Binding of Transcription Factor Sp1 to GC Box DNA Revealed by Footprinting Analysis: Different Contact of Three Zinc Fingers and Sequence Recognition Mode†

Jun Kuwahara, Atsuo Yonezawa, Mayumi Futamura, and Yukio Sugiura*

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

Received December 1, 1992; Revised Manuscript Received March 24, 1993

ABSTRACT: Transcription factor Sp1 has three tandem repeats of a Cys_2His_2-type zinc finger motif and specifically binds to GC box DNA. We investigated the interaction of Sp1 with GC box DNA by several footprinting techniques. Methylation of four guanine bases in the sequence 5′-GGGCG-3′ is strongly protected by Sp1 binding, whereas a guanine base flanked at the 5′ end by the above sequence is extremely hypermethylated. Methylation interference experiments explicitly show that four guanine bases from the guanine-rich strand, and one from the cytosine-rich strand, in the sequence 5′-GGGCG-3′ are crucially required for GC box recognition by Sp1. In footprinting using the 1,10-phenanthroline–copper complex, binding of Sp1 clearly alters the cleavage patterns by the metal complex. Footprints of the protein did not cover the full length of each GC box sequence, and the protein strongly masked scission in the sequence 5′-GCGG(A/G)(G/A)-3′. In cleavage of GC box DNA by the bleomycin–iron complex, Sp1 binding induces new cutting at a 5′-GA-3′ site within the box. The results indicate that (i) the three zinc fingers do not contribute equivalently to the binding of Sp1 to the GC box, namely, important base contacts arise from the second and third fingers, and (ii) the protein binding induces local but significant structural distortion of the 3′ region of the guanine-rich strand in the GC box. These features are clearly distinct from those of Zif268 and Krox20, which are three-zinc-finger proteins closely related to Sp1.

Transcription factor Sp1 isolated from HeLa cell extracts is a sequence-specific DNA binding protein that binds to proximal promoter sequences of various cellular and viral genes and activates transcription of these genes by RNA polymerase II (Gidoni et al., 1984; Kadonaga et al., 1986, 1987). The basic recognition unit of Sp1 has been proposed to be a decanucleotide (GC box) with the consensus sequence 5′-(G/T)GGGCGG(G/A)(G/A)(C/T)-3′ (Briggs et al., 1986; Kadonaga et al., 1986), although the apparent binding affinity for Sp1 varies with the individual base sequence. Sp1-responsive promoters contain multiple Sp1 binding sites with different spacing. However, a single binding site appears to be sufficient for a promoter stimulated by Sp1 (Dynan & Tjian, 1985). The DNA binding domain located at the C-terminus of the protein contains three contiguous repeats of a classical Cys_2His_2-type zinc finger motif. Multiple repeats of this module are typical of the Cys_2His_2-type zinc finger proteins. A crystal structure of a three-zinc-finger protein complexed with DNA has become available for mouse immediate early protein Zif268 and has revealed that each individual finger engages equivalently in base recognition (Pavletich & Pabo, 1991). Previous studies using divergent Sp1-responsive promoters and mutated oligonucleotides harboring GC box sequences have given useful information concerning the DNA sequence recognized by Sp1 (Letovsky & Dynan, 1989; Thiesen & Bach, 1990; Chavrier et al., 1990). Nevertheless, DNA binding and the base contacts made by the protein are not known in detail. In order to clarify the mode of interaction of Sp1 with GC box DNA, we have applied several footprinting techniques to the Sp1(167*)–GC box DNA complex. Herein, we demonstrate the roles of the three zinc fingers in GC box DNA recognition and propose the DNA binding mode of Sp1.

MATERIALS AND METHODS

Materials. All enzymes were purchased from Takara Shuzo (Kyoto, Japan), except for *Nco*I and *Bst*NI, which were obtained from New England Biolabs (Boston, MA). The Sp1 fragment, Sp1(167*), was prepared and purified as described previously (Kuwahara & Coleman, 1990). The second-generation bleomycin antibiotic peplomycin (PEM) was supplied by Nippon Kayaku (Tokyo, Japan). All other chemicals were of commercial reagent grade.

Preparation of the ^{32}P-End-Labeled DNA Restriction Fragments. The end-labeled DNA fragments for footprinting experiments were prepared from the plasmids pBR322 and pIBISV. Plasmid pIBISV (3171 bp)[1] containing six GC boxes from the SV40 early promoter region was constructed by insertion of a *Pvu*II–*Hin*dIII fragment (344 bp; location 272–5243–5171) of SV40 DNA between a unique *Sma*I site and a *Hin*dIII site of pIBI24 (International Biotechnologies Inc., New Haven, CT). To generate the fragment labeled at the 5′ end of the guanine-rich strand (G strand) of SV40 DNA, a *Nco*I-digested pIBISV restriction fragment was dephosphorylated, labeled with T4 polynucleotide kinase and [γ-^{32}P]-ATP, and recut with *Eco*RI. For the fragment labeled at the

† This study was supported in part by a Grant-in-Aid for Scientific Research on Priority Area from the Ministry of Education, Science, and Culture of Japan.

* To whom correspondence should be addressed.

[1] Abbreviations: bp, base pair; PEM, peplomycin; G strand, guanine-rich strand of GC box sequences in SV40 early promoter DNA or pBR322 plasmid DNA; C strand, cytosine-rich strand of GC box sequences in SV40 early promoter DNA or pBR322 plasmid DNA; DMS, dimethyl sulfate; EDTA, ethylenediaminetetraacetic acid (disodium salt); OP, 1,-10-phenanthroline.

