Volume 17 Number 7 1989 Nucleic Acids Research

Measurement of the binding of transcription factor Sp1 to a single GC box recognition sequence

John Letovsky and William S.Dynan*

Department of Chemistry and Biochemistry, Campus Box 215, University of Colorado, Boulder, CO 80309, USA

Received December 14, 1988; Revised and Accepted March 1, 1989

ABSTRACT

The equilibrium constant was determined for the binding of the transcription factor Sp1 to a single consensus GC box DNA recognition site, (5'-GGGGCGGGGC-3'). For these experiments, single copies of the recognition site were synthesized and cloned in a standard plasmid background. Binding was measured either by a footprinting assay modified so that the binding reaction was at equilibrium, or by a gel mobility shift assay. The concentration of active Sp1 in the reactions and the dissociation constant were determined by computer-assisted fitting to theoretical curves. Values for the dissociation constant obtained in different experiments ranged from 4.1 X 10^{-10} M to 5.3 X 10^{-10} M. Several variants of the consensus recognition site were also tested. An A-substituted variant (5'-GGGGAGGGGC-3') and a T-substituted variant (5'-GGGGTGGGGC-3') were bound 3-fold and 6-fold more weakly than the consensus site, respectively. A G-substituted variant (5'-GGGGGGGGGC-3') was bound at least 30-fold more weakly than the consensus site. These findings help distinguish between alternative models for Sp1-DNA recognition. They are consistent with the presence of specific hydrogen-bond contacts between Sp1 and the central C-G base pair, but provide no particular evidence to support a model where local DNA structure is the dominant factor in the interaction.

INTRODUCTION

The transcription factor Sp1 was first detected as an activity in HeLa cell extracts that selectively stimulated transcription in vitro from Simian Virus 40 (SV40) promoters (1). Sp1 binds to an essential region of the SV40 early promoter, approximately 50 to 110 bp upstream of the RNA start site, and this binding has been shown to be essential for activation of transcription (2). Subsequent studies have shown that Sp1 binds to and activates transcription from a number of other viral and cellular promoters (3,4,5,6). The Sp1 protein has been extensively purified, and activity has been shown to be associated with two related polypeptides with apparent molecular weights of 95 kd and 105 kd (7). A partial cDNA clone of Sp1 was recently obtained, and analysis of the derived amino acid sequence suggests the presence of three zinc finger motifs within the DNA-binding portion of the molecule (8).

Comparison of the different sequences recognized by Sp1 suggests that the basic recognition unit is a 10 bp motif, the GC box, with a consensus sequence 5'-GGGGCGGGGC-3' (5). The number and spacing of these sequences varies widely between promoters, and the sequence may be present in either orientation, relative to the direction of mRNA synthesis. Many studies have noted that the apparent affinity for Sp1 varies between different sites. In general, however, these

binding studies have not been quantitative, and a rigorous interpretation of the data has not been attempted.

In the work presented here, Sp1 recognition oligonucleotides of the desired sequence were inserted in a standard plasmid background, and standard footprinting and gel mobility shift assays were modified so that the binding reaction was as close to equilibrium as possible. These studies used both the consensus Sp1 recognition site and variants with all 3 possible substitutions in the central nucleotide of the GC box (5'-GGGGCGGGGC-3').

MATERIALS AND METHODS

Preparation of DNA fragments containing a single GC box.

Complementary synthetic oligonucleotides containing the consensus sequence for Sp1 binding (5'-GATCGGGGCGGGGC-3' and 5'-GATCGCCCCGCCCC-3') or one of 3 mutant sequences 5'-GATCGGGGAGGGGC-3' and 5'-GATCGCCCCTCCCC-3'; (5'-GATCGGGGTGGGGCGATC-3' and 5'-GATCGCCCCACCCCGATC-3'; 5'-GATCGGGGGGGGGGC-3' and 5'-GATCGCCCCCCCCCC-3') were annealed and ligated into the Bam HI site within the polylinker of the pUC19 cloning vector. Plasmids containing a single insert in the desired orientation (G-rich strand is in the anti-sense of the lac Z gene) were identified by DNA sequencing and designated pGC1, pGA1, pGT1, and pGG1 respectively. To prepare fragments for binding assays, CsCl-purified plasmid DNA was digested with Hind III and treated with calf intestinal phosphatase. The reaction was terminated by adjustment to 10 mM EGTA and incubation at 65° C for 10 minutes. The DNA was digested with Pvu II and subjected to preparative polyacrylamide gel electrophoresis. The desired 155 bp fragment was recovered by electroelution, followed by membrane filtration (0.22 μm pore size) and precipitation from ethanol. The DNA was resuspended in water and quantitated by UV spectrophotometry. An absorbance of 1.0 at 260 nm was assumed to represent a concentration of 50 μg/ml. The isolated fragments were stored frozen at -20° C.

DNA fragments were 5'-end-labeled with [γ-^{32}P] ATP (ICN, 1.5 x 10^7dpm/ pmol) and T4 polynucleotide kinase. The Hind III end was labeled selectively, as the blunt, phosphorylated Pvu II end was not a substrate for the kinase. Reactions were terminated by incubating at 65° C for 15 minutes, diluted with water and stored at 4° C. This protocol avoided the need for ethanol precipitation after labelling, thus assuring that the concentration of the DNA remained unchanged after the initial determination.

Purification of Sp1.

