On the other hand, osteosarcoma cells expressing apparently normal RB protein, such as U-2OS and TE85 (*22, 23*), were not significantly affected by infection with Rb virus. This implies that the RB gene is functional in these cells and that alternative pathways in osteosarcoma genesis do not involve RB gene inactivation. Our preliminary studies also support this concept in human breast cancers (*26*). Other proposed cancer suppressor genes, including those for Wilms' tumor on chromosome 11 (*27*), renal cell carcinoma and small cell lung carcinoma on chromosome 3 (*28, 29*), and colon cancer on chromosome 5 (*30*), may function analogously to the RB gene in suppressing different kinds of cancer.

It is not known whether inactivation of one or more cancer suppressor genes in a cell is sufficient to cause cancer. Regardless of this uncertainty, replacement of suppressor genes in tumor cells, as demonstrated here, could be a novel strategy for the treatment of clinical malignancy. Unlike conventional, cytotoxic cancer therapies, gene therapy would be based on permanent correction of an underlying defect in tumor cells (*31*). Therapy may not need to be targeted because cancer suppressor genes should not harm normal cells. The ultimate utility of this approach will depend on progress in obtaining other cancer suppressor genes, in understanding their involvement in human tumors, and in improving the technology of exogenous gene expression.

Note added in proof: We have established 12 retinoblastoma cell clones that stably express RB protein for more than 4 months. Thus, expression of the foreign protein is not lethal to the cells.

REFERENCES AND NOTES

1. H. Harris, O. J. Miller, G. Klein, P. Worst, T. Tachibana, *Nature* **223**, 363 (1969); G. Klein, U. Bregula, F. Wiener, *J. Cell Sci.* **8**, 659 (1974).
2. E. J. Stanbridge, *Adv. Viral Oncol.* **6**, 83 (1987).
3. B. E. Weissman *et al.*, *Science* **236**, 175 (1987).
4. A. G. Knudson, *Cancer* **35**, 1022 (1975); A. Koufos *et al.*, *Nature* **316**, 330 (1985).
5. W. F. Bodmer *et al.*, *Nature* **328**, 614 (1987); C. Larsson, B. Skogseid, K. Oberg, Y. Nakamura, M. Nordenskjold, *ibid.* **332**, 85 (1988); G. A. Rouleau *et al.*, *ibid.* **329**, 246 (1987); B. R. Seizinger, R. L. Martuza, J. F. Gusella, *ibid.* **322**, 644 (1986).
6. J. M. Bishop, *Science* **235**, 305 (1987); G. Klein, *ibid.* **238**, 1539 (1987); H. Harris, *Nature* **323**, 582 (1986); R. Sager, *Cancer Res.* **46**, 1573 (1986).
7. W.-H. Lee, R. Bookstein, E. Y.-H. P. Lee, *J. Cell. Biochem.*, in press.
8. J. J. Yunis and N. Ramsey, *Am. J. Dis. Child.* **132**, 161 (1978); R. S. Sparkes *et al.*, *Science* **208**, 1042 (1980); P. Ward *et al.*, *J. Med. Genet.* **21**, 92 (1984).
9. W. K. Cavenee *et al.*, *Nature* **305**, 779 (1983).
10. S. H. Friend *et al.*, *ibid.* **323**, 643 (1986).
11. W.-H. Lee *et al.*, *Science* **235**, 1394 (1987).
12. Y.-K. T. Fung *et al.*, *ibid.* **236**, 1657 (1987).
13. R. Bookstein *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2210 (1988); E. Y.-H. P. Lee *et al.*, *ibid.*, p. 6017.
14. W.-H. Lee *et al.*, *Nature* **329**, 642 (1987).
15. P. Whyte *et al.*, *ibid.* **334**, 124 (1988); J. A. De Caprio *et al.*, *Cell* **54**, 275 (1988).
16. J. Toguchida *et al.*, *Cancer Res.* **48**, 3939 (1988); A. E. Mendoza, J.-Y. Shew, E. Y.-H. P. Lee, R. Bookstein, W.-H. Lee, *Hum. Pathol.* **19**, 487 (1988); S. H. Friend *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 9059 (1987); E. Y.-H. P. Lee *et al.*, *Science* **241**, 218 (1988); J. W. Harbour *et al.*, *ibid.* **241**, 353 (1988).
17. R. Weiss, N. Teich, H. Varmus, J. Coffin, *RNA Tumor Viruses* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1985).
18. A. D. Miller *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 4709 (1983); P. J. Southern and P. Berg, *J. Mol. Appl. Genet.* **1**, 327 (1982).
19. J. R. de Wet *et al.*, *Mol. Cell. Biol.* **7**, 725 (1987).
20. A. D. Miller, M.-F. Law, I. M. Verma, *ibid.* **5**, 431 (1985); L. Xu, J. K. Yee, J. D. Wolf, T. Friedmann, in preparation.
21. R. Mann *et al.*, *Cell* **33**, 153 (1983); C. L. Cepko, *ibid.* **37**, 1053 (1984).
22. H.-J. S. Huang *et al.*, unpublished data.
23. J.-Y. Shew *et al.*, manuscript in preparation.
24. P. K. Vogt, *Virology* **46**, 939 (1971); D. A. Zarling and H. M. Temin, *J. Virol.* **17**, 74 (1976); M. Emerman and H. M. Temin, *ibid.* **50**, 42 (1984); D. J. Jolly, R. C. Willis, T. Friedmann, *Mol. Cell Biol.* **6**, 1141 (1986).
25. W.-H. Lee *et al.*, *J. Virol.* **38**, 1064 (1981).
26. E. Y.-H. P. Lee *et al.*, in preparation.
27. S. H. Orkin, D. S. Goldman, S. E. Sallan, *Nature* **309**, 172 (1984); A. Koufos *et al.*, *ibid.*, p. 170; E. R. Fearon *et al.*, *ibid.*, p. 176.
28. B. Zbar *et al.*, *ibid.* **327**, 721 (1987).
29. S. L. Naylor, B. E. Johnson, J. D. Minna, A. Y. Sakaguchi, *ibid.* **329**, 451 (1987).
30. E. R. Fearon, S. H. Hamilton, B. Vogelstein, *Science* **238**, 193 (1987); E. Solomon *et al.*, *Nature* **328**, 616 (1987).
31. M. A. Eglitis and W. F. Anderson, *BioTechniques* **6**, 608 (1988).