0006-2960/93/0432-5994$04.00/0 © 1993 American Chemical Society

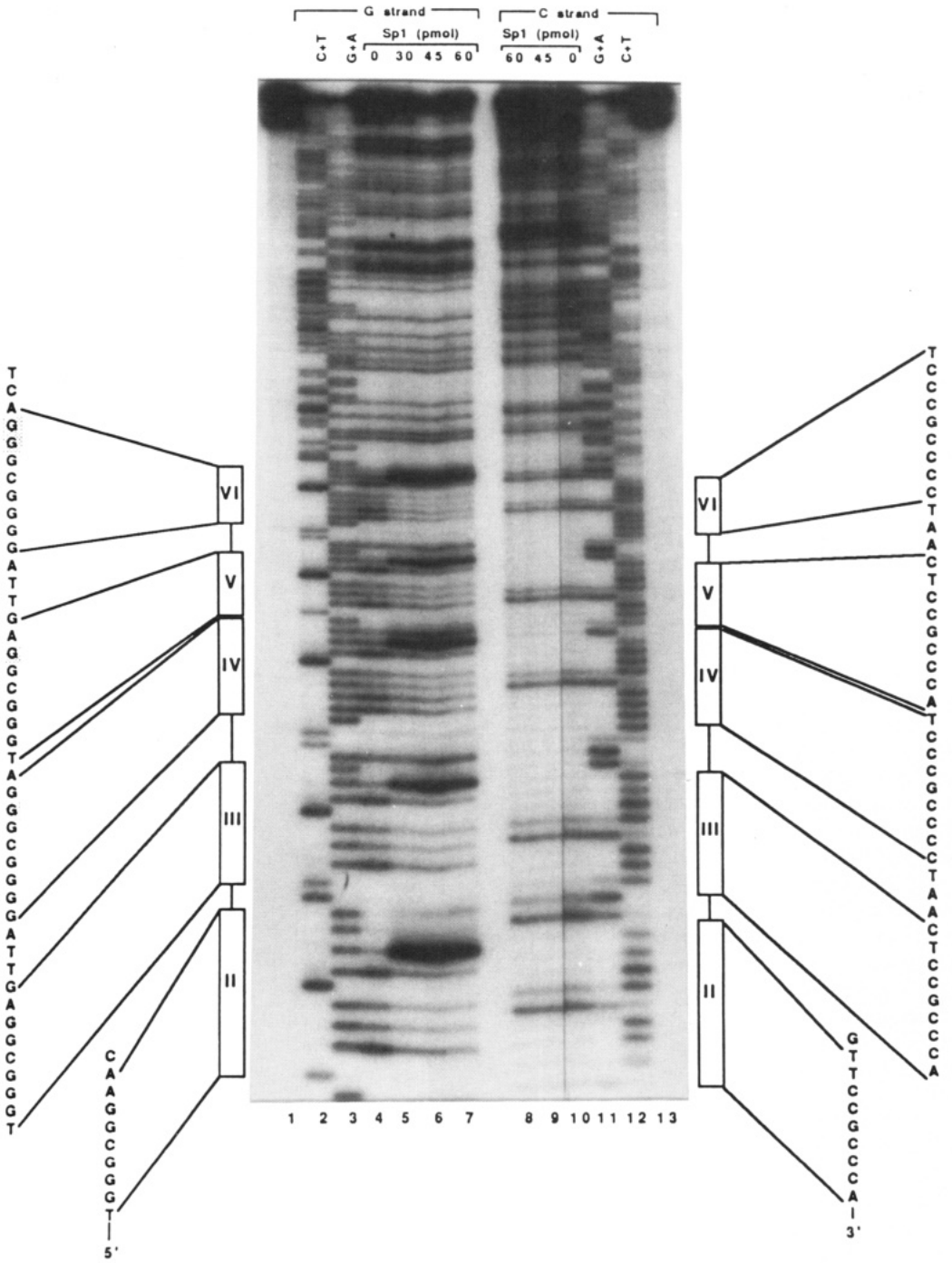

FIGURE 1: Methylation protection analysis for Sp1(167*)-bound SV40 early promoter DNA. The DNA fragments labeled at the 5′ end of the G strand (lanes 1–7) and the 3′ end of the C strand (lanes 8–13) were used. The samples (20 μL) contained the following amounts of Sp1(167*): none (lanes 4 and 10), 30 pmol (lane 5), 45 pmol (lanes 6 and 9), and 60 pmol (lanes 7 and 8). Lanes 1 and 13: intact DNA. Lanes 2 and 12: DNA sequencing (C+T) by the Maxam-Gilbert method. Lanes 3 and 11: DNA sequencing (G+A). Enhanced methylation sites are marked by stippling on the base sequence.

3′ end of the cytosine-rich strand (C strand), a *Nco*I-digested pIBISV restriction fragment was labeled using the Klenow fragment of DNA polymerase I and [α-^{32}P]dCTP and was cut again with *Eco*RI. To prepare the end-labeled fragment of pBR322 DNA, a *Bst*NI-digested restriction fragment was phosphorylated at the 5′ end with T4 polynucleotide kinase and [γ-^{32}P]ATP and then recut with *Hae*III for the G strand or TaqI for the C strand, providing a 105- or a 72-bp restriction fragment, respectively. The fragments containing the Sp1 binding site were gel-purified.

Methylation Protection Analysis. The reaction mixtures (final volume, 20 μL) contained the radiolabeled DNA fragment (pIBISV *Nco*I–*Eco*RI (253 bp), pBR322 *Bst*NI–*Hae*III (105 bp), or *Bst*NI–TaqI (72 bp)), sonicated calf thymus DNA (0.4 μg), and Sp1 (167*) (0–60 pmol) in a cacodylate buffer (25 mM sodium cacodylate (pH 7.3) and 10 mM $MgCl_2$). After preincubation at 20 °C for 30 min, the sample solutions were mixed with 0.5% dimethyl sulfate (DMS), and the reaction was allowed to proceed at 20 °C for 5 min. The reaction was terminated by the addition of 5 μL of a DMS stop solution (1.5 M sodium acetate (pH 7.0), 1 M β-mercaptoethanol, and 250 μg/mL yeast tRNA), and the DNA was precipitated with ethanol. Subsequently, the samples were treated with 1 M piperidine (100 μL) at 90 °C for 30 min and lyophilized. The cleavage products were redissolved in a loading buffer (98% deionized formate, 0.025%

Table I: GC Box Sequences Used in This Work

nucleotide									
1	2	3	4	5	6	7	8	9	GC boxes[a]
5′-G	G	G	G	C	G	G	A	G-3′	"pseudo GC" and I
5′-T	G	G	G	C	G	G	A	G-3′	III and V
5′-T	G	G	G	C	G	G	A	A-3′	II
5′-G	G	G	G	C	G	G	G	A-3′	IV and VI

[a] Boxes I–VI donate six Sp1-binding sites in SV40 early promoter DNA. "Pseudo GC" box indicates nine base pairs of Sp1 bound to pBR322 plasmid (location 2553–2545).

xylene cyanol, and 0.025% bromophenol blue) and analyzed by electrophoresis on a 10% polyacrylamide/7 M urea sequencing gel. DNA sequencing reactions were carried out by the Maxam–Gilbert method (Maxam & Gilbert, 1980). The bands were visualized by autoradiography.