A 32 liter culture of HeLa cells was harvested and a nuclear extract was prepared as described (9), except that the ammonium sulfate precipitation step was omitted. In preparation for chromatography, extracts were diluted to the same conductivity as TM buffer containing 0.15 M KCl (TM buffer: 50mM Tris-HCl pH 7.9, 12.5 mM $MgCl_2$, 1mM EDTA, 1mM dithiothreitol and 20% (v/v) glycerol). A 32 ml heparin-agarose column was prepared as described (1,10) and equilibrated with TM buffer containing 0.15 M KCl. The column was loaded and washed with TM buffer containing 0.15

M KCl. Bound protein was eluted with TM buffer containing 0.4 M KCl. Fractions containing the peak of protein were pooled, mixed with 800 µg d(I-C), diluted to a conductivity equivalent to TM buffer containing 0.15 M KCl, and used for DNA squence affinity chromatography. Affinity columns were prepared as described by Kadonaga and Tjian (11). Heparin-agarose fractions were passed first over a 4 ml AP-1 affinity column (oligonucleotide sequence: 5'-GATCATGGTTGCTGACTAATTGAGA-3'). Material flowing through this column was applied to a 4 ml Sp1 affinity column (oligonucleotide sequence 5'-GATCGGGGCGGGGC-3'). Sp1 activity was eluted from the column by a linear gradient of 0.15M KCl to 1.2 M KCl in buffer Z (Buffer Z: 25mM Hepes pH 7.8, 12.5 mM $MgCl_2$, 1mM dithiothreitol, 20% (v/v) glycerol and 0.1% (v/v) Nonidet P-40 (NP-40)). Fractions were assayed for Sp1 activity by DNAse I footprinting. Peak fractions typically eluted at 0.3-0.5 M KCl. These were pooled, diluted, and applied to the column for a second round of chromatography. Binding studies were conducted using pooled fractions from the second affinity column.

Equilibrium DNAse I Footprinting Method.

Reactions contained 7.5 x 10^{-11} M DNA fragment, 10 µg/ml poly d(I-C), 10 mM Hepes [K^+] pH 7.8, 1 mM $CaCl_2$, 5 mM $MgCl_2$, 100 mM KCl, 0.2 mM dithiothreitol, 0.02% NP40, and 4% glycerol. Sp1-containing fractions were added last in the amounts indicated in the figure legends. The final volume was 200 µl. The reactions were incubated at 10° C for 20 min., which preliminary experiments showed was sufficient to allow the binding reaction to reach equilibrium. DNAse I was added in a small volume (5 µl) to a final concentration of 18 nM, and incubation was continued at 10° C for one min. The reaction was terminated by addition of 180 µl of a solution containing 1% SDS, 200 mM NaCl, 20 mM EDTA and 100 µg/ml yeast tRNA. Samples were extracted with an equal volume of phenol: chloroform (1:1, v:v) and precipitated by addition of 1 ml ethanol. Pellets were resuspended in a small volume of redistilled formamide containing marker dyes, incubated in a boiling water bath for 2 minutes, loaded immediately onto a thin, denaturing polyacrylamide gel, and subjected to electrophoresis. Gels were dried on Whatman 3MM filter paper and exposed to Kodak XAR-5 X-ray film. Binding was quantitated as described in Results.

This protocol incorporates several modifications from the methods used in earlier studies of Sp1-DNA binding (1, 9, 12). Most notably, the temperature and ionic conditions were held constant throughout the binding and DNAse digestion periods, rather than subjecting the reaction to a jump in temperature and ionic conditions at the time DNase is added.

Gel Mobility Shift Binding Assay.

Standard reactions contained 2.9 X 10^{-11} M DNA fragment, 40 µg/ ml poly d(I-C), 7.1 mM Hepes [K^+] pH 7.8, 3.6 mM $MgCl_2$, 100 mM KCl, 5.7% glycerol and 0.03% NP-40. Sp1-containing fractions were added last as indicated in the figure legends. The final volume was 35 µl. The reactions were incubated at 10° C for 20 min. to allow the binding reaction to reach equilibrium. 2 µl of a solution containing 48% glycerol and 0.05% bromphenol blue dye was added to the reaction, and samples were loaded immediately on a native 5% polyacrylamide gel (49:1 (w/w) acrylamide:bis-

acrylamide). The gel and electrophoresis buffer contained 40 mM Tris base, 306 mM glycine and 0.1% NP-40. Samples were electrophoresed at 4° C for 1.75 hours at 260 V. The gel was dried on Whatman 3MM paper and exposed to Kodak XAR 5 X-ray film. Portions of the gel were excised and radioactivity was determined by liquid scintillation counting, as described in Results. It was necessary to include NP-40 in the gel to resolve the bound complexes into discrete bands. Presumably, the detergent blocks aggregation of the proteins present in the sample.

Calculation of a Binding Constant for Sp1

It was assumed that Sp1 bound with 1:1 stoichiometry to the single recognition site present in each DNA molecule. Using this assumption, the binding may be described by equations 1 and 2.

$$(1) \qquad P + D \underset{}{\overset{K}{\rightleftharpoons}} PD$$

$$(2) \qquad K\,[P]\,[D] = [PD]$$

where P represents Sp1, D represents the GC box-containing DNA fragment, PD represents the Sp1-DNA complex, and K is the association constant that governs the reaction. One can express the total Sp1 concentration in the reaction, $[P_{tot}]$, simply as $[P_{tot}] = [P] + [PD]$. [PD] can be substituted from (2.) and the equation rearranged to give

$$(3) \qquad [P_{tot}] = [P]\,(1 + K\,[D]).$$

The concentration of total DNA in the binding reaction can be expressed similarly as $[D_{tot}] = [D]\,(1 + K\,[P])$ or by solving for [D]

$$(4) \qquad [D] = \frac{[D_{tot}]}{1 + K[P]}$$

Substituting the value of [D] from (4) into (3) gives

$$(5) \qquad [P_{tot}] = [P]\left[1 + \frac{K[D_{tot}]}{1 + K[P]}\right]$$

Expressed as a quadratic, (5) becomes

$$(6) \qquad K[P]^2 + (K[D_{tot}] - K[P_{tot}] + 1)[P] - [P_{tot}] = 0.$$

Using the quadratic formula, this equation can then be solved for [P] in terms of K, $[P_{tot}]$, and $[D_{tot}]$. In our experiments, the value for $[D_{tot}]$ was known, but the values for K and $[P_{tot}]$ were not.

