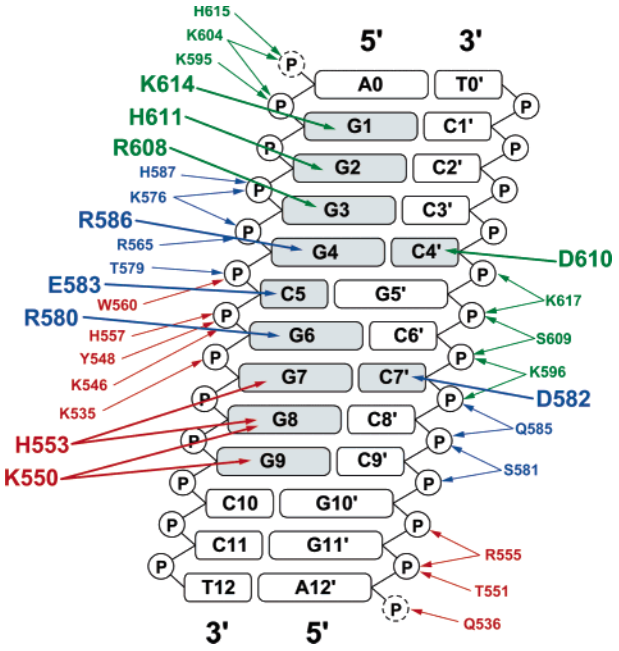

FIGURE 7: Summary of predicted DNA-binding mode of Sp1(530−623). The boxes represent DNA bases, and the phosphate backbone is represented by a chain of circles. The recognized bases are colored in gray. The phosphate groups at the 5′-terminus, which were absent in the synthesized DNA, are represented in dashed circles. Arrows indicate possible interactions between residue and DNA base or phosphate group. The residues that recognize a DNA base are represented in large characters, and the residues from different fingers are color-coded: finger 1, red; finger 2, blue; and finger 3, green. As described in the Discussion section, Lys550 and His553 may form alternative hydrogen bonds to two base positions.

from the ethylation interference experiments, it has been suggested that these residues interact with the phosphate group of G-strand (*18*). Large chemical shift changes upon DNA binding are consistent with the specific interaction between linker regions and the phosphate backbone. Interestingly, in Lys535, which is not in the linker region but located at the corresponding position of finger 3 relative to the following finger, and in Gln536, ethylation interference experiments demonstrate possible interactions with G-strand and C-strand, respectively (*18*). These residues show moderate chemical shift perturbations. In the DNA complex, the N-terminal region of Sp1(530−623) would probably become rigid to interact with phosphate group as linker regions. Such interactions have been reported in Lys11 and Arg12 of finger 1 of TFIIIA (*26*, *27*). In addition to the above-mentioned interactions, the residues at position 1 (Thr551, Ser581, and Ser609) and position 5 (Arg555, Gln585, and Lys596) are likely involved by interactions with the C-strand in our model, while the G-strand possibly interacts with Trp560 and Thr579.

DNA Base Recognition Modes of Fingers 2 and 3. The amino acid sequence of the N-terminal region of a recognition helix in finger 2 of Sp1, RSDELQR, is quite similar to that of finger 1, RSDELTR, or finger 3, RSDERKR, of Zif268. Three-dimensional structure of this region fits with small pairwised rmsd values (0.95 ± 0.07 and 0.91 ± 0.05 Å). This structural similarity and the binding preference to GCG common to these fingers suggest that the DNA recognition modes of these fingers are very similar to each other. The GCG sequence of target DNA would be recognized by following interactions: two hydrogen bonds between the guanidyl group of Arg580 at position −1 to O7 and N7 of Gua6, hydrophobic interaction between Cγ, Cδ of Glu583 at position 3 and C5, C6 of Cyt5, and two hydrogen bonds between the guanidyl group of Arg586 at position 6 to O7 and N7 of Gua4. At the same time, as observed in X-ray structures of Zif268, two hydrogen-bonding interactions between side chains of Arg580 and Asp582 at position 2 may be formed and fix side chain conformation of Arg580. In this model, the preference to G−C pair at seventh position of DNA (*3*, *31*) can be accounted for the possible interaction between Asp582 and N4 of Cyt7′.