32. We thank D. Miller, I. Verma, B. Weinstein, and M. Krigler for supplying retroviral vectors; T. Sery for WERI-Rb27; H. To and F. Hong for their technical assistance; and P. Scully for pathological examination. Supported by grants from the National Institutes of Health (EY-05758 and HD-20034) and the American Cancer Society to W.-H.L. and T.F.

4 October 1988; accepted 4 November 1988

Distinct Regions of Sp1 Modulate DNA Binding and Transcriptional Activation

JAMES T. KADONAGA,* ALBERT J. COUREY,† JOSEPH LADIKA, ROBERT TJIAN

Sp1 is a sequence-specific DNA binding protein that activates RNA polymerase II transcription from promoters that contain properly positioned GC boxes. A series of deletion mutants of Sp1 were expressed in *Escherichia coli* and used to identify separate regions of the protein that are important for three different biochemical activities. The sequence-specificity of DNA binding was conferred by Zn(II) fingers, whereas a different region of Sp1 appeared to regulate the affinity of DNA binding. The *E. coli*–synthesized Sp1 was able to stimulate initiation of RNA synthesis in vitro, and at least two distinct segments of the protein contributed to its transcriptional activity.

TRANSCRIPTIONAL REGULATION OF genes in eukaryotes is mediated in part by sequence-specific factors that bind to promoter and enhancer elements (*1*). Many of these proteins have been characterized and, in some instances, the factors have been shown to participate directly in regulation of mRNA synthesis by in vitro transcription analysis. Promoters and enhancers typically contain multiple binding sites for several sequence-specific transcription factors, and it is likely that these factors act in conjunction with each other to specify a unique program of transcription for each of the thousands of genes in a eukaryote. Consequently, it is important to understand both how these factors modulate RNA polymerase II transcription as well as how these DNA binding proteins interact with each other.

Sp1 is a sequence-specific transcription factor that recognizes GGGGCGGGGC and closely related sequences, which are often referred to as GC boxes (*2*). Sp1 was initially identified as a factor from HeLa cells that selectively activates in vitro transcription from the SV40 early promoter (*3*) and binds to the multiple GC boxes in the 21-bp repeated elements in SV40 (*4*). The protein was then purified by sequence-specific DNA affinity chromatography (*5, 6*). Sp1 consists of two species of 95 and 105 kD (as determined by SDS–polyacrylamide gel electrophoresis), which appear to be variants of a single polypeptide (*7, 8*). More recently, by isolation of a partial cDNA encoding Sp1 and localization of the DNA binding domain, it was shown that Sp1 binds to DNA by interaction of contiguous Zn(II) finger motifs (*7*).