The ratio of bound to total DNA was determined experimentally under various conditions. This ratio, represented as θ, can be written as

$$(7) \qquad \theta = \frac{[PD]}{[PD] + [D]}$$

Substituting [PD] from (2) gives

$$(8) \qquad \theta = \frac{[P]K}{1 + [P]K}$$

Using equation (8), the experimentally determined values for θ can be related to the quadratic (6). Values for the two unknown parameters $[P_{tot}]$ and K can then be adjusted to give the best fit to the

two equations using a non-linear least squares fitting program. Binding is described in the text in terms of the dissociation constant, K_D.

$$(9) \qquad K_D = 1/K.$$

RESULTS

Determination of the equilibrium constant for the binding of Sp1 to a single GC box.

The plasmid pGC1 contained a synthetic insert with one copy of the consensus GC box recognition sequence. A preliminary footprinting analysis was carried out to determine whether the insert was recognized by Sp1 as an authentic binding site. The region protected by Sp1 was aligned with Maxam-Gilbert sequence markers (data not shown) and the 21 bp protected sequence was determined to be "GGGGATCGGGGCGGGGCGATC" (GC box is underlined). Both the size of the footprint and its location centered on the GC box were consistent with expectations. Moreover, the amount of Sp1 required to protect the pGC1 fragment was similar to that required to protect the Sp1-binding region of SV40 (data not shown).

The dissociation constant for the Sp1-DNA interaction was determined in two steps, using the equilibrium footprinting protocol (see Materials and Methods). In the first step, the concentration of Sp1 was held constant, and total DNA concentration was increased by mixing labelled probe with known amounts of an unlabelled, but otherwise identical, DNA fragment. Addition of as much as a 10-fold excess of unlabelled DNA had no visible effect on binding, but as the concentration of the DNA was increased further, the protection of the labeled fragment was relieved (Fig 1A.). This effect was quantitated by excising blocks of bands from the gel and determining the amount of radioactive label present by liquid scintillation counting. Block I is within the protected region; block II is outside and serves as an internal control to compensate for any small differences in the amount of label loaded in each lane. After subtraction of background, the ratio of counts in block I to block II was computed. This ratio was then scaled to give values for fractional saturation of the binding site. It was assumed that at a fractional saturation of 1.0, there would be no radioactivity above background in block I.

We found that essentially the same results were obtained when radioactivity in each block was determined using an AMBIS radioanalytic imaging system, which uses a spatially resolved gas ionization detector (Automated Microbiology systems, San Diego CA). Densitometry of an autoradiographic exposure was less satisfactory as a method of determining radioactivity, perhaps because of the limitations of the available instrumentation or nonlinearity of the film response.

A theoretical curve was fit to the fractional saturation data by adjusting parameters representing total Sp1 concentration and the binding constant. The concentration of active Sp1 in this preparation was determined to be $(4.9 \pm 0.1) \times 10^{-8}$ M. The uncertainty indicated here and elsewhere was calculated by standard linear approximation (13). The accuracy of the Sp1 concentration determined by this method is relatively good and is limited principally by the accuracy with which DNA fragment concentration can be determined experimentally. By contrast, the fit between the

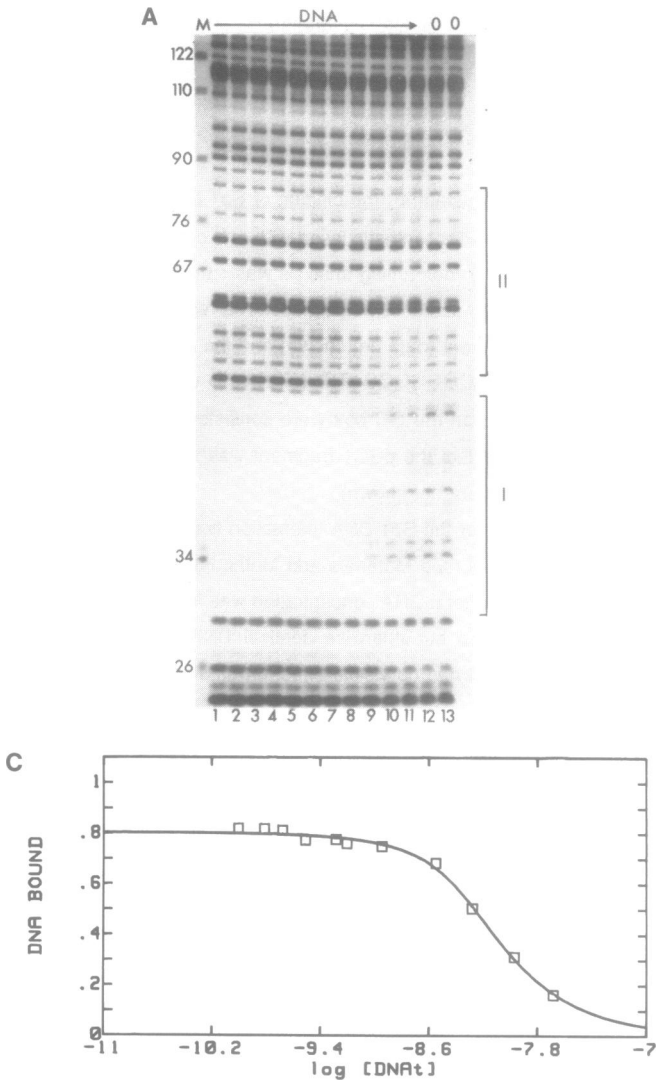
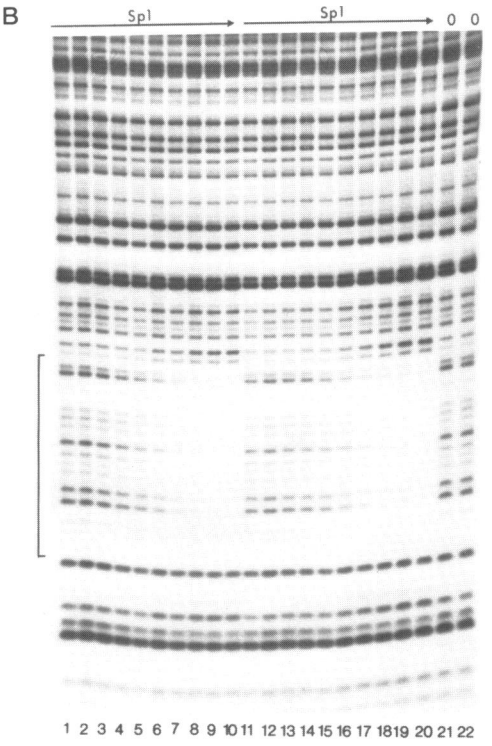
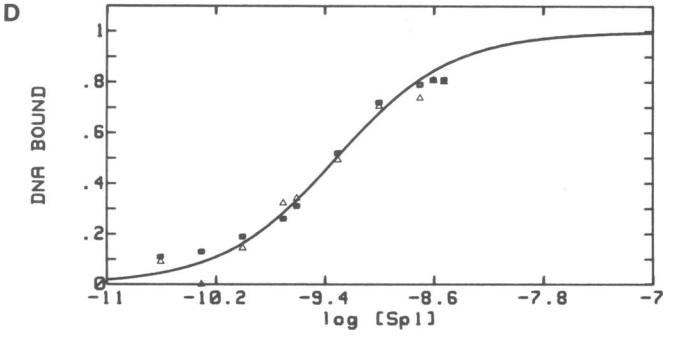