Methylation Interference Analysis. The interference experiment was carried out according to the modified method of Siebenlist and Gilbert (1980). The 5′-end-labeled *Bst*NI–*Hae*III or *Bst*NI–TaqI restriction fragment of pBR322 in 200 μL of a cacodylate buffer was methylated by 0.5% DMS at 20 °C for 5 min. The reaction was terminated by the addition of 40 μL of DMS stop solution. The methylated DNA was precipitated and lyophilized. Binding reaction mixtures (20 μL) contained the methylated restriction fragment (approximately 5×10^5 cpm), sonicated calf thymus DNA (1 μg), Sp1(167*) (60 pmol), and 10 mM Tris buffer (pH 8.0). After incubation at 20 °C for 15 min, the protein-bound and free DNAs were separated on a 5% non-denaturing polyacrylamide gel and eluted from the gel with an elution buffer (0.5 M ammonium acetate, 10 mM magnesium acetate, 0.1% SDS, and 1 mM EDTA (pH 8.0)). The recovered DNAs were treated with 1 M piperidine (100 μL) at 90 °C for 30 min and then lyophilized. The cleavage products were analyzed on a 10% polyacrylamide/7 M urea sequencing gel. The bands were visualized by autoradiography. Densitometric analysis was performed with a Pharmacia LKB 2222 Ultra Scan XL laser densitometer. Based on the obtained relative cutting probabilities (F and B), the extent of interference was calculated for each base as the logarithm of the ratio of these two probabilities.

DNA Cleavage Analysis by PEM–Fe(II) Complex. The iron complex of peplomycin (PEM–Fe) was freshly prepared by mixing equal moles of PEM and iron sulfate ($FeSO_4 \cdot 7H_2O$) in a buffer containing 10 mM Tris-HCl (pH 7.5). The reaction mixtures (final volume, 20 μL) contained Sp1(167*) (0–45 pmol), the radiolabeled restriction fragment, and sonicated calf thymus DNA (0.2 μg) in a buffer containing 90 mM Tris–borate (pH 8.0). After incubation at 20 °C for 30 min, cleavage of the DNA samples was initiated by the addition of the PEM–Fe complex (final concentration, 0.5 μM) and sodium ascorbate (final concentration, 0.1 mM) and was allowed to proceed at 20 °C for 2 or 3 min. The cutting reaction was terminated by the addition of EDTA (0.5 mM) and sodium acetate (0.3 M). The DNAs were recovered as precipitate from ethanol and analyzed by electrophoresis on a 10% polyacrylamide/7 M urea sequencing gel.

Footprinting Analysis by $(OP)_2$–Cu(I) Complex. Reaction samples (16 μL) containing a 5′- or a 3′-end-labeled pIBISV DNA fragment (10 fmoL), sonicated calf thymus DNA (0.2 μg), Sp1(167*) (0, 5, 10, and 20 pmol), and 10 mM Tris-HCl (pH 8.0) were incubated at 20 °C for 30 min. Stock solution (2 μL) containing 2 mM 1,10-phenanthroline (OP) and 0.45 mM copper nitrate ($Cu(NO_3)_2 \cdot 3H_2O$) was added to the samples. The cleavage reaction was started by the addition of 2 μL of 58 mM mercaptopropionic acid, allowed to proceed at 20 °C for 2 or 4 min, and terminated by the addition of calf thymus DNA (1 μg), EDTA (0.5 mM), and sodium acetate (0.3 M). The cleavage products were recovered as precipitate from ethanol and analyzed by electrophoresis on a 10% polyacrylamide/7 M urea sequencing gel.

