Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 5670–5674, July 1991
Biochemistry

DNA looping and Sp1 multimer links: A mechanism for transcriptional synergism and enhancement

(transcription factor Sp1/enhancer mechanism/scanning transmission electron microscopy)

IRIS A. MASTRANGELO*, ALBERT J. COUREY†, JOSEPH S. WALL*, STEPHEN P. JACKSON‡§, AND PAUL V. C. HOUGH*¶

*Biology Department, Brookhaven National Laboratory, Upton, NY 11973; †Department of Chemistry and Biochemistry, University of California at Los Angeles, Los Angeles, CA 90024; ‡Howard Hughes Medical Institute, Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720

Communicated by Robert G. Roeder, March 15, 1991

ABSTRACT **Using conventional and scanning transmission electron microscopy, we have examined the physical basis of long-range enhancer effects between distal and proximal elements in a eukaryotic promoter. Specifically, we have studied binding of human transcription factor Sp1 to 10-base-pair G+C-rich elements ("GC boxes") located at −100 and +1700 relative to the RNA start site. It was previously observed that the distantly located site functions in synergism with the promoter-proximal site to strongly activate transcription *in vivo*. Here we demonstrate that this synergism is likely to be a direct consequence of interactions between remote and local Sp1, the remote Sp1 translocated to the promoter by a DNA loop. Scanning transmission electron microscopy shows that Sp1 initially forms a tetramer and subsequently assembles multiple tetramers stacked in register at the DNA loop juncture. This unexpected finding not only provides the physical basis for loop formation but also defines a biological process leading to strongly increased concentration of activator protein at the promoter. The mechanism may unify the problem of transcriptional activation by removing enhancer action as a separate class of regulatory activity.**

Control of eukaryotic gene expression often occurs at initiation of transcription. This process involves interactions between promoter- and enhancer-binding proteins and the RNA polymerase II transcriptional machinery (1–5). In many instances, the spatial and temporal specificity of transcription is controlled by enhancer elements that are located several kilobases from the RNA start site. Hence, to understand how genes are turned on and off, it is necessary to understand the mechanism(s) of long-range transcription regulation. DNA looping mediated by protein–protein interaction has been widely considered a likely mechanism for establishing a presence at the proximal promoter of distally bound protein (6).

In this work, we employed electron microscopy (EM) to examine the mechanistic basis of long-range synergism that is mediated by human transcription factor Sp1. Sp1 is a promoter-selective transcription factor, for both cellular and viral genes, that binds to G+C-rich elements with three zinc fingers and activates transcription via glutamine-rich domains (7–13). Although Sp1 is generally considered to be a protein that interacts with promoter-proximal elements, Sp1 bound 1700 nucleotides downstream of the RNA start site of the herpes simplex virus thymidine kinase gene can act synergistically with the factor bound at the normal upstream position at −100 (13). Thus, Sp1 also possesses the ability to mediate long-range activation of transcription in a manner similar to that of enhancer-binding transcription factors. These experiments and others (14–16) suggest that there is no clear distinction between promoter- and enhancer-binding factors and that long-range vs. short-range effects will be found to result from differing means for quaternary association.

In the transcription experiments above, showing synergism *in vivo*, direct contact between distal and proximal Sp1 was considered a likely mechanism for functional enhancement and DNA looping a probable means for protein translocation. However, neither here nor in other eukaryotic systems showing functional enhancement has direct evidence been found for DNA loops or mediating link structures (2, 6, 17). On the other hand, in prokaryotic systems, apposition of separated DNA sites through looping has been implicated in processes as diverse as regulation of transcription, initiation of DNA replication, and recombinational inversion, transposition, and integration (refs. 18–26; ref. 27 and references therein). For repressor proteins of phage λ and the *lac* and *deo* operons of *Escherichia coli*, and for the bacterial enhancer protein NtrC, DNA loops generated by protein–protein contacts have been observed by EM (28–31). In a study of binding to the upstream region of the mammalian uteroglobin gene, DNA looping between progesterone receptors bound approximately 2450 and 2600 base pairs (bp) from the transcription start site has been demonstrated by EM (32). As noted by those authors, there is so far no direct evidence relating receptor–receptor interaction to function. The present report establishes activator protein Sp1 as a physical link between functionally interacting proximal and distal binding sites in a eukaryotic promoter and describes a highly ordered, multimeric link structure that appears able to carry out many of the functions normally ascribed to enhancers.