Figure 1. Determination of the dissociation constant for Sp1-DNA complexes using the equilibrium DNAse I footprinting method. The DNA fragment used in this experiment was derived from pGC1 and contained a single consensus Sp1 recognition site. Equilibrium footprinting was carried out as described in Materials and Methods. A. Titration with increasing amounts of DNA. Reactions in lanes 1-11 contained15 µl of affinity purified Sp1; reactions in lanes 12 & 13 were controls that contained no Sp1. All reactions contained 1 X 10^{-10} M of the labeled DNA fragment, and were supplemented with the identical non-labeled DNA fragment to the following total DNA concentrations in lane1-11, respectively: 1.00 x 10^{-10} M, 1.50 x 10^{-10} M, 2.03 x 10^{-10} M, 3.06 x 10^{-10} M, 5.12 x 10^{-10} M, 6.15 x 10^{-10} M, 1.13 x 10^{-9} M, 2.68 x 10^{-9} M, 5.25 x 10^{-9} M, 1.04 x 10^{-8} M, 2.07 x 10^{-8} M. Binding was quantitated by excising and counting blocks I and II as described in the text. Lane M contained Hpa II digested pBR322 DNA markers. B. Titration with increasing





amounts of affinity-purified Sp1. All reactions contained 7.5 X 10^{-11} M end-labeled DNA fragment. Sp1 concentrations in lanes 1-20, respectively, were: 2.47 x 10^{-11} M, 4.93 x 10^{-11} M, 9.86 x 10^{-11} M, 1.97 x 10^{-10} M, 2.47 x 10^{-10} M, 4.93 x 10^{-10} M, 9.86 x 10^{-10} M, 1.97 x 10^{-9} M, 2.47 x 10^{-9} M, 2.96 x 10^{-9} M, 2.47 x 10^{-11} M, 4.93 x 10^{-11} M, 9.86 x 10^{-11} M, 1.97 x 10^{-10} M, 2.47 x 10^{-10} M, 4.93 x 10^{-10} M, 9.86 x 10^{-10} M, 1.97 x 10^{-9} M, 2.47 x 10^{-9} M, 2.96 x 10^{-9} M. No Sp1 was added to reactions shown in lanes 21 & 22. Protected region is bracketed at left. Binding was quantitated as above. C. Quantitation of data in panel A; theoretical curve was generated as described in Materials and Methods. Concentration of active Sp1 in the affinity-purified preparation was determined to be (4.9 ± 0.1) x 10^{-8} M. D. Quantitation of data in panel B. The K_D for binding to single consensus site was determined to be (4.6 ± 0.3) x 10^{-10} M.