We had previously described isolation of a partial Sp1 cDNA, designated Sp1-1, that encodes the COOH-terminal 696 amino acid residues of Sp1 (*7*). To obtain the remainder of the Sp1 coding sequence, we prepared Sp1-enriched cDNA libraries by

Howard Hughes Medical Institute, Department of Biochemistry, University of California, Berkeley, CA 94720.

*Present address: Department of Biology, B-022, University of California, San Diego, La Jolla, CA 92093.
†To whom correspondence should be addressed.

using a primer that is complementary to the Sp1 cDNA. The Sp1 extension libraries were screened with a synthetic oligonucleotide that is complementary to a region 5′ of the primer sequence to yield only isolates that result from extension of the Sp1-specific primer on the Sp1 mRNA. By this approach, we generated 27 Sp1 5′ extension clones, and the longest of these isolates was designated Sp1-E7 (*9*). The Sp1 cDNA sequence is in GenBank (accession number J03133). We have attempted several times to determine the NH_2-terminal amino acid sequence of Sp1, but the protein appears to be derivatized at the NH_2-terminus. As a consequence, we will assign the first in-frame Met codon, which conforms moderately well with the consensus for eukaryotic initiation sites proposed by Kozak (*10*), as Met1. Beginning at Met1, Sp1 is a polypeptide of 778 amino acid residues with a calculated molecular mass of 79,902 daltons.

We carried out DNA blot analysis of the Sp1 gene by probing human genomic DNA with two fragments of the cDNA that encode either the Zn(II) fingers (designated "Zn fingers") or the region that is immediately NH_2-terminal to the Zn(II) fingers (designated "no fingers") (Fig. 1) (*11*). Under high stringency conditions, the Zn fingers probe hybridized to a unique Hind III fragment and a unique Pst I fragment (Fig. 1, lanes 1 and 2). Thus, Sp1 appears to be a single polypeptide that is encoded by a single gene. When the experiment was carried out under low stringency conditions, several fragments that weakly hybridized to either the Zn fingers or no fingers probes were identified (Fig. 1, compare lanes 3 and 4 with 1 and 2 and lanes 7 and 8 with 5 and 6). Since Zn(II) fingers have been identified in a variety of proteins (*12*), we were not surprised to find genomic DNA fragments that contain sequences that are similar to the Zn fingers probe. On the other hand, the region of Sp1 that is encoded by the no fingers probe (from 327 to 168 residues from the COOH-terminus) (*11*) did not possess any significant homology with other known proteins (*7*) and is important for transcriptional activity in vivo (*13*) and in vitro. Consequently, these data suggest that the Sp1 gene is a member of a new family of related genes that may encode transcription factors.

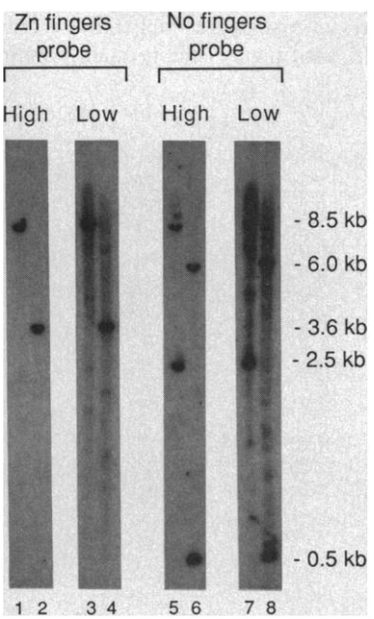

Fig. 1. DNA blot analysis of the Sp1 gene. Human genomic DNA (10 μg) was digested with either Hind III (lanes 1, 3, 5, and 7) or Pst I (lanes 2, 4, 6, and 8), subjected to electrophoresis on a 0.8% agarose gel, and transferred to GeneScreen hybridization membrane (Du Pont, Biotechnology Systems). The immobilized DNA was then hybridized either to the Zn fingers probe (lanes 1 to 4) or to the no fingers probe (lanes 5 to 8). The filters were treated either at high stringency conditions (lanes 1, 2, 5, and 6) or at low stringency conditions (lanes 3, 4, 7, and 8) (*11*). The apparent lengths of the DNA fragments are indicated.