MATERIALS AND METHODS

Binding Reactions. Changes in binding reactions made for microscopy—e.g., lowering the Nonidet P-40 concentration in the binding buffer from 0.1% to 0, use of glutaraldehyde crosslinking and higher protein and DNA concentrations—were tested in DNase I footprinting reactions to show that saturation binding was retained. With high binding shown by footprinting from the droplet applied to the EM grid, the fraction of DNA substrate molecules having one or more specifically bound Sp1 molecules was increased ≈6-fold. Double binding was increased from undetectable to up to 40%.

In typical binding reactions, 1–2 μl of vaccinia virus-expressed HeLa cell Sp1, 75 μg/ml, was added to Z^U buffer,

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: EM, electron microscopy; STEM, scanning transmission EM; SV40, simian virus 40.

§Present address: The Wellcome Trust, Cambridge University, Tennis Court Road, Cambridge, England CB21QR.
¶To whom reprint requests should be addressed.

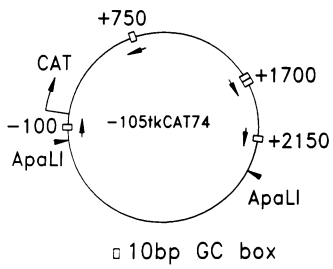

FIG. 1. Schematic diagram of the 4355-bp reporter plasmid −105tkCAT74. Boxes represent G+C-rich sequences recognized by Sp1. The "GC box" labeled −100 is in the herpes simplex thymidine kinase promoter that drives the chloramphenicol acetyltransferase (CAT) gene. At +1700, GC boxes III and IV from the simian virus 40 (SV40) 21-bp repeats have been inserted (13). The GC box in the vector at +2150 has the same sequence as an Sp1 recognition element in the dihydrofolate reductase gene (9). The site in the CAT gene coding region at +750 is opposite in orientation to the other three. Length of the upper linear fragment released by *Apa*LI digestion is 2612 bp. Distance from the −100 GC box to the left *Apa*LI cut is 96 bp, and that from the +1700 boxes to the right cut is 684 bp.

bringing the volume to 35 μl (33). Z^U buffer is standard DNase I footprinting buffer without Nonidet P-40 [25 mM Hepes·KOH, pH 7.5/100 mM KCl/10 μM $ZnSO_4$/1 mM dithiothreitol/20% (vol/vol) glycerol (10)]. Thirty-five microliters of probe DNA was added to give 1–3 ng/μl in the reaction mixture. Following incubation on ice for 15–20 min, 70 μl of 10 mM $MgCl_2$ was added. Droplets (3 μl) were placed on EM grids.

In parallel experiments using endogenous HeLa cell Sp1, observed complexes were essentially the same as those described here.

EM and Scanning Transmission EM (STEM). Thin carbon substrate on grids was air-discharged and coated with poly(L-lysine) (34). Complexes adhering to the film were crosslinked for 2 min with 0.1% glutaraldehyde and washed with 0.1× Z^U buffer lacking glycerol, and with 20 mM ammonium acetate. For conventional microscope study, grids were stained and shadowed with uranyl acetate and tungsten.

For STEM, structures were observed on a stage cooled to −150°C (35, 36). For the magnifications used in these experiments, the electron beam intersects a molecule at positions separated by 5 Å × 5 Å or 10 Å × 10 Å in a 512 × 512 matrix. The molecular mass of an unstained, unshadowed molecule as measured by STEM is proportional to the sum of the number of scattered electrons over all incident beam positions that intersect the molecule (35–37).

RESULTS

Quaternary Association of Remote and Local Sp1. To examine Sp1-dependent DNA looping by EM, we used the plasmid DNA containing Sp1 binding sites at −100 and +1700 relative to the RNA start site that was employed in the *in vivo* experiments, except for the presence of two, not six, GC boxes at +1700 (Fig. 1). Conventional EM analysis revealed that Sp1 bound specifically to its recognition sites (GC boxes) on a topologically relaxed plasmid template with high frequency (64% of all plasmids contained bound Sp1). A fraction of the plasmids (27%) contained Sp1 bound specifically at both proximal and distal GC boxes (Fig. 2*A*), and in 79% of those plasmids the distant GC boxes were joined in a DNA loop to proximal GC boxes by Sp1 (Fig. 2 *B–D*). Significantly, a great majority of the plasmids exhibiting DNA looping contained loops with contour lengths consistent with Sp1 bound at −100 and +1700 relative to the start site (Fig. 2*B* and *C*; Table 1).