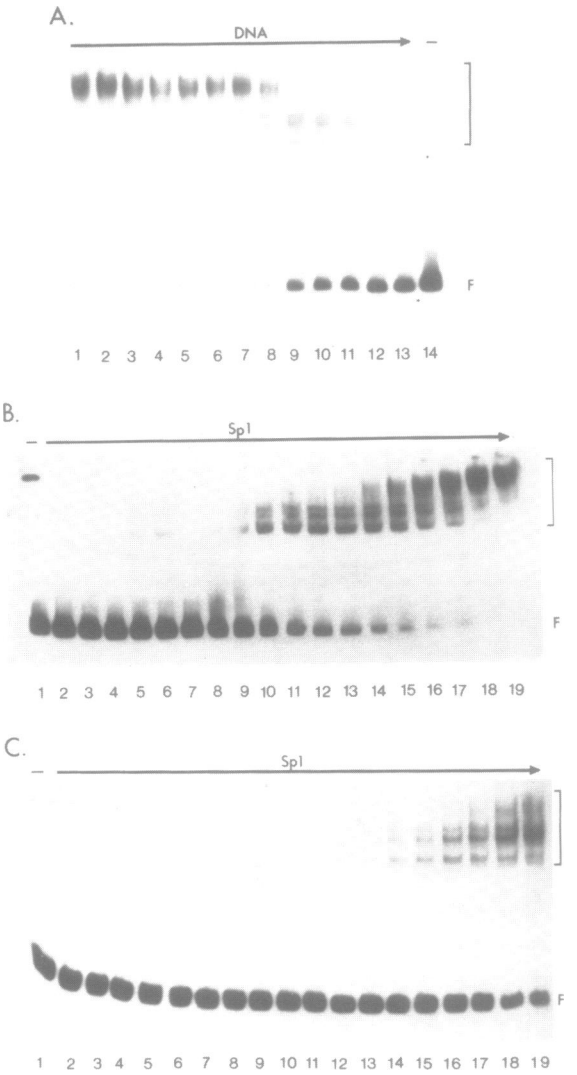
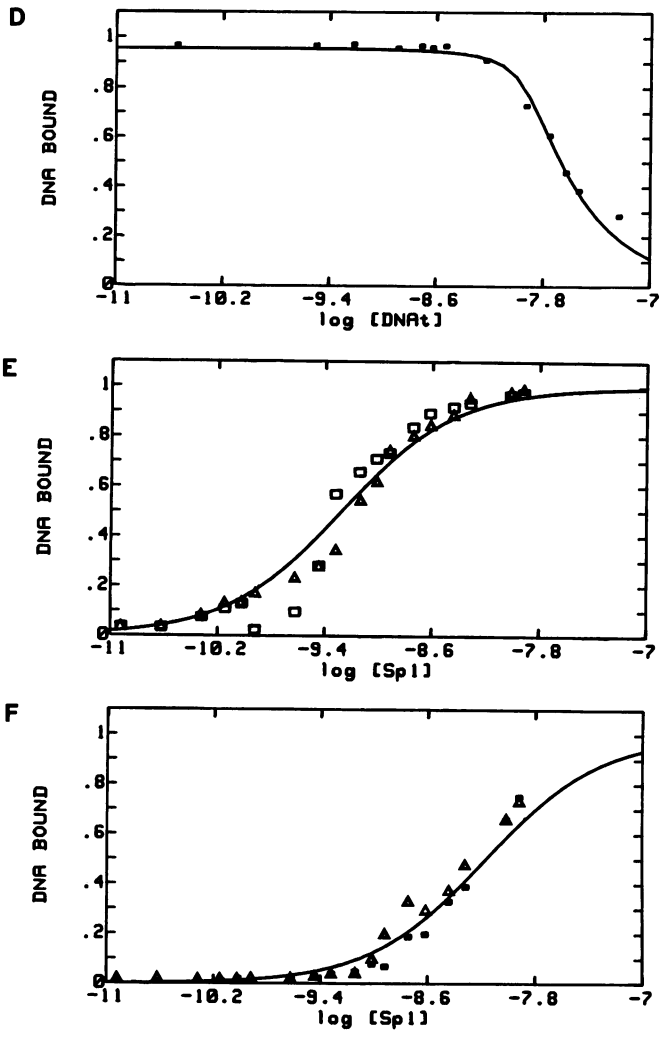


Figure 2. Determination of the dissociation constant for Sp1-DNA complexes using the gel mobility shift method. Gel mobility shift assay was carried out as described in Materials and Methods. A. Titration with increasing amounts of DNA. Reactions in lanes 1-13 contained 10 µl affinity purified Sp1; the reaction in lane 14 was a control that contained no Sp1. All reactions contained 2.85 X 10^{-11} M of the labeled pGC1 DNA fragment containing a single consensus Sp1 recognition site, and were supplemented with the identical non-labeled DNA fragment to the following total DNA concentrations in lanes 1-13, respectively: 2.86 x10^{-11} M, 3.23 x10^{-10} M, 6.17 x 10^{-10} M, 1.20 x 10^{-9} M, 1.80 x 10^{-9} M, 2.40 x 10^{-9} M, 2.98 x 10^{-9} M, 5.93 x 10^{-9} M, 1.18 x 10^{-8} M, 1.77 x 10^{-8} M, 2.35 x 10^{-8} M, 2.94 x 10^{-8} M, 5.89 x 10^{-8} M. Binding was quantitated by excising regions of the gel containing Sp1/DNA complexes (bracket) and free DNA fragment (F) as



described in the text. B. Titration with increasing amounts of affinity-purified Sp1. All reactions contained 2.85 X 10^{-11} M end-labeled pGC 1 DNA fragment. No Sp1 was added to the reaction in lane 1. Sp1 concentrations in lanes 2-19, respectively, were: 1.14 x 10^{-11} M, 2.28 x 10^{-11} M, 4.56 x 10^{-11} M, 6.84 x 10^{-11} M, 9.12 x 10^{-11} M, 1.14 x 10^{-10} M, 2.28 x 10^{-10} M, 3.42 x 10^{-10} M, 4.56 x 10^{-10} M, 6.84 x 10^{-10} M, 9.12 x 10^{-10} M, 1.14 x 10^{-9} M, 1.71 x 10^{-9} M, 2.28 x 10^{-9} M, 3.42 x 10^{-9} M, 4.56 x 10^{-9} M, 9.12 x 10^{-9} M, 1.14 x 10^{-8} M. Regions containing Sp1/ DNA complexes (bracket) and free DNA fragment (F) are indicated. C. Titration with Sp1; experiment is the same as that shown in panel B except the labeled DNA is a 141 bp fragment of pUC19 that lacks the oligonucleotide insert. D. Quantitation of data in panel A. Concentration of active Sp1 in the affinity-purified preparation was calculated to be (4.0 ± 0.2) x10^{-8} M. E. Quantitation of data. Different symbols represent data from panel B and a companion experiment. The K_D for Sp1 binding to the single site DNA fragment was determined to be (5.3 ± 0.4) x 10^{-10} M. F. Quantitation of data. Different symbols represent data from panel C and a companion experiment.

theoretical binding curve and the experimental data is relatively insensitive to assumed values of K_D, meaning that the value of K_D obtained is quite imprecise. To better determine K_D, a separate titration was carried out where DNA concentration was held constant and the amount of Sp1 added to the reaction was varied (Figure 1B, 1D). A theoretical curve was fit to these data assuming that the Sp1 concentration determined from the first titration was correct and adjusting a single parameter representing K_D. In this experiment, a value for K_D of $(4.6 \pm 0.3) \times 10^{-10}$ M was obtained. In both the DNA and protein titration experiments, the theoretical curve could be aligned closely with the data.