In finger 3 of Sp1, except for a basic Lys residue at position 6, the amino acid sequence of the N-terminal region of a recognition helix, RSDHLSK, is similar to that of finger 2, RSDHLTT, of Zif268. The superposition of the backbone structure onto that of finger 2 of Zif268 with a small rmsd of 0.79 ± 0.10 Å also indicates both fingers recognize bases at positions −1, 2, and 3 in a similar way. Thus, it is suggested that the guanidyl group of Arg608 forms two hydrogen bonds to O7 and N7 of Gua3 and Nϵ of His611 forms a hydrogen bond to N7 of Gua2 or Ade2. The interaction between arginine at position −1 and aspartic acid at position 2 may be formed as the case of finger 2 of Sp1. On the other hand, Lys residue of position 6 would form a hydrogen bond to N7 of Gua1 at the Nξ atom, as found in finger 3 of TFIIIA (*26*, *27*). It has been reported that Sp1-(530-623) is able to recognize Thy1 (*3*, *31*). In such a case, a hydrogen bond between Nξ of Lys614 and O2 of Thy base may be formed.

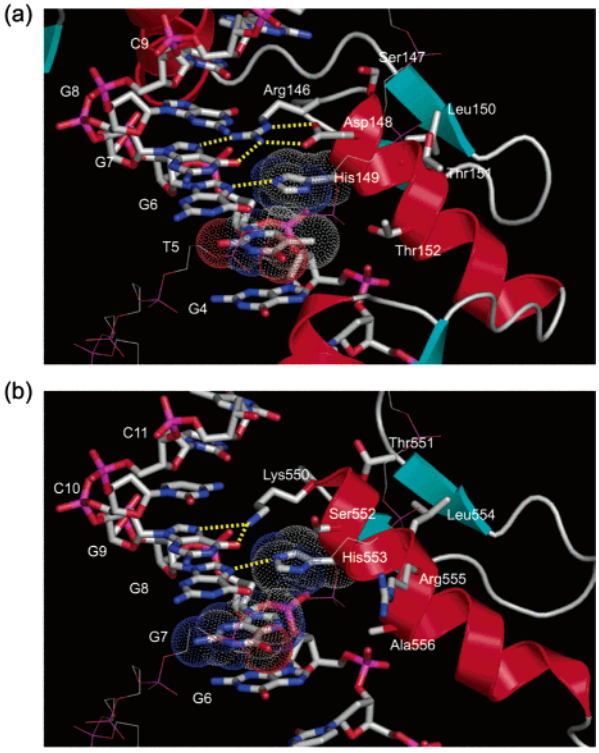

FIGURE 8: X-ray structure of Zif268-finger 2 and DNA binding model structure of Sp1-finger 1. The structures of Zif268-finger 2 and Sp1-finger 1 are represented in ribbon model. The side chains at positions −1 to 6 of the recognition helix and G-strands are represented in sticks. The backbone atoms of C-strands are represented by lines. (a) Base recognition mode of Zif268-finger 2. Hydrogen bonds observed in the X-ray structure are represented by dashed lines. The hydrogen bonds between Arg146 and Asp148 at positions −1 and 2 of the helix fix the orientation of the arginine side chain to increase the specificity of the arginine−guanine interaction. The van der Waals contact between the imidazole ring of His149 and the methyl group of Thy5 (represented by dots) may play important role in site-specific recognition (*6*). (b) Base recognition model of Sp1-finger 1. The model structure was built by appropriate modifications of base composition in X-ray structure of Zif268 (PDB 1AAY) and replacement of Zif268-finger 2 by Sp1-finger 1. Conflicted atoms were corrected manually. The possible hydrogen bonds corresponding to those concerned with base recognition in Zif268-finger 2 are represented by dashed lines. In contrast to Arg148 of Zif268, Lys550 at position −1 of helix lacks hydrogen-bonding interactions to fix the orientation of its side chain. In the case of guanine base at position 7 of the G-strand, the van der Waals contact between His553 and methyl group no longer exists.

Unique DNA Recognition Mode of Finger 1. In canonical finger, position 6 of the recognition helix is a basic residue such as Arg or Lys interacting with the Gua base (*29*). However, the corresponding residue is Ala556 in finger 1 of Sp1: that is not likely to recognize a base. Such substitution is also found in finger 2 of Zif268 (PDB code 1AAY), where Thr occurs in position 6 (*6*). Finger 2 of Zif268 recognizes Gua bases at −1 and 3 positions of the recognition helix with Arg and His residues, respectively (Figure 8a). Interestingly, the corresponding residues in finger 1 of Sp1, Lys550 and His553, are basic residues, as occurs in finger 2 of Zif268. These facts suggest that both fingers have a similar DNA recognition mode, in which Nξ of Lys550 forms a hydrogen bond to O7 or N7 of Gua9 base and Nϵ of His553 forms a hydrogen bond to N7 of Gua/Ade8. Our NMR structure indicates that finger 1 of Sp1 has