We next investigated the binding of Sp1 to supercoiled or linearized plasmid DNA. With supercoiled DNA, link ("nexus") structures were observed with loop contour lengths identical to that expected for Sp1 bound at −100 and +1700 (Fig. 2*E*). Sp1 binding to linear DNA was similar to that observed with relaxed and supercoiled DNA (Fig. 2*F*; Table 1). In addition, because of the known location of restriction enzyme digestion sites of the plasmid, it was possible to determine unambiguously the positions of Sp1 DNA binding on the linear DNA. The binding statistics for the relaxed circular and linear templates are comparable (Table 1), and thus it appears that whether the DNA is in a circular or linear

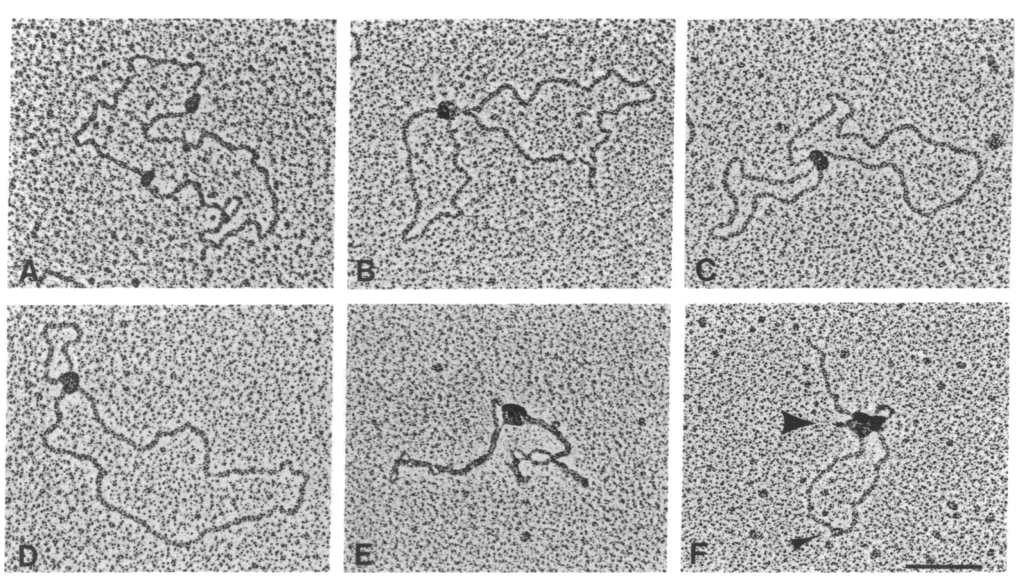

FIG. 2. Sp1 assembles both unlinked and nexus structures on DNA templates of differing topology, as observed by conventional EM. Binding statistics are reported in Table 1. (Bar = 100 nm.) (*A*) Contour length of DNA separating the two Sp1 assemblies is 1700 bp, or 40% of the plasmid length. (*B–D*) Nexus structures divide relaxed plasmids into two loops of unequal length. Contour length of the smaller loop, 1700 bp in *B* and *C* and 400 bp in *D*, indicates the Sp1 nexus at −100/+1700 and +1700/+2150 GC boxes, respectively. (*E*) On a supercoiled template, the large Sp1 assembly divides the plasmid into 1750-bp and 2500-bp lengths, left and right, respectively. (*F*) On the *Apa*LI fragment, Sp1 oligomers extending 40–60 bp along the DNA link the −100 GC box and +1700 GC boxes measuring, respectively, 90 bp from the end of the short arm (large arrowhead) and 690 bp from the end of the long arm. A low multimer of Sp1 is bound at the +750 GC box (small arrowhead).