Gel mobility shift assay using DNA containing a single GC box. As an alternative, independent method to measure the binding constant of Sp1, we used the gel mobility shift assay. Conditions in the binding reactions were adjusted to be reasonably close to those used in the equilibrium footprinting reactions, in order to facilitate direct comparison of the results obtained with the two techniques (see Materials and Methods). Following the 20 min. binding reaction, a small volume of loading dye was added and the reaction was loaded immediately onto a polyacrylamide gel. The presence of bound Sp1 reduced the mobility of the DNA probe, allowing the separation of bound and free DNA. The Sp1-DNA complexes were heterogeneous; possible interpretations of this heterogeneity will be discussed below. To estimate the amount of binding, the entire region that contained Sp1-bound DNA (Fig. 2A, 2B, 2C, brackets) was excised from each lane of the gel and radioactivity was determined by liquid scintillation counting. The region that contained the free probe was also excised and counted. The ratio of bound to total DNA was calculated from this data and used to determine binding parameters in the same way as in the footprinting experiments.

To estimate the concentration of active Sp1 in the affinity purified preparation, a titration was carried out where increasing amounts of DNA fragment were added (figure 2A, 2D). A value of $(4.0 \pm 0.2) \times 10^{-8}$ M, was obtained, which was about 20% lower than with the footprinting method. Using this value for Sp1 concentration, the K_D for Sp1 binding was determined by fitting a theoretical curve to the protein titration data (fig. 2B, 2E). The value for K_D was $(5.3 \pm 0.4) \times 10^{-10}$ M, which was about 15% larger than the value obtained in the footprinting experiments.

In a separate experiment, nonspecific binding of proteins in the Sp1 preparation was measured. The nonspecific DNA probe was derived from pUC vector DNA, and was identical to the single-site probe, except that it lacked the Sp1 binding site insert. Results are shown in Fig. 2C and 2F. Binding to the nonspecific fragment was weaker than to the specific fragment, and about 10 times as much protein was required to give equivalent levels of saturation. Although this experiment confirms that binding is dependent on the presence of the GC box insert, it also illustrates that significant nonspecific binding occurs at higher protein concentrations. The amount of nonspecific binding suggests that the gel mobility shift method is not appropriate for the characterization of mutant Sp1 binding sites that are more than a few-fold weaker than the consensus site.

The Sp1-DNA complexes that were obtained by the gel mobility shift technique were heterogeneous. Because several retained species were seen even at the lowest concentrations of

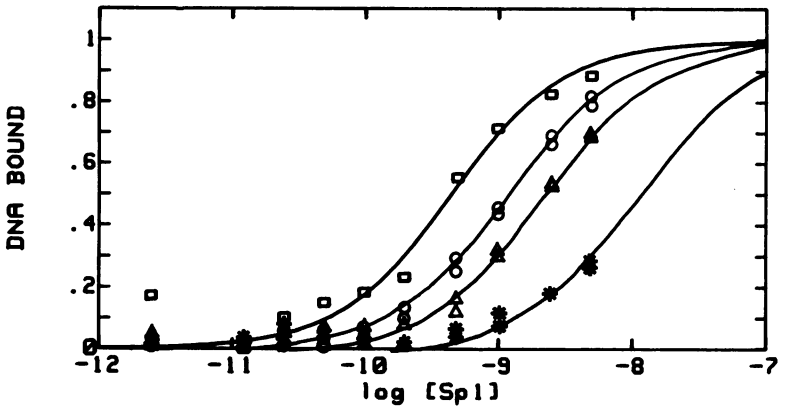


Figure 3. Determination of the dissociation constant for complexes between Sp1 and variant recognition sites. The equilbrium footprinting method was used. Consensus site from pGC1 (□), K_D = (4.1 ± 0.7) x 10^{-10} M; T-substituted variant from pGT1 (Δ), K_D = (2.3 ± 0.1) x 10^{-9} M; A-substituted variant from pGA1 (O), K_D = (1.2 ± 0.1)x10^{-9}M; G-substituted variant from pGG1 (*), K_D = (1.2 ± 0.04)x10^{-8}M.

Sp1, it is likely that there was pre-existing heterogeneity in the protein preparation, probably reflecting post-translational modifications (8). Additional more slowly migrating species appeared as the protein concentration was increased. This may reflect aggregation of Sp1 or the binding of multiple Sp1 molecules to the DNA at high concentrations of protein. It is unlikely that the additional bands represent a contaminant in the preparation, as the activity copurified with Sp1 at several steps (data not shown). The presence of more slowly migrating, possibly aggregated, complexes had relatively little influence on the derived K_D, because these complexes become significant only at protein concentrations well above the midpoint of the binding titration.

Synthetic oligonucleotides were also tested in the gel mobility shift assay (not shown). Binding appeared to be at least ten-fold weaker with the oligonucleotides than with the corresponding 155 bp pGC1 restriction fragment, suggesting that Sp1 was unable to interact normally with such short duplexes. Moreover, complementary oligonucleotides of 18 or 30 bp were found to convert between single-stranded and double-stranded conformations, even when maintained at 4° C and at high ionic strength, perhaps because of favorable intrastrand interactions available within the G-rich strand. These data suggest that binding studies using Sp1 recognition oligonucleotides should be interpreted cautiously.

Binding of Sp1 to variants of the consensus recognition site.

In order to assess the contribution of individual nucleotide residues in the Sp1 binding reaction, binding site variants were constructed containing each of the three possible base changes in the central C residue (5'-GGGGCGGGGC-3') of the consensus Sp1 recognition sequence. DNA fragments containing each of the single sites were isolated and used as probes in the equilibrium

footprinting assay. Results of a protein titration experiment are shown in Fig. 3.

Distinct differences in binding were seen between the variants. The consensus fragment had a K_D of (4.1 ± 0.7) x 10^{-10} M, in close agreement with the results of prior experiments. The A-substituted variant (5'-GGGGAGGGGC-3') had a K_D of (1.2 ± 0.05) x10^{-9} M, and the T-substituted variant (5'-GGGGTGGGGC-3') had a K_D of (2.3 ± 0.1) x 10^{-9} M. Thus, Sp1 bound to these two variants approximately 3- and 6-fold less tightly, respectively, than to the consensus site. The G-substituted variant showed very little ability to be bound by Sp1. Based on a small amount of protection seen with the largest amount of Sp1 tested, a K_D of (1.2 ± 0.04) x 10^{-8} M was estimated, about 30-fold weaker binding than with the consensus site. The true strength of the interaction may be even weaker, but could not be accurately measured because we were unable to saturate the binding site in these experiments.