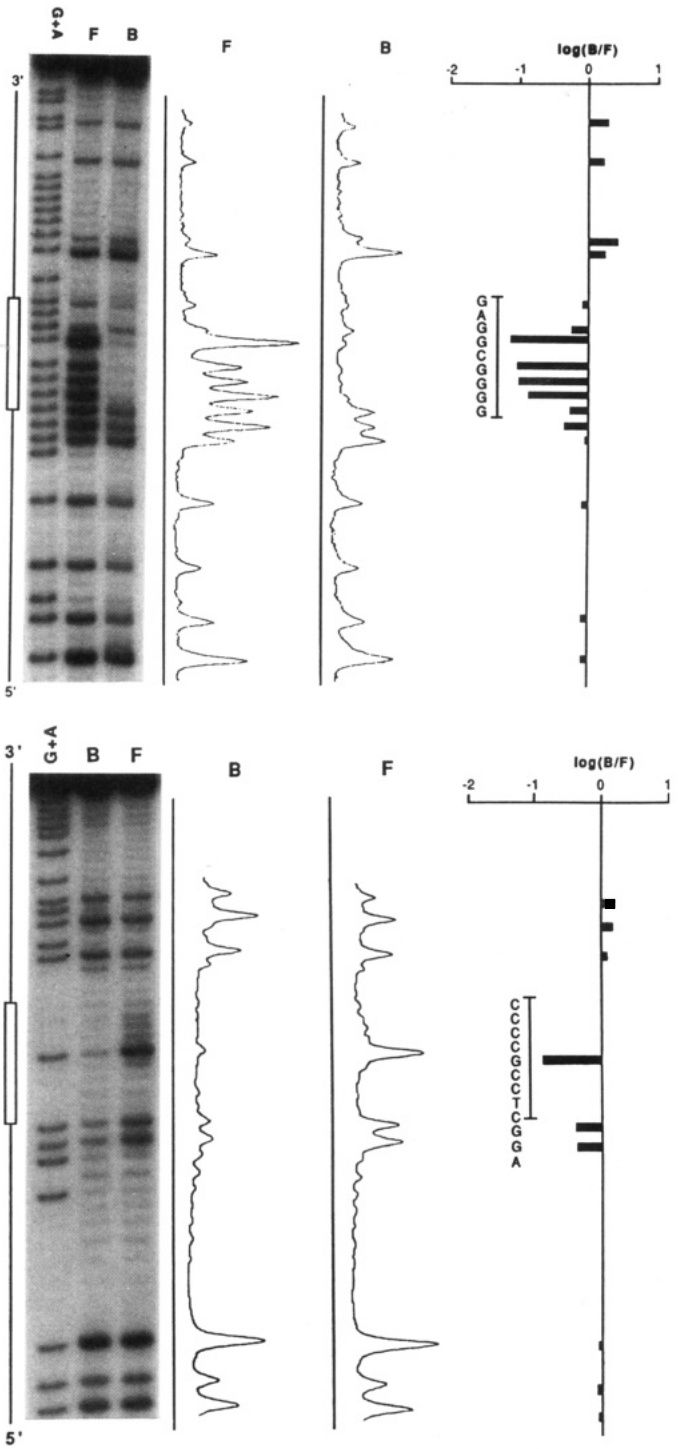

FIGURE 2: Methylation interference analyses for the G strand (panel a, top) and the C strand (panel, bottom) of Sp1(167*)-bound pseudo GC box DNA. The protein-bound (B) and free (F) DNAs were isolated, cleaved, and analyzed on a denaturing polyacrylamide gel (left). Lane G+A presents the Maxam–Gilbert sequence ladder prepared from the same fragment. An autoradiogram of the gel was scanned with a densitometer (middle). The scans for the free and the protein-bound DNAs were adjusted to match at the edge. The extent of interference was calculated as the logarithm of the ratio of the cutting probabilities for the two bands (log(B/F)), indicated by a histogram (right).

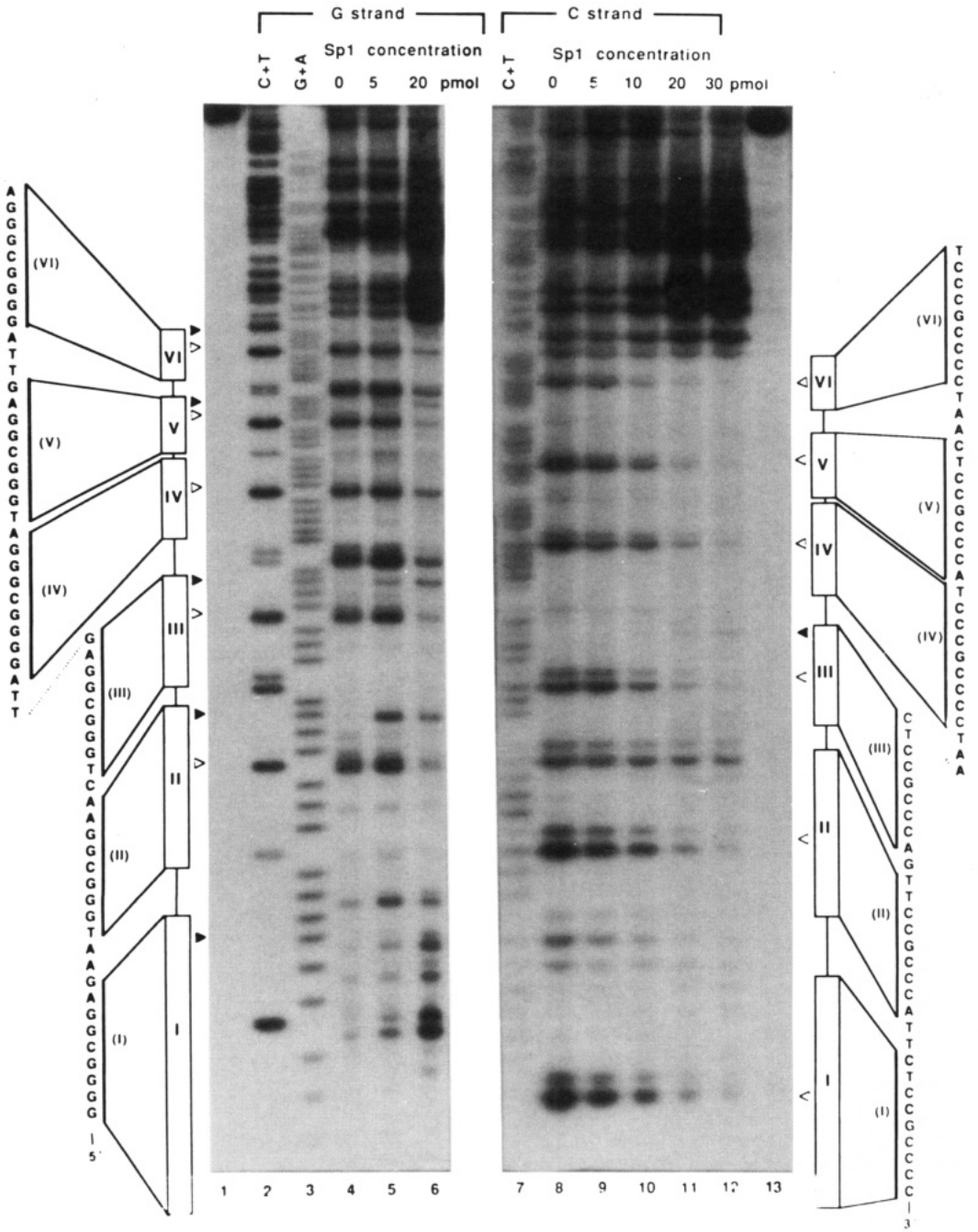

FIGURE 3: Cleavage analysis for Sp1(167*)-bound SV40 early promoter DNA by PEM–Fe(II) complex. DNA fragments labeled at the 5′ end of the G strand (lanes 1–6) and the 3′ end of the C strand (lanes 7–13) were used. The samples (20 μL) contained the following amounts of Sp1(167*): none (lanes 4 and 8), 5 pmol (lanes 5 and 9), 20 pmol (lanes 6 and 11), 10 pmol (lane 10), and 30 pmol (lane 12). Lanes 2 and 7: DNA sequencing (C+T) by the Maxam–Gilbert method. Lane 3: DNA sequencing (G+A). Lanes 1 and 13: intact DNA. Inhibited cleavage sites and new cutting sites are indicated with open and closed arrowheads, respectively, on the side of each panel.