Table 1. Binding of Sp1 to the relaxed circular plasmid −105tkCAT74 and its 2600-bp *Apa*LI linear fragment

	Relaxed circular	Linear
Total number	296	96
Sp1 at one or more GC boxes	189 (64%)	77 (80%)*
Sp1 at two separated GC boxes	79 (27%)	35 (36%)
Nexus, all† (% is of previous row)	62 (79%)	27 (77%)
Nexus, −100/+1700 (% is of previous row)	45 (73%)	12 (44%)

A molar ratio of seven Sp1 monomers per GC box was used. Statistics of binding and nexus formation for supercoiled −105tkCAT74 are qualitatively similar. Contour lengths were measurable on ≈50% of these plasmids.

*Nonspecific binding occurred in an additional 5–10% of fragments.

†Comparison of normalized loop contour lengths with sequence separations [square brackets] leads to identification of the GC-box locations linked: 1754 ± 45 bp, n = 27 [1816] (−100/+1700); 2128 ± 80 bp, n = 6 [2272] (−100/+2150); 466 ± 56 bp, n = 6 [456] (+1700/+2150); 896 ± 74 bp, n = 4 [966] (+750/+1700); 1247 ± 108 bp, n = 4 [1412] (+750/+2150). Fourfold higher precision position measurements are available by STEM, though usually for smaller numbers of events.

state is not an important factor in the formation of the nexus structures. Furthermore, the contour lengths of DNA molecules with or without bound Sp1 are indistinguishable (data not shown), which suggests that there is no significant compression due to the wrapping of the DNA around the Sp1 molecule.

The Link Is Effected by a Single Tetramer or Several in Register. In the conventional micrographs (e.g., Fig. 2*F*), it is clear by comparison with the unbound Sp1 multimers nearby on the foil that the link structure is composed of many monomer units of Sp1. STEM allows us to make this observation quantitative by electron-scattering determination of molecular mass (*Materials and Methods* and references given there). The methodology for determination of multimeric state has recently been strengthened by use of the "mass ladder," defined as the measurement of the masses of all the low multimers of protein, not bound to DNA, which are found deposited from solution onto the foil (34). The mass ladder for Sp1 is shown in Fig. 3*A*. Well-defined peaks are obtained for monomer through tetramer, and the higher mass region shows some evidence for the octamer. The mean mass for each Sp1 peak, plotted vs. oligomer number, gives a straight line (Fig. 3*A Inset*). The pattern of one to four and eight monomer units suggests that the tetramer is significant, and this suggestion is strongly reinforced by measurement of structures occurring at nexus. There, Sp1 was present as at least the tetramer and more often as two, three, and higher multiples of the tetramer (Fig. 3*B*). The tetrameric form may be essential for assembly of the link; alternatively, the binding of Sp1 to apposed GC boxes may promote formation of tetramers and multiples of tetramers.

In STEM, additional information is available as quantitative micrographs (Fig. 4). In these experiments, the scattered electron count at a particular probe position determines the mass thickness (37). Unlike the total molecular mass measurements, the observed distribution of mass thickness depends upon molecular orientation. To clarify the presentation, we have expressed mass thickness as a linear thickness of "compact" protein (see legend to Fig. 4). The species shown in Fig. 4 were chosen because they provided a particularly good "top view" of Sp1 (see Fig. 5). Comparison of the mass thickness of the octamer not bound to DNA (Fig. 4*C*) with that of the 12-mer at nexus (Fig. 4*A* and *B*) reveals that the 12-mer has 1.5 times the mass thickness of the octamer. This finding, along with the quadrangular appear-