also typical folding of a C_2H_2-type zinc finger, and chemical shift perturbations of Sp1 suggest that the overall folding of each finger might not change upon its binding to target DNA. Thus, as shown in Figure 8b, we have constructed a hypothetical model of DNA bound finger 1 based on its structural similarities to finger 2 of Zif268 by manual rigid body fitting. In contrast to Zif268, Sp1(530−623) is known to recognize Ade9 in addition to Gua9 with the same affinity (*31*, *32*). According to our model, such tolerance can be attributed to the substitutions of Arg at position −1 in Zif268 fingers, which specifically recognizes O7 and N7 of Gua base by two hydrogen bonds, to Lys in finger 1 of Sp1. On the other hand, some differences from finger 2 of Zif268 could be expected. Position 2 in finger 2 of Zif268, where an Asp residue is important for the recognition of Cyt10′ through a hydrogen bond and, at the same time, fixes the side chain of Arg residue at position −1 by two hydrogen bonds, is occupied by a Ser residue in finger 1 of Sp1 (Figure 8a). Because of this substitution, Sp1(530−623) lacks preference for Cyt at 10′ (*3*, *31*) and the side-chain conformation of Lys550 at position −1, which is important for base recognition, would be unfixed. A similar case occurs in finger 1 of TFIIIA (PDB code 1TF3); it was shown that Lys at position −1 of the helix has several hydrogen-bonding partners (*26*). Thus, it is suggested that Lys at position −1 of finger 1 of Sp1 also has several hydrogen-bonding partners. The result of methylation interference analysis of Sp1(530−623), which suggests 2-fold interactions between Lys550 and Gua8/9, also supports this suggestion (*17*). In addition, in finger 2 of Zif268, a stacking interaction between a methyl group of Thy base at G-strand and the imidazole ring of His residue at position 3 may play an important role for fixing the side-chain conformation and site-specific recognition (*33*). But in the case of Sp1, the corresponding base at 7 is not necessarily a Thy base. Recent MD calculation of Sp1 DNA binding domain reported that the lack of this His−Thy interaction may change the hydrogen-bonding partner of His553 to the preceding base (*34*). The lack of these interactions to fix the side chains of Lys550 and His553, which are concerned with base recognition, would allow these residues to have several hydrogen-bonding partners at multiple base positions (Figure 7). Consequently, finger 1 of Sp1 may have less sequence and site specificity than finger 2 of Zif268. It was known that the lack of finger 1 causes remarkable loss of affinity for target DNA (*17*). Because of this relaxed base recognition of finger 1, Sp1-(530−623) can bind more various sequences than other multi-C_2H_2-type zinc fingers, and such a property may be required for the ubiquitous transcription factor Sp1, which activates transcription of many genes.