To analyze the biochemical activities of Sp1, we initially expressed Sp1-778C protein in *Escherichia coli*. This protein consists of the NH_2-terminal 34 amino acid residues of the *lacZ′* gene encoded by pBluescript SK(+) fused to the 778 amino acid residues of Sp1 (*14*). Sp1-778C was purified to greater than 50% homogeneity (*15*) and subjected to SDS-gel electrophoresis and silver staining. The molecular size of Sp1-778C (97 kD; Fig. 3B) is similar to that of HeLa Sp1 (95 and 105 kD) (*5, 6*) and is significantly larger than the calculated molecular mass of 83,428 daltons (3,625 daltons of *lacZ′*+ 79,902 daltons of Sp1). Because of the similar SDS gel mobility of Sp1-778C and HeLa Sp1, it is likely that Sp1-778C contains most, if not all, of the Sp1 polypeptide.

A distinguishing feature of Sp1 is its ability to act as a promoter-selective transcription initiation factor in vitro. Therefore, we tested whether or not bacterially synthesized Sp1-778C protein possessed transcriptional activity. Sp1-778C can activate in vitro RNA synthesis from the SV40 early promoter, but not to the same extent as HeLa Sp1 (Fig. 2A). At first, we felt that the Sp1-778C protein might have been partially inactivated by the 6*M* urea treatment that occurs during its isolation (*15*). To address this possibility, we incubated HeLa Sp1 with 6*M* urea, removed the urea by gel filtration, and measured the transcriptional activity of the resulting protein (Fig. 2A, lanes 4 and 5). This treatment did not affect the ability of Sp1 to stimulate RNA synthesis. Thus, similar urea treatment of Sp1-778C is probably not responsible for its lower transcriptional activity relative to HeLa Sp1. Furthermore, human Sp1 expressed in *Drosophila* tissue culture cells activates RNA synthesis in vitro to the same extent as HeLa Sp1 (*13*). Thus, our longest Sp1 cDNA appears to contain all of the coding sequence that is necessary for full transcriptional activity. We therefore conclude that *E. coli*–synthesized Sp1 has the inherent ability to stimulate RNA synthesis in vitro, but that full transcriptional activity may require additional modifications of the protein that occur in higher organisms. HeLa Sp1, but not *E. coli*–synthesized Sp1, appears to possess multiple post-translational modifications, including multiple *O*-linked *N*-acetylglucosamine monosaccharide residues (*16*). The binding of wheat germ agglutinin, a lectin that recognizes *N*-acetylglucosamine residues, to HeLa Sp1 inhibits in vitro transcription to the level observed with *E. coli* Sp1, whereas DNA binding is unaffected by the lectin. It is possible that the lower specific activity of bacterial Sp1 is, in part, due to the lack of glycosylation.

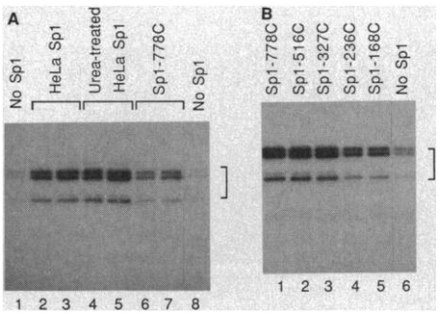

Fig. 2. In vitro transcription analysis of Sp1 expressed in *E. coli*. (**A**) HeLa Sp1 activates in vitro transcription from the SV40 early promoter to a greater extent than *E. coli*–synthesized Sp1. The SV40 early promoter was transcribed in vitro according to the procedure of Dynan and Tjian (*3*) with no added Sp1 (as a negative control; lanes 1 and 8), HeLa Sp1 (>90% purity, lanes 2 and 3) (*5, 6*), urea-treated HeLa Sp1 (lanes 4 and 5), or Sp1-778C (>50% purity, lanes 6 and 7) (*15*). The in vitro–synthesized RNA was detected by primer-extension analysis. (**B**) In vitro transcription analysis with NH_2-terminal deletion mutants of Sp1 expressed in *E. coli* (*14*). The SV40 early promoter was transcribed in vitro with Sp1-778C (lane 1), Sp1-516C (lane 2), Sp1-327C (lane 3), Sp1-236C (lane 4), and Sp1-168C (lane 5), or with no added Sp1 (as a negative control, lane 6). In these experiments, the DNA binding activity of each of the Sp1 proteins was determined by DNase I footprinting, and identical DNA binding activity units of each protein were added to the reactions. Saturating levels of Sp1 were used in these experiments.