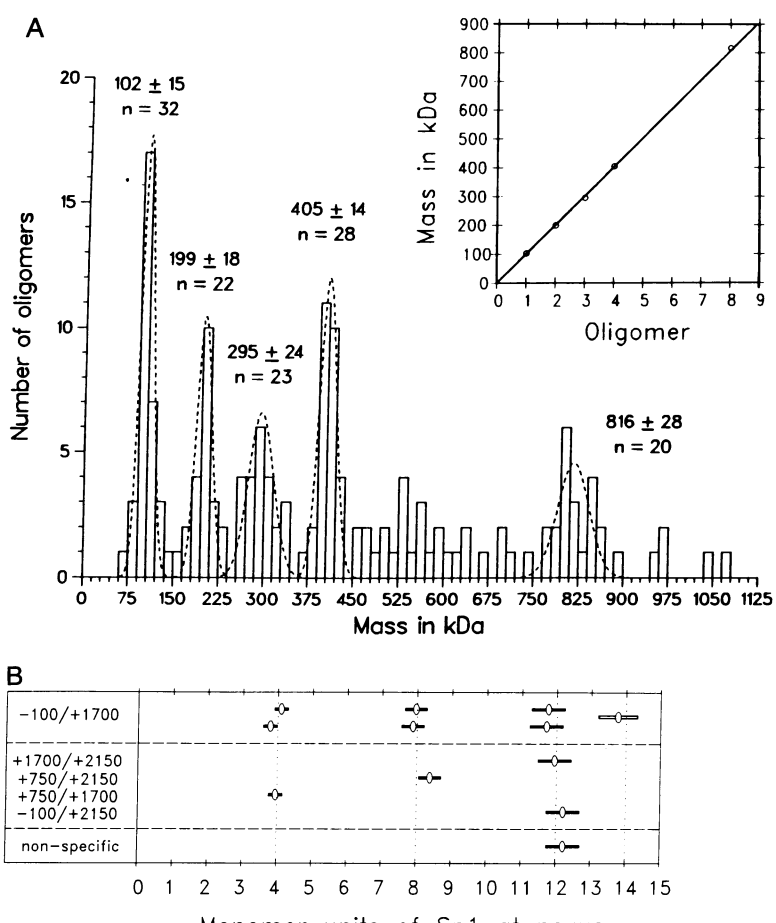

FIG. 3. Identification of oligomeric states of Sp1. (*A*) STEM mass measurements of 159 Sp1 molecules not bound to DNA. Gaussian distributions fitted to individual peaks have standard deviations decreasing from ≈15% at 100 kDa to ≈4% at 800 kDa, in accord with theoretical expectations (35, 36). In the *Inset*, the mean for each peak is plotted vs. oligomer number. Slope of the linear least-squares fit is 101 ± 6 kDa. This is the calibration procedure's unit mass for Sp1 as referenced to TMV (34). Masses, at nexus or elsewhere, are referred to the least-squares straight line for determination of number of monomer units. (*B*) Number of monomer units of Sp1 in nexus structures. The open ellipse indicates measured mass and the horizontal bars a standard error of 4%, taken from the Gaussian fits in *A*. All but the 14-mer have masses consistent with integer x tetramer. The 14-mer may not be an intact structure; additional mass was found nearby on the foil.