SUPPORTING INFORMATION AVAILABLE

Chemical shifts for Sp1(530−623). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Dynan, W. S., and Tjian, R. (1983) Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II, *Cell 32*, 669−80.
2. Kadonaga, J. T., Carner, K. R., Masiarz, F. R., and Tjian, R. (1987) Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain, *Cell 51*, 1079−90.
3. Kadonaga, J. T., Jones, K. A., and Tjian, R. (1986) Promoter-specific activation of RNA polymerase II transcription by Sp1, *Trends Biochem. Sci. 11*, 20−23.
4. Bucher, P. (1990) Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences, *J. Mol. Biol. 212*, 563−78.
5. Pavletich, N. P., and Pabo, C. O. (1991) Zinc finger−DNA recognition: crystal structure of a Zif268−DNA complex at 2.1 A, *Science 252*, 809−17.
6. Elrod-Erickson, M., Rould, M. A., Nekludova, L., and Pabo, C. O. (1996) Zif268 protein−DNA complex refined at 1.6 A: a model system for understanding zinc finger−DNA interactions, *Structure 4*, 1171−80.
7. Desjarlais, J. R., and Berg, J. M. (1992) Toward rules relating zinc finger protein sequences and DNA binding site preferences, *Proc. Natl. Acad. Sci. U.S.A. 89*, 7345−9.
8. Desjarlais, J. R., and 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. U.S.A. 90*, 2256−60.
9. Rebar, E. J., and Pabo, C. O. (1994) Zinc finger phage: affinity selection of fingers with new DNA-binding specificities, *Science 263*, 671−3.
10. Choo, Y., and Klug, A. (1994) Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage, *Proc. Natl. Acad. Sci. U.S.A. 91*, 11163−7.
11. Choo, Y., and Klug, A. (1994) Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions, *Proc. Natl. Acad. Sci. U.S.A. 91*, 11168−72.
12. Greisman, H. A., and Pabo, C. O. (1997) A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites, *Science 275*, 657−61.
13. Segal, D. J., and Barbas, C. F., 3rd (2000) Design of novel sequence-specific DNA-binding proteins, *Curr. Opin. Chem. Biol. 4*, 34−9.
14. Narayan, V. A., Kriwacki, R. W., and Caradonna, J. P. (1997) Structures of zinc finger domains from transcription factor Sp1. Insights into sequence-specific protein−DNA recognition, *J. Biol. Chem. 272*, 7801−9.
15. Thiesen, H. J., and Bach, C. (1990) Target Detection Assay (TDA): a versatile procedure to determine DNA binding sites as demonstrated on SP1 protein, *Nucleic Acids Res. 18*, 3203−9.
16. Shi, Y., and Berg, J. M. (1995) A direct comparison of the properties of natural and designed zinc-finger proteins, *Chem. Biol. 2*, 83−9.
17. Yokono, M., Saegusa, N., Matsushita, K., and Sugiura, Y. (1998) Unique DNA binding mode of the N-terminal zinc finger of transcription factor Sp1, *Biochemistry 37*, 6824−32.
18. Nagaoka, M., and Sugiura, Y. (1996) Distinct phosphate backbone contacts revealed by some mutant peptides of zinc finger protein Spl: effect of protein-induced bending on DNA recognition, *Biochemistry 35*, 8761−8.
19. Kay, L. E. (1995) Pulsed field gradient multidimensional NMR methods for the study of protein structure and dynamics in solution, *Prog. Biophys. Mol. Biol. 63*, 277−99.
20. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J. Biomol. NMR 6*, 277−93.
21. Cornilescu, G., Delaglio, F., and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology, *J. Biomol. NMR 13*, 289−302.
22. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr. D54* (Pt. 5), 905−21.
23. Diakun, G. P., Fairall, L., and Klug, A. (1986) EXAFS study of the zinc-binding sites in the protein transcription factor IIIA, *Nature 324*, 698−9.
24. Foster, M. P., Wuttke, D. S., Clemens, K. R., Jahnke, W., Radhakrishnan, I., Tennant, L., Reymond, M., Chung, J., and Wright, P. E. (1998) Chemical shift as a probe of molecular interfaces: NMR studies of DNA binding by the three amino-

terminal zinc finger domains from transcription factor IIIA, *J. Biomol. NMR 12*, 51−71.
25. Schmiedeskamp, M., Rajagopal, P., and Klevit, R. E. (1997) NMR chemical shift perturbation mapping of DNA binding by a zinc-finger domain from the yeast transcription factor ADR1, *Protein Sci. 6*, 1835−48.
26. Foster, M. P., Wuttke, D. S., Radhakrishnan, I., Case, D. A., Gottesfeld, J. M., and Wright, P. E. (1997) Domain packing and dynamics in the DNA complex of the N-terminal zinc fingers of TFIIIA, *Nat. Struct. Biol. 4*, 605−8.
27. Nolte, R. T., Conlin, R. M., Harrison, S. C., and Brown, R. S. (1998) Differing roles for zinc fingers in DNA recognition: structure of a six-finger transcription factor IIIA complex, *Proc. Natl. Acad. Sci. U.S.A. 95*, 2938−43.
28. Suzuki, M., Gerstein, M., and Yagi, N. (1994) Stereochemical basis of DNA recognition by Zn fingers, *Nucleic Acids Res. 22*, 3397−405.
29. Wolfe, S. A., Greisman, H. A., Ramm, E. I., and Pabo, C. O. (1999) Analysis of zinc fingers optimized via phage display: evaluating the utility of a recognition code, *J. Mol. Biol. 285*, 1917−34.
30. Pavletich, N. P., and Pabo, C. O. (1993) Crystal structure of a five-finger GLI−DNA complex: new perspectives on zinc fingers, *Science 261*, 1701−7.
31. Briggs, M. R., Kadonaga, J. T., Bell, S. P., and Tjian, R. (1986) Purification and biochemical characterization of the promoter-specific transcription factor, Sp1, *Science 234*, 47−52.
32. Jones, K. A., Kadonaga, J. T., Luciw, P. A., and Tjian, R. (1986) Activation of the AIDS retrovirus promoter by the cellular transcription factor, Sp1, *Science 232*, 755−9.
33. Elrod-Erickson, M., Benson, T. E., and Pabo, C. O. (1998) High-resolution structures of variant Zif268−DNA complexes: implications for understanding zinc finger−DNA recognition, *Structure 6*, 451−64.
34. Marco, E., Garcia-Nieto, R., and Gago, F. (2003) Assessment by molecular dynamics simulations of the structural determinants of DNA-binding specificity for transcription factor Sp1, *J. Mol. Biol. 328*, 9−32.

BI048438P