RESULTS

Methylation Protection Analysis of the Sp1(167)–GC Box DNA.* Figure 1 shows methylation protection patterns of SV40 early promoter DNA complexed with Sp1(167*). As a function of increasing concentrations of Sp1(167*), runs of four guanine bases (G(2), G(3), G(4), and G(6)) at a central sequence, 5′-GGGCG-3′, in the G strand were clearly protected by the protein. In contrast, methylation at the G(7) base adjacent to the 3′ side of the above protected region was extremely enhanced by Sp1(167*) binding (lanes 5–7). These phenomena were reproducibly observed for all six GC boxes in SV40 DNA. In closer detail, additional protection was detected at the G(1) bases in boxes I, IV, and VI. Methylation of G(8) bases in boxes IV and VI was slightly promoted. On the other hand, similar apparent protection and enhancement were not observed in the C strand. A pBR322 restriction fragment that contains a single copy of a pseudo GC box sequence (location, 2545–2553) was also used for this purpose (Table I). The specific binding of Sp1(167*) to this fragment was carefully checked by gel mobility shift assay (data not shown). The methylation protection analysis for this fragment showed that five guanine bases (5′-*GGGGCGGAG*-3′) were protected from chemical modification, while the G(7) base was hypermethylated by the protein binding (data not shown).

Methylation Interference Analysis of the Sp1(167)–GC Box DNA.* In order to finely determine which guanine bases are contacted by Sp1(167*), methylation interference analysis was invstigated by use of a pBR322 restriction fragment

involving a single GC box sequence (Table I). Panels a and b of Figure 2 show the results of interference analysis for the G and the C strand, respectively, of the pseudo GC box DNA. Densitometric scans for the bands corresponding to 41 bases (location, 2527–2568) on the autoradiogram were performed. The extent of interference is indicated by a histogram. Of special interest is the fact that methylation of only four guanine bases (G(2), G(3), G(4), and G(6)) in the 5′-GGGGCGGAG-3′ site severely interfered with the binding of Sp1(167*) to the pseudo GC box. In addition, the protein binding was slightly weakened by premethylation of G(1), G(7), and G(9) in 5′-GGGGCGGAG-3′. Modification of the G base immediately preceding G(1) in 5′-GGGGGC-GGAG-3′ faintly affected Sp1(167*) binding. In the C strand, methylation of the G(5′) base in the center of the pseudo GC box strongly disturbed the binding of Sp1(167*) to this box. Additional interference at the G bases near the GC box, 5′-GGCTCCGCCCC-3′, was also observed.

Cleavage Analysis of the Sp1(167)–GC Box DNA by the PEM–Fe(II) Complex.* Figure 3 shows sequence-specific cleavage patterns of SV40 DNA by the PEM–Fe complex. The activated PEM–Fe complex cuts double-stranded DNA preferentially at 5′-G-pyrimidine-3′ sequences and much less effectively at 5′-GA-3′ and 5′-GG-3′ sites (lanes 3 and 7). The protein binding strongly inhibits cleavage at most of the bases in the G strand, especially at 5′-GC(5)-3′ sites in the center (open arrowheads). It is very interesting that new cleavages by the PEM–Fe(II) complex were induced at the 5′-GA(8) sites in boxes I, II, III, and V and slightly induced at 5′-GG(8) sites in boxes IV and VI (closed arrowheads). In the C strand of SV40 DNA, the digestion at the 5′-GC-(4′)-3′ site in the center of each GC box was clearly suppressed by increasing the concentration of Sp1(167*). The cleavage at the 5′-C(9′)T(8′)-3′ site in box III was slightly promoted. In the experiment using a fragment containing a single pseudo GC box, we also observed clear enhancement of DNA cleavage by Sp1(167*) binding at the 5′-GA(8)-3′ site in the G strand and at the 5′-GGCTCCGCCCC-3′ site in the C strand (Figure 4).

Footprinting Analysis of the Sp1(167)–GC Box DNA by $(OP)_2$–Cu(I) Complex.* Figure 5 shows footprinting patterns of Sp1(167*) with the 1,10-phenanthroline cuprous complex ($(OP)_2$–Cu(I)) on the SV40 early promoter DNA. In the absence of Sp1(167*), the $(OP)_2$–Cu(I) complex cuts the SV40 DNA fragment at every base on both strands and more effectively at $(dG)_n$ (n = 2–4) sites, reflecting its intrinsic order for base preference in digestion of DNA, G > C,T > A (lanes 3, 5, and 9). Binding of Sp1(167*) clearly alters the cleavage patterns by the $(OP)_2$–Cu(I) complex. In the G strand, footprints of Sp1(167*) did not cover the full length of each GC box sequence. The protein strongly masked cleavages at the sequence 5′-GCGG(A/G)(G/A)-3′, especially in boxes III, IV, V, and VI, whereas two or three nucleotides from the 5′ end of a GC box were unaffected by Sp1(167*) binding. In the C strand, the dinucleotides at 5′-G(5′)C(4′)-3′ in the center of each box (boxes III–VI) were faintly exposed to cleavage by the $(OP)_2$–Cu(I) complex. In comparison with the scission events in the G strand, footprints in the complementary C strand were very weak.

DISCUSSION

Specific Contacts with Guanine Bases in GC Box DNA. The Sp1 fragment, Sp1(167*), can specifically bind to an oligonucleotide containing a GC box sequence, 5′-GGGGC-GGGG-3′ (Kuwahara & Coleman, 1990), and its methylation protection pattern on the SV40 early promoter is remarkably close to that for native Sp1. In the methylation protection experiments using other GC box sequences in the monkey β promoter and the ARV-2 (HIV) retroviral long terminal repeat (LTR), similar protection and enhancement patterns for native Sp1 have also been observed (Gidoni et al., 1984; Jones et al., 1986). Thus, Sp1(167*) appears to bind GC boxes in the same manner as native Sp1.