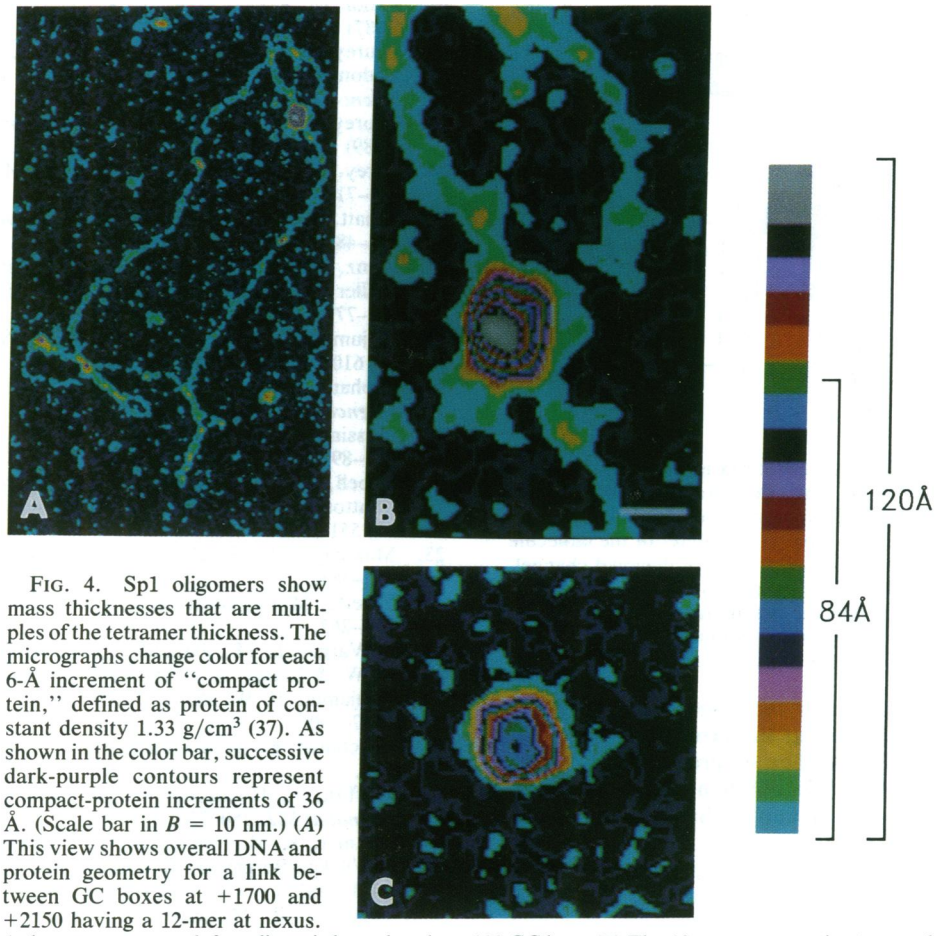

FIG. 4. Sp1 oligomers show mass thicknesses that are multiples of the tetramer thickness. The micrographs change color for each 6-Å increment of "compact protein," defined as protein of constant density 1.33 g/cm^3 (37). As shown in the color bar, successive dark-purple contours represent compact-protein increments of 36 Å. (Scale bar in *B* = 10 nm.) (*A*) This view shows overall DNA and protein geometry for a link between GC boxes at +1700 and +2150 having a 12-mer at nexus. At bottom, extreme left, a dimer is bound at the −100 GC box. (*B*) The 12-mer at nexus in *A* extends 120 Å above the thin carbon substrate, with color contours passing through three purple levels and ending in two white shades. (*C*) An 8-mer, not bound to DNA, extends 84 Å above the thin carbon substrate, with color contours passing through two purple levels and ending in blue and green.

ance of the oligomers, suggests that the Sp1 tetramers are stacked in register. Mass distributions of most if not all oligomers of Sp1 are consistent with the structures described for the 8- and 12-mers on the basis of the most symmetrical micrographs (I.A.M. and P.V.C.H., unpublished data). The observed preference for more than one tetramer suggests favorable tetramer–tetramer interactions, or Sp1–DNA interaction at sequences adjacent to GC boxes‖, or both.

DISCUSSION

The Sp1 tetramer as an important component of the link structure was an unexpected finding. Courey *et al.* (13) had obtained evidence for Sp1 dimers, but not higher multimers, based on glutaraldehyde crosslinking followed by SDS/polyacrylamide gel electrophoresis. The power and robustness of mass measurement within the framework of the mass ladder, as a method for determining multimeric state, have been demonstrated in a study of SV40 large tumor (T) antigen. The T-antigen mass ladder showed ATP-dependent assembly of hexamers in solution and hexamers and double hexamers surrounding origin DNA (34). Biochemical investigation over the preceding decade had concluded that T antigen assembled tetramers in solution and bound at the origin as one or several tetramers (references in ref. 34). Recently, confirmation of the hexamer in solution and the hexamer and double hexamer at the SV40 origin have been obtained by native gel electrophoresis (ref. 38; C. Prives, personal communication), in good agreement with STEM micrographs of the bound double hexamer.

Possible Structures for the Link Tetramer. Tetramers consisting of identical subunits can assemble only as two types of ring structure (39). The first ring type, C_4, has fourfold rotational symmetry (i.e., is identical after any number of 90° rotations about an axis perpendicular to the plane of the ring). The second type, D_2, has successive monomers, including zinc-finger domains, inverted, as illustrated in Fig. 5. D_2 is more likely, since 9 out of 11 homotetramers whose structures have been solved crystallographically are of this type (refs. 40–42; ref. 43 and references therein). EM cannot distinguish D_2 from C_4, since micrographs describe projections onto a plane. For either symmetry the four zinc-finger domains provide a strong capability for linking remote and promoter-proximal GC boxes. The D_2 structure is interesting in that it provides automatically the orientation independence observed in enhancer action—a remote GC box that is opposite in orientation can simply choose the next zinc-finger domain.