DISCUSSION

Estimation of the K_D for Sp1-DNA complexes.

DNase I footprinting has been widely used to obtain qualitative information about the binding sites for mammalian transcription factors, including Sp1. It is a versatile method, as only a small amount of protein is required, specific and nonspecific binding are easily differentiated, and satisfactory results may be obtained using partially purified fractions. Moreover, recent work with purified prokaryotic DNA-binding proteins has shown that thermodynamically rigorous binding results can be obtained using the DNase I footprinting method, provided the reactions are run at equilibrium (14, 15).

DNA titration experiments using the equilibrium footprinting method showed that the Sp1 concentration in our affinity-purified preparation was approximately 4.9 X 10^{-8} M. Two independent protein titrations, also using the equilibrium footprinting method (Fig. 1, Fig. 3) gave values for the dissociation constant of 4.6 and 4.1 X 10^{-10} M, respectively. To confirm these results by an independent method, binding was measured by the gel mobility shift assay. The binding constant measured by gel retention was 5.3 X 10^{-10} M. The small difference between this value and the values from the footprinting experiments probably reflects the error intrinsic to the two methods.

The value obtained for K_D has several practical implications. Most Sp1 binding experiments that have been reported in the past have used DNA probe concentrations well below the K_D, with free protein in fairly large excess over DNA. These conditions are suitable for routine measurements of the amount of activity in a preparation and for experiments to compare the strength of binding to different sites. It is useful to know that under these conditions, the fractional saturation of the DNA is expected to be relatively insensitive to small changes in DNA concentration. Many experiments involve competition for binding to different sites; in these experiments, it is the molar concentration of competitor DNA, rather than the frequently reported "fold-excess" that controls the amount of observed competition. Finally, the total mass of protein that is present in complex

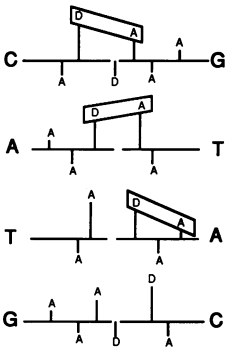


Figure 4. Top diagram: potential hydrogen bond contacts between Sp1 and central C-G base pair of consensus recogniton site (GGGGCGGGGC). D indicates donor, A indicates acceptor. Top edge, major groove, bottom edge, minor groove. Box indicates proposed contacts (see text). Remaining diagrams, from top to bottom: analagous potential contacts in A-variant (GGGGAGGGGC), T-variant (GGGGTGGGGC), and G-variant (GGGGGGGGGC). Drawing is based on analysis in reference 24 and styled after reference 25.

with DNA, which is important for certain types of experiments, can readily be increased by increasing the DNA concentration in the binding reaction.

The number of Sp1 molecules in the cell nucleus has been estimated to be on the order of 5,000-10,000 (5), corresponding to an intranuclear concentration of 1.25×10^{-7} to 2.5×10^{-7} M (16), well above the K_D determined in vitro. It is difficult to extrapolate from in vitro to in vivo circumstances, because of the many differences in ionic and other conditions. Moreover, theoretical arguments suggest that in vivo binding may sometimes be controlled by the amount of competition from nonspecific sites (17). Although these factors make quantitative predictions difficult, we can conclude that the value for K_D obtained in vitro is at least consistent with the ability of Sp1 to efficiently occupy specific binding sites in vivo.

The affinity of Sp1 for its recognition site appears to be somewhat weaker than the affinities of other purified mammalian transcription factors for their recognition sites. Although binding conditions were somewhat different in different studies, the adenovirus major late transcription factor (MLTF) has been reported to bind to its recognition site with a K_D of 10^{-10} to 4×10^{-12} M (16), the transcription factor NF-1 binds to its site with a K_D of 2.1×10^{-11} M (18), and the Drosophila heat shock transcription factor binds to its site with a K_D of 4×10^{-12} M (19). On the other hand, the affinity of Sp1 for its recognition site appears to be on the same order as the affinity of E. coli RNA polymerase for strong bacterial promoters (20).

Binding to Variant Recognition Sites.

Because most known Sp1 sites contain a C at the central position of the recognition consensus (5'-GGGGCGGGGC-3'), it seemed likely that substitution at this position might have a substantial effect on the binding. Such variants were therefore selected as the first to be tested in the

quantitative assay. The A-substituted variant (GGGGAGGGGC) was able to interact quite strongly with Sp1, with a dissociation constant only three times greater than that with the consensus site. This is consistent with reports that functional A-substituted sites occur in natural promoter sequences (6, 21). Somewhat surprisingly, the T-substituted variant (5'-GGGGTGGGGC-3') also showed a reasonable binding affinity, with a dissociation constant six times greater than for the consensus. The G-substituted variant (5'-GGGGGGGGGC-3') was by far the weakest site tested in our experiments, with at least a 30-fold difference from the consensus. This difference between the binding to the strongest and weakest sites corresponds to a minimum ΔG on the order of 2 kcal/ mol, consistent with the loss of one or two hydrogen bonds.

These data allow us to draw several tentative conclusions about the structural features that are recognized by the Sp1 protein when it contacts the central C-G base pair of the recognition site. First, the large difference between the consensus and G-substituted sites suggests that major groove contacts are important for recognition. The potential hydrogen bond contacts in the minor groove are insufficient to distinguish between C-G and G-C base pairs (24). This conclusion is consistent with prior methylation protection studies showing major groove contacts between Sp1 and DNA at several adjacent bases, and an apparent absence of any minor groove contacts in the recognition site(14). It is unlikely that the C-5 position of the cytosine is important for recognition, as two recent studies showed that Sp1 is able to bind to 5-methyl cytosine-substituted sites with an affinity indistinguishable from homologous cytosine-containing sites (22,23). Finally, the N-7 of the guanine in the central C-G base pair probably does not participate in a hydrogen bond to the protein, because it is not protected from dimethylsulfate modification (14). By elimination, there are two remaining functional groups in the major groove that may participate in hydrogen bonding. The 4-amino group of cytosine is a potential hydrogen-bond donor and the 6-carbonyl group of guanine is a potential hydrogen-bond acceptor (24). We suggest that Sp1 probably interacts with both of these functional groups.