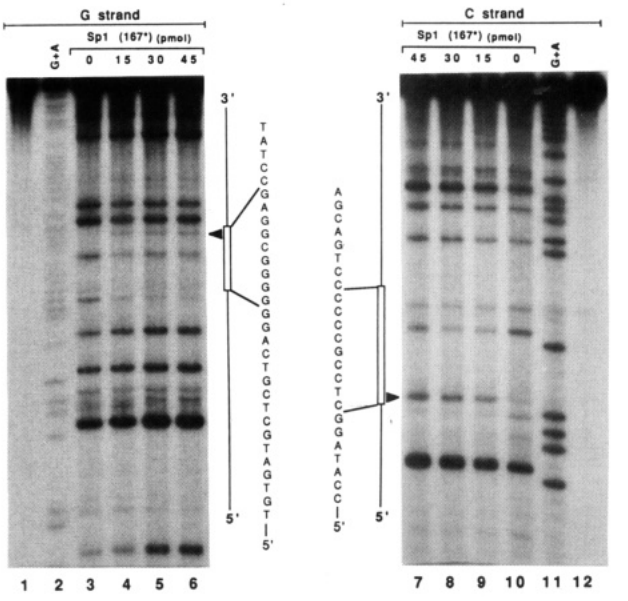

FIGURE 4: Cleavage analysis for Sp1(167*)-bound pseudo GC box DNA by PEM–Fe(II) complex. DNA fragments labeled at the 5′ end of the G strand (lanes 1–6) and the 3′ end of the C strand (lanes 7–12) were used. The samples contained the following amounts of Sp1(167*): none (lanes 3 and 10), 15 pmol (lanes 4 and 9), 30 pmol (lanes 5 and 8), and 45 pmol (lanes 6 and 7). Lanes 1 and 13: intact DNA. Lanes 2 and 11: DNA sequencing (G+A) by the Maxam–Gilbert method. The enhanced cleavage site is indicated with a closed arrowhead on the side of each panel.

From the methylation interference experiment, it turned out that runs of five guanine bases (G(2), G(3), G(4), G(6), and G(5′)) on both strands in the sequence 5′-GGGCG-3′ dominate GC box recognition by Sp1(167*). The G(1), G(7), and G(9) bases are only weakly contacted by Sp1(167*). Indeed, the first and ninth positions are not highly conserved in Sp1-responsive promoters and can be replaced with thymine or adenine bases, respectively (Kadonaga et al., 1986; Briggs et al., 1986). The G(7) base in the GC box is exposed to DMS molecules (Gidoni et al., 1984; Jones et al., 1986), and hence this base is not likely to be recognized by the protein, despite the fact that this guanine base is highly conserved (Kadonaga et al., 1986). In an interference experiment using the metal-responsive element (MRE) of the mouse metallothionein I gene, the methylation of five guanines on both strands was also severely interfered with by the binding of Sp1 in a HeLa cell nuclear extract (Westin & Scaffner, 1988). Additional interferences were observed at the 5′-TCGGGCGGAGTGC-3′ site in the G strand. The inconsistencies with our results may be due to the use of a nuclear extract.

There are complicated observations for the fifth C(5)·G-(5′) base pair in the center of the GC box, namely, the absence of apparent methylation protection at the G(5′) base on the C strand (Figure 1), yet the presence of a clear interference with Sp1(167*) binding by premodification of the G(5′) base with DMS. Sp1 weakly protected the methylation of guanine

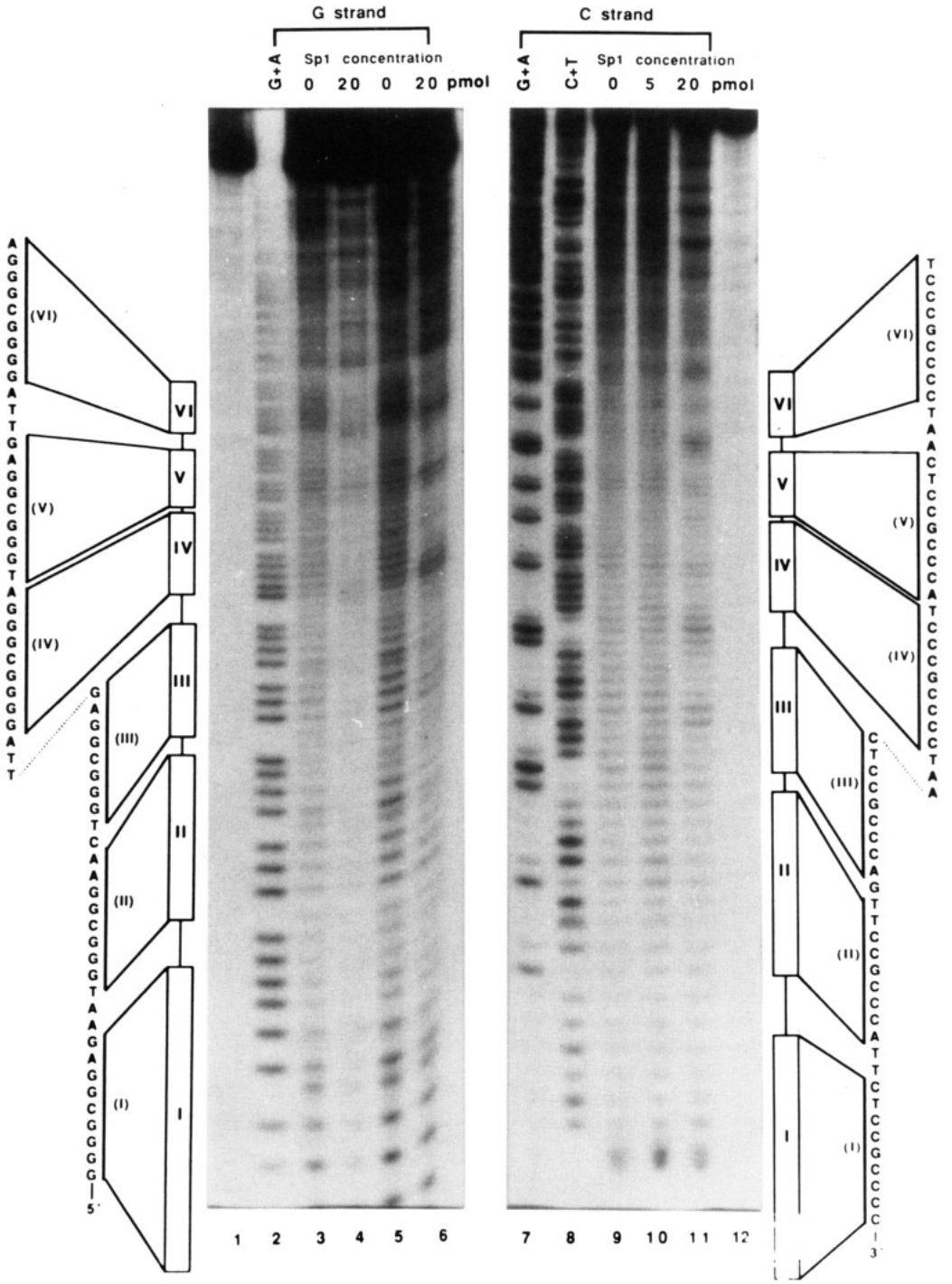

FIGURE 5: Footprinting analysis for Sp1(167*)-bound SV 40 early promoter DNA by $(OP)_2$–Cu(I) complex. DNA fragments labeled at the 5′ end of the G strand (lanes 1–6) and the 3′ end of the C strand (lanes 7–12) were used. The reaction samples (20 μL) contained the following amounts of Sp1(167*): 5 pmol (lane 10) and 20 pmol (lanes 4, 6 and 11). Cleavage reactions by the $(OP)_2$–Cu(I) complex were carried out for 2 min (lanes 3 and 4) or 4 min (lanes 5, 6, 9, 10, and 11). Lanes 2 and 7: DNA sequencing (G+A) by the Maxam–Gilbert method. Lanes 8: DNA sequencing (C+T). Lanes 1 and 12: intact DNA.