Since Sp1-778C protein can activate RNA synthesis in vitro, we were able to localize regions of Sp1 that are important for transcriptional activation. First, we expressed a series of NH_2-terminal deletion

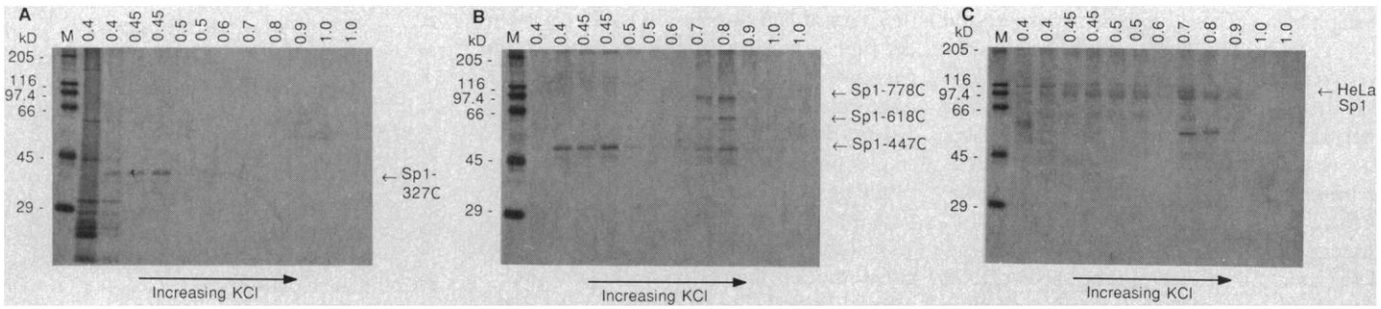

Fig. 3. Sp1 binds to DNA in low- and high-affinity forms (*18*). Fractions from sequence-specific DNA affinity chromatography of Sp1-327C, Sp1-778C, and HeLa Sp1 (*15*) were analyzed by SDS-gel electrophoresis and silver staining. The KCl concentrations in the buffer used to elute the protein are shown above each lane (as molarity). The molecular mass of each of the marker proteins is indicated in kilodaltons. (**A**) Elution of Sp1-327C. (**B**) Elution of Sp1-778C. The proteins designated Sp1-618C and Sp1-447C probably derive from internal translation initiation (*15*). (**C**) HeLa Sp1 contains both low and high salt-eluting species.

mutants of Sp1 in *E. coli* that contain 516, 327, 236, and 168 COOH-terminal amino acid residues of Sp1 (*14*). Then, the *E. coli*–synthesized Sp1 proteins were purified to greater than 50% homogeneity by sequence-specific DNA affinity chromatography (*5, 15*). We had previously shown that the COOH-terminal 168 amino acid residues of Sp1, which contain its three Zn fingers, are sufficient for sequence-specific binding to DNA (*7*). Consistent with the earlier work, Sp1-778C, Sp1-516C, Sp1-327C, Sp1-236C, and Sp1-168C all bind selectively to the GC box elements in the SV40 early promoter by deoxyribonuclease I (DNase I) footprint assays (*8*). Next, we carried out in vitro transcription reactions with the purified proteins (Fig. 2B). Sp1-778C stimulated RNA synthesis to the same extent as Sp1-516C and Sp1-327C and to a greater extent than Sp1-236C and Sp1-168C, which have low, but detectable, levels of transcriptional activity. Some portion of the segment of Sp1 between 327 and 236 amino acid residues from the COOH-terminus appears to be important for transcriptional activity in vitro, whereas a region that can weakly activate RNA synthesis is also present in the COOH-terminal 168 amino acid residues of Sp1. It thus appears that there are at least two regions of Sp1 that are involved in transcriptional activation in vitro. These findings are also consistent with studies in which we have localized segments of Sp1 that are important for transcriptional activity in vivo by transient transfection of wild-type and mutant variants of the Sp1 gene into *Drosophila* Schneider line 2 cultured cells, which provide an Sp1-deficient background (*13*).