This model allows rationalization of the relative affinity of Sp1 for the different variants of the consensus site (refer to Figure 4). The A-T base pair in the A-substituted variant contains an donor-acceptor pair analogous to that in the consensus site. The 6-amino group of adenine is the potential hydrogen bond donor, and the 4-carbonyl of thymine is the potential hydrogen bond acceptor. The availability of these contacts plausibly accounts for the strength of the binding to the A-substituted variant. The 6-amino group and N-7 of adenine allow analogous contacts to be made with a T-A base pair, although they are somewhat displaced in space. With the G-C transversion mutant, however, the 4-amino group of cytosine is displace to the wrong side of the major groove, and moreover, there is no potential for a pairwise interaction with a donor and acceptor analogous to those in the consensus site.

An alternative model for Sp1 binding is suggested by the unusual base composition and asymmettry of the consensus recognition site, which consists of a homopurine tract punctuated by the central C-G base pair. It is possible that this DNA has a distinctive local structure that serves as

a basis for Sp1 recognition. Such a structure would be expected to be disrupted in the G-variant, which lacks the interrupting pyrimidine in the purine rich strand. Indeed, this variant was bound very poorly by Sp1. However, the A-variant, which also lacks the interrupting pyrimidine, was bound relatively well, and the T-variant, which has the interrupting pyrimidine, was bound relatively more weakly. Thus, the data provide no particular evidence to support a model that Sp1 recognizes a distinctive local DNA structure.

ACKNOWLEDGMENTS

We thank Professor Stanley Gill and his associates, Charles Robert and Enrico DiCera, for derivation of the equations used to calculate the theoretical binding curve and for help in fitting binding parameters to the experimental data. We thank Cheryl Grosshans for oligonucleotide synthesis and for instruction in the use of the AMBIS radioanalytic imaging system. We also thank Henri Sasmor and Gerald Hertz for helpful discussions, and Professor Gill and Rhea-Beth Markowitz for comments on the manuscript. This work was supported by U.S. Public Health Service grant no. GM 35866.

*To whom correspondence should be addressed

REFERENCES

1. Dynan, W.S. and Tjian, R. (1983) Cell **32**, 669-680.
2. Dynan, W.S. and Tjian, R. (1983) Cell **35**, 79-87.
3. Jones, K.A., Yamamoto, K.R. and Tjian, R. (1985) Cell **42**, 559-572.
4. Dynan, W.S.,Sazer, S.,Tjian,R., Schimke,R.T. (1985) Nature **319**, 246-248.
5. Kadonaga, J. T., Jones, K. A., Tjian, R. T. (1986) Trends Biochem. Sci. **11**, 20-23.
6. Jones, K. A., Kadonaga, J. T., Luciw, P. A., and Tjian, R. T. (1986) Science **232**, 755-759.
7. Briggs, M.R., Kadonaga, J.T., Bell, S.P., Tjian, R. (1986) Science **234**, 47-52.
8. Kadonaga, J. T., Carner, K. R., Masiarz, F. R., and Tjian, R. (1987) Cell **51**, 1079-1080.
9. Dynan, W.S. (1987) In Setlow, J. (ed.) Genetic Engineering, Principles and Methods, Plenum Publishing, New York, vol. 9, pp. 75-87.
10. Davison, B. L., Leighton, T., and Rabinowitz, J. C. (1979) J. Biol. Chem. **254**, 9220-9226.
11. Kadonaga, J. and Tjian, R. T. (1986) Proc. Natl. Acad. Sci. USA **83**, 5889-5893.
12. Gidoni, D., Dynan, W. S., and Tjian, R. T. (1984) Nature **312**, 409-413.
13. Bevington, P. R. (1969) In Data Reduction and Error Analysis for the Physical Sciences, McGraw-Hill, New York.
14. Brenowitz, M., Senear, D. F., Shea, M. A., and Akers, G. K. (1986) Meth. Enzymol. **130**, 132-181.
15. Brenowitz, M., Senear, D.F., Shea, M.A., and Ackers, G. K. (1986) Proc. Natl. Acad. Sci. USA **83**, 8462-8466.
16. Chodosh, L. A., Carthew, R. W., and Sharp, P. A. (1986) Mol. Cell Biol. **6**, 4723-4733.
17. Lin, S. Y., and Riggs, A. D. (1975) Cell **4**, 107-111.
18. Rosenfeld, P. J., and Kelly, T. J. (1986) J. Biol. Chem. **261**, 1398-1408.
19. Wu,C., Wilson, S., Walker, B., Dawid, I., Paisley;, T., Zimarino, V.,, and Ueda, H. (1977) Science **238**, 1247-1233.
20. McClure, W. R., and Hoopes, B. C. (1987) In Reznikoff, W. S., Burgess, R. R., Dahlberg, J. E., Gross, C. A., Record, M. T., and Wickens, M. P. (eds.) RNA Polymerase and the Regulation of Transcription, Elsevier, New York, pp. 85-103.
21. Evans, T., DeChiara, T., and Efstradiatis, A. (1988) J. Mol. Biol. **199**,61-81.
22. Harrington, M. A, Jones, P. A., Imagawa, M., and Karin, M. (1988) Proc. Natl. Acad. Sci. USA **85**, 2066-2070.
23. Holler, M., Westin, G., Jiricny, J. and Schaffner, W. (1988) Genes Dev. **2**, 1127-1135.
24. Seeman, N. C., Rosenberg, J. M., and Rich, A. (1976) Proc. Natl. Acad. Sci. USA **73**, 804-808.
25. Schleif, R. (1988) Science **241**, 1182-1187.