bases in the C strand of U2 RNA gene enhancer (Janson et al., 1987). Site-specific incorporation of a methyl group into the fifth position of C(5) and/or C(6′) does not affect GC box recognition of Sp1 (Harrington et al., 1988). Nevertheless, the C(5)·G(5′) base pair is quite well conserved in the center of a GC box. These results suggest that particular induced fitting of the GC box DNA occurs upon Sp1 binding. Indeed, DNA binding of Zif268 is interfered with by premethylation of guanines in the secondary strand 5′-C*G*CCCC*G*C-3′, whereas specific contact of these guanine bases with the protein is not detected in the crystal structure of Zif268–DNA complex (Christy & Nathans, 1989; Pavletich & Pabo, 1991).

Implications for the Base Recognition Mode of Sp1. The crystal structure of the Zif268–DNA complex reveals that (i) the three-finger domain associates with a nine-base-pair sequence with triplet subsites and is arranged in antiparallel orientation to the primary interacting strand of DNA, (ii) there is one-to-one correspondence between three base pairs in a subsite and the three base-contacting residues in a helical region of each finger domain, and (iii) the protein makes contacts only with guanine bases and uses only arginine and histidine side chains for their recognition (Pavletich & Pabo, 1991). Sp1 and serum-inducible proteins Zif268 and Krox20 have several common features with respect to amino acid

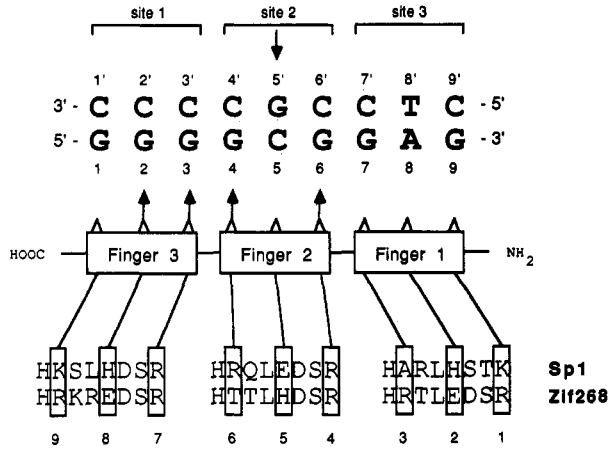

FIGURE 6: Plausible interaction mode of Sp1 with GC box DNA. Arrows indicate strong guanine contacts of Sp1 shown in Figure 2. Amino acid residues predicted by reference to the previous study of Zif268 (Pavletich & Pabo, 1991) to be involved in specific base recognition are enclosed with boxes.

sequences of the three-finger domains and their target sequences of DNA (Nardelli et al., 1991). The middle finger (finger 2) of Sp1 has the same set of three key residues (R(6), E(5), and R(4)) as those in the first and the third finger of Zif268 (see Figure 6). In addition, The 5′-GCG-3′ subsite recognized by these two fingers in Zif268 is exceptionally well conserved in divergent GC box DNAs (Kadonaga et al., 1986).

On the basis of this information and the present results, we predict a plausible interaction mode of Sp1 with the GC box as shown in Figure 6. The G(2) and G(3) bases in site 1 probably contact histidine (H(8)) and arginine (R(7)) residues in the third finger (finger 3), and the two guanine bases, G(4) and G(6), in site 2 also associate with the R(6) and R(4) residues, respectively, in finger 2. Surprisingly, specific contacts with all of the critical four guanines (G(2), G(3), G(4), and G(6)) in the G strand are provided by the two C-terminal finger domains (finger 2 and finger 3). On the contrary, the G(7) and G(9) bases in site 3 weakly contact Sp1(167*). When the eighth base in site 3 is guanine, as in boxes IV and VI of SV40 early promoter, methylation of this base is promoted, as in the previous study (Gidoni et al., 1984) and in our observations (see Figure 1). Therefore, the first finger (finger 1) seems to contribute to DNA base contacts less significantly than fingers 2 and 3. Several lines of evidence support our prediction: (i) deletion of one-half of the amino acid residues in finger 3 greatly reduces the binding affinity of Sp1 (Courey & Tjian, 1988); (ii) removal of 1 equivalent of zinc ion from Zn_3Sp1(167*) does not abolish the DNA binding affinity of Sp1 severely (Kuwahara & Coleman, 1990; Kuwahara et al., 1991); (iii) a very recent primer extension/mobility shift assay for Sp1 peptides reveals that the 5′ portion of the GC box DNA sequence (5′-GGGCG-3′) contributes more strongly to the total binding energy than the 3′ portion (5′-GGGC-3′) (Kriwacki et al., 1992). DMS protection and interference experiments suggest that the G(1) base is not as well recognized by Sp1(167*) as the above-mentioned four guanine bases. In addition, the first G base can be replaced by thymine. Therefore, the lysine residue (K(9)) at the base-contacting position in finger 3 might not be an effective determinant for guanine recognition, although lysine is well utilized in G recognition by other zinc finger proteins, such as GAL4 and glucocorticoid receptor, and by λ repressor (Marmorstein et al., 1992; Luisi et al, 1991; Jordan & Pabo, 1988).