In the course of DNA affinity purification of the *E. coli*–synthesized Sp1 proteins, we found that Sp1 binds to DNA in both low- and high-affinity forms. Sp1-778C (Fig. 3B) and Sp1-516C (*17*) eluted from the DNA affinity resin at a higher salt concentration than the smaller Sp1-327C (Fig. 3A), Sp1-236C (*17*), and Sp1-168C (*17*), whereas HeLa Sp1 eluted from the DNA affinity resin in both low- and high-affinity forms (Fig. 3C). In addition, the proteins designated Sp1-618C and Sp1-447C, which probably derive from internal translation initiation (*15*), eluted at high salt and at both low and high salt, respectively. This low- or high-affinity binding of Sp1 is distinct from its sequence-specific DNA binding activity because both the low- and high-affinity *E. coli*–synthesized proteins were able to bind to DNA with the same sequence specificity as HeLa Sp1 (*7, 18*). Since Sp1-778C and Sp1-516C bind to DNA with high affinity and Sp1-327C, Sp1-236C, and Sp1-168C bind to DNA with low affinity, a region of Sp1 between 516 and 327 amino acid residues from the COOH-terminus is likely to influence the high-affinity binding to DNA. This segment of Sp1 is distinct from the three Zn(II) fingers, and it appears that the Zn(II) fingers are responsible for sequence-specific DNA binding, whereas a separate domain of the protein may regulate the affinity of Sp1 binding to DNA.

What is the mechanism by which the DNA binding affinity of Sp1 is modulated? First, because of an entropic effect, multimers of Sp1 might cooperatively bind to DNA with higher affinity than monomers,

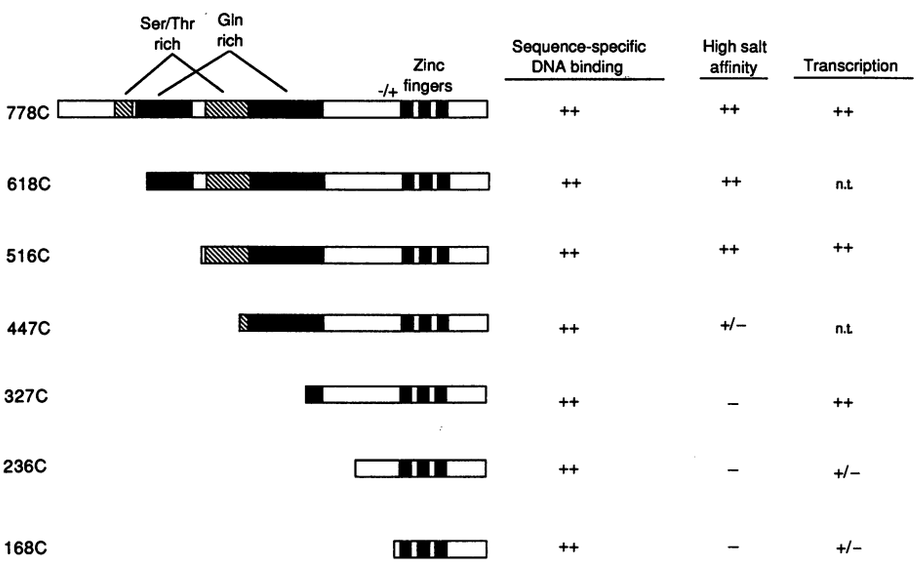

Fig. 4. Localization of distinct activities of Sp1. The structures of the COOH-terminal 778 amino acid residues of Sp1(778C) and of a number of NH_2-terminal deletion mutants of Sp1 (618C, 516C, 447C, 327C, 236C, and 168C) are shown on the left and the activities of these proteins in several different assays are tabulated on the right. The three Zn(II) finger motifs and segments of Sp1 that are rich in Ser/Thr or Gln (*7*) are depicted. The sequence-specific DNA binding activity was assayed by DNase I footprinting. The high salt affinity and in vitro transcriptional activity of the mutants were assayed as described in the text. In each case, ++ indicates activity indistinguishable from that of the 778C protein. In the case of the high salt affinity assay, +/– indicates that approximately half the protein elutes from the affinity column at low salt and the remainder elutes only at high salt, whereas – indicates that all the protein elutes at low salt. For the transcription assay, +/– indicates approximately 25% of the activity of the 778C protein, and n.t. means not tested.

in a manner similar to the λ repressor (*19*). In this case, the Sp1 segment between 516 and 327 amino acid residues from the COOH-terminus might be important for multimerization of the protein. Alternatively, by an enthalpic effect, direct protein-protein interaction (either intermolecular or intramolecular) between a region of Sp1 (516C to 327C) and the Zn(II) fingers could affect the strength of the binding of the fingers to DNA. Third, the amino acid residues that are NH_2-terminal to the fingers in Sp1-327C, Sp1