Function of the Nexus. The function of nexus structures may be to increase the local concentration of activator proteins (20) in the vicinity of the transcription start site where the activation domain can facilitate assembly of the initiation complex. Although a nexus can form with a single tetramer, we have found that there are typically two or more tetramers at the loop juncture. Multimerization through

‖Looping induced by Sp1 was first observed by us in a study of binding to the six GC boxes of the SV40 early promoter. Linked sites occurred 150–750 bp distant, indicating that the nexus structure can join specific and nonspecific sites.

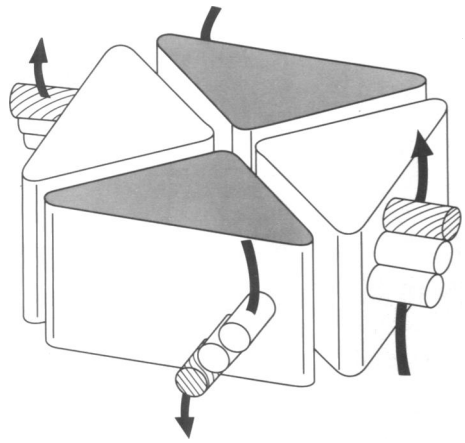

FIG. 5. A model of the Sp1 tetramer at nexus, drawn as a D_2 ring. Monomers are alternatively up (unshaded), rotated 180° to down (shaded), up, and down. The D_2 tetramer has twofold rotational symmetry around a vertical axis through the center of the molecule and also about horizontal axes centered on each diagonal channel. Zinc fingers, shown as cylinders, are inverted in adjacent monomers. Each monomer's zinc-finger domain can interact with a single GC box shown as an arrow. The GC-box orientation changes by 180° for binding to the next monomer of the ring.

stacking of tetramers at nexus increases the quantity of activator via an ordered array at the promoter. Since it is known that an increased number of proximal GC boxes increases activation strongly (11), our demonstration of high-multimer, remote–local association of Sp1 may unify the mechanisms for enhancer action and local activation. Furthermore, a 12-mer of Sp1 is capable of spanning at least 36 bp of DNA and could directly interact with the RNA polymerase initiation complex. Control of multimerization may provide a mechanism for regulation.

In the future, the experimental approach described in this study could be applied and extended to the analysis of other transcription factors. For instance, various combinations of different purified factors could be examined for nexus formation, and a correlation may be established between nexus formation and long-range transcriptional activation. These types of experiments could lead to greater insight concerning specific and productive interactions between different sequence-specific DNA-binding factors, including those involved in local activation (1, 44) and assembly of the RNA polymerase II initiation complex (45).

This work was supported by the Office of Health and Environmental Research of the U.S. Department of Energy and by grants from the National Institutes of Health (to P.V.C.H.). S.P.J. is a Lucille P. Markey Visiting Fellow. The Brookhaven STEM is a National Institutes of Health Biotechnology Resource.