DNA Structural Alteration by Sp1 Binding. In this study, we observed conspicuous enhancement of methylation at the G(7) base (Figure 1) and new cleavage at the 5′-G(7)A(8)-3′ site by the PEM–Fe complex (Figures 3 and 4). In general, enhanced reactivity of the DMS molecule with guanine bases in DNA complexed with protein is explained by either protein binding-induced structural alteration of the DNA or the formation of a so-called hydrophobic pocket by the protein on the DNA (Wissmann & Hillen, 1991). Although the possibility of changes in other helical parameters, such as propeller twist and slide, cannot be excluded, these phenomena might be interpreted as follows: significant increase of helical twist angle at the base-pair step G(6)-G(7) in the sequence 5′-GGA-3′ and concomitant relative decrease of helical twist angle in the next base-pair step, G(7)-A(8), mediated by Sp1-(167*) binding. The above-mentioned structural perturbation will cause base pair 7 to slip away from base pair 6 and make the N7 nitrogen atom of G(7) more accessible to the DMS molecule in the major groove of the G(6)-G(7) step. Protein-induced methylation enhancement is also observed in the λ repressor–operator interaction (Humayun et al., 1977). Of nine guanine bases in the O_L1 operator (17 bp), the methylation of only one guanine in the center is promoted by binding of λ repressor. The crystal structure of the protein–DNA complex demonstrates that the base-pair step on the 5′ side of the guanine has an unusually large helical twist angle (47.0°, av 34.4°) and that the G-C base pair has the highest propeller twist angle in the O_L1 site (21.2° av 11.6°). In addition, this guanine base makes no contacts with the λ repressor (Jordan & Pabo, 1988). These structural data would support our explanation.

The PEM–Fe complex specifically cleaves pyrimidine bases in the 5′-GC and 5′-GT sequences. The adenine base at the 5′-GA site is usually a less preferred cutting site. The convex environment comprising the 2-amino group of guanine and the 2-carbonyl group of the pyrimidine base in a 5′-G-pyrimidine site appears to be preferable for the primary binding of metallobleomycin and a snug fit to the bithiazole–carboxamide moiety of the drug (Kuwahara & Sugiura, 1988). The relative decrease of helical twist in the base-pair step G(7)-A(8) improves the stacking of these two base pairs at the 5′-GA site. This may also achieve the protruded environment at this step and the comfortable steric geometry between amino group of G(7) and N3 nitrogen atom of A(8) which results in new cutting by the PEM–Fe complex.

Global helical distortion of DNA, such as unwinding, does not appear to occur in Sp1(167*) binding, because none of the guanine bases were modified with diethyl pyrocarbonate, which specifically modifies the N7 positions of guanine and adenine bases in the unwound region of DNA (unpublished results). In addition, significant enhancement of cleavage by the $(OP)_2$–Cu(I) complex was not detected in Sp1(167*) binding (Figure 5), as seen in the *Escherichia coli* RNA polymerase–DNA complex (Spassky & Sigman, 1985).

In conclusion, we demonstrate that (i) five guanine bases on both strands are important for GC box recognition by Sp1, (ii) the contributions of individual zinc finger domains to specific binding of Sp1 to GC box are nonequivalent, and (iii) the region near the 3′ end of GC box is locally distorted by the protein binding. These features are distinct from those of the three-zinc-finger proteins Zif268 and Krox20, which are closely related to Sp1. The present results would provide good information for DNA interaction by the more extended zinc finger, such as TFIIIA.

ACKNOWLEDGMENT

We thank Dr. J. E. Coleman (Yale University, New Haven, CT) for providing plasmids containing Sp1 fragment and for comments on the manuscript.

REFERENCES

Briggs, M. R., Kadonaga, J. T., Bell, S. P., & Tjian, R. (1986) *Science 234*, 47–52.

Chavrier, P., Vesque, C., Galliot, B., Vigneron, M., Dolle, P., Duboule, D., & Charnay, P. (1990) *EMBO J. 9*, 1209–1218.

Christy, B., & Nathans, D. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 8737–8741.

Courey, A. J., & Tjian, R. (1988) *Cell 55*, 887–898.

Dynan, W. S., & Tjian, R. (1985) *Nature 316*, 774–778.

Gidoni, D., Dynan, W. S., & Tjian, R. (1984) *Nature 312*, 409–413.

Harrington, M. A., Jones, P. A., Imagawa, M., & Karin, M. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 2066–2070.

Humayun, Z., Kleid, D., & Ptashne, M. (1977) *Nucleic Acids Res. 4*, 1595–1607.

Janson, L., Bark, C., & Pettersson, U. (1987) *Nucleic Acids Res. 15*, 4997–5016.

Jones, K. A., Kadonaga, J. T., Luciw, P. A., & Tjian, R. (1986) *Science 232*, 755–759.

Jordan, S. R., & Pabo, C. O. (1988) *Science 242*, 893–899.

Kadonaga, J. T., Jones, K. A., & Tjian, R. (1986) *Trends Biochem. Sci. 11*, 20–23.

Kadonaga, J. T., Carner, K. R., Masiarz, F. R., & Tjian, R. (1987) *Cell 51*, 1079–1090.

Kriwacki, R. W., Schultz, S. C., Steitz, T. A., & Caradonna, J. P. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 9759–9763.

Kuwahara, J., & Sugiura, Y. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 2459–2463.

Kuwahara, J., & Coleman, J. E. (1990) *Biochemistry 29*, 8627–8631.

Kuwahara, J., Futamura, M., Sogo, N., Yonezawa, A., & Sugiura, Y. (1991) *Nucleic Acids Symp. Ser. 25*, 121–122.

Letovsky, J., & Dynan, W. S. (1989) *Nucleic Acids Res. 17*, 2639–2653.

Luisi, B. F., Xu, W. X., Otwinowski, Z., Freedman, L. P., Yamamoto, K. R., & Sigler, P. B. (1991) *Nature 352*, 497–505.

Marmorstein, R., Carey, M., Ptashne, M., & Harrison, S. C. (1992) *Nature 356*, 408–414.

Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499–560.

Nardelli, J., Gibson, T. J., Vesque, C., & Charnay, P. (1991) *Nature 349*, 175–178.

Pavletich, N. P., & Pabo, C. P. (1991) *Science 252*, 809–817.

Siebenlist, U., & Gilbert, W. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 122–126.

Spassky, A., & Sigman, D. S. (1985) *Biochemistry 24*, 8050–8056.

Thiesen, H.-J., & Bach, C. (1990) *Nucleic Acids Res. 19*, 3203–3209.

Westin, G., & Schaffner, W. (1988) *EMBO J. 7*, 3763–3770.

Wissmann, A., & Hillen, W. (1991) *Methods Enzymol. 208*, 365–379.