1. Mitchell, P. & Tjian, R. (1989) *Science* **245,** 371–378.
2. Serfling, E., Jasin, M. & Schaffner, W. (1985) *Trends Genet.* **1,** 224–230.
3. Reinberg, D., Horikoshi, M. & Roeder, R. G. (1987) *J. Biol. Chem.* **262,** 3322–3330.
4. Gillies, S. D., Morrison, S. L., Oi, V. T. & Tonegawa, S. (1983) *Cell* **33,** 717–728.
5. Banerji, J., Olson, L. & Schaffner, W. (1983) *Cell* **33,** 729–740.
6. Ptashne, M. (1986) *Nature (London)* **322,** 697–701.
7. Dynan, W. S. & Tjian, R. (1983) *Cell* **32,** 669–680.
8. Gidoni, D., Dynan, W. S. & Tjian, R. (1984) *Nature (London)* **312,** 409–413.
9. Kadonaga, J. T., Jones, K. A. & Tjian, R. (1986) *Trends Biochem. Sci.* **11,** 20–23.
10. Kadonaga, J. T., Carner, K. R., Masiarz, F. R. & Tjian, R. (1987) *Cell* **51,** 1079–1090.
11. Courey, A. J. & Tjian, R. (1988) *Cell* **55,** 887–898.
12. Kadonaga, J. T., Courey, A. J., Ladika, J. & Tjian, R. (1988) *Science* **242,** 1566–1570.
13. Courey, A. J., Holtzman, D. A., Jackson, S. P. & Tjian, R. (1989) *Cell* **59,** 827–836.
14. Carey, M., Leatherwood, J. & Ptashne, M. (1990) *Science* **247,** 710–712.
15. Schatt, M. D., Rusconi, S. & Schaffner, W. (1990) *EMBO J.* **9,** 481–487.
16. Bienz, M. & Pelham, H. R. B. (1986) *Cell* **45,** 753–760.
17. Müller, H.-P., Sogo, J. M. & Schaffner, W. (1989) *Cell* **58,** 767–777.
18. Majumdar, A. & Adhya, S. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 6100–6104.
19. Popham, D. L., Szeto, D., Keener, J. & Kustu, S. (1989) *Science* **243,** 629–635.
20. Mossing, M. C. & Record, M. T., Jr. (1986) *Science* **233,** 889–892.
21. Lobell, R. B. & Schleif, R. F. (1990) *Science* **250,** 528–532.
22. Chattoraj, D. K., Mason, R. J. & Wickner, S. H. (1988) *Cell* **52,** 551–557.
23. Mukherjee, S., Erickson, H. & Bastia, D. (1988) *Cell* **52,** 375–383.
24. Surette, M. G., Buch, S. J. & Chaconas, G. (1987) *Cell* **49,** 253–262.
25. de Vargas, L. M., Pargellis, C. A., Hasan, N. M., Bushman, E. W. & Landy, A. (1988) *Cell* **54,** 923–929.
26. Benjamin, H. W. & Cozzarelli, N. R. (1988) *EMBO J.* **7,** 1897–1905.
27. Heichman, K. A. & Johnson, R. C. (1990) *Science* **249,** 511–517.
28. Griffith, J., Hochschild, A. & Ptashne, M. (1986) *Nature (London)* **322,** 750–752.
29. Kramer, K. M., Niemoller, M., Amouyal, M., Revel, B., v.Wilcken-Bergmann, B. & Müller-Hill, B. (1987) *EMBO J.* **8,** 1481–1491.
30. Amouyal, M., Mortensen, L., Buc, H. & Hammer, K. (1989) *Cell* **58,** 545–551.
31. Su, W., Porter, S., Kustu, S. & Echols, H. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 5504–5508.
32. Théveny, B., Bailly, A., Rauch, C., Rauch, M., Delain, E. & Milgrom, E. (1987) *Nature (London)* **329,** 79–81.
33. Jackson, S. P., MacDonald, J. J., Lees-Miller, S. & Tjian, R. (1990) *Cell* **63,** 155–165.
34. Mastrangelo, I. A., Hough, P. V. C., Wall, J. S., Dodson, M., Dean, F. B. & Hurwitz, J. (1989) *Nature (London)* **338,** 658–662.
35. Wall, J. (1979) in *Introduction to Analytical Electron Microscopy*, eds. Hren, J. J., Goldstein, J. I. & Joy, D. C. (Plenum, New York), pp. 333–342.
36. Mastrangelo, I. A., Hough, P. V. C., Wilson, V. G., Wall, J. S., Hainfeld, J. F. & Tegtmeyer, P. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 3626–3630.
37. Hough, P. V. C., Mastrangelo, I. A., Wall, J. S., Hainfeld, J. F., Sawadogo, M. & Roeder, R. G. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 4826–4830.
38. Parsons, R. E., Stenger, J. E., Ray, S., Welker, R., Anderson, M. E. & Tegtmeyer, P. (1991) *J. Virology*, in press.
39. Hamermesh, M. (1962) in *Group Theory and Its Application to Physical Problems* (Addison–Wesley, Reading, MA), pp. 32–67.
40. Miller, S. (1989) *Protein Eng.* **3,** 77–83.
41. Varghese, J. N., Laver, W. G. & Colman, P. M. (1983) *Nature (London)* **303,** 35–40.
42. Xia, Z.-X., Shamala, N., Bethge, P. H., Lim, L. W., Bellamy, H. D., Xuong, N. Y., Lederer, F. & Mathews, F. S. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 2629–2633.
43. Janin, J., Miller, S. & Chothia, C. (1988) *J. Mol. Biol.* **204,** 155–164.
44. Sawadogo, M. & Roeder, R. G. (1985) *Cell* **43,** 165–175.
45. Buratowski, S., Hahn, S., Guarente, L. & Sharp, P. A. (1989) *Cell* **56,** 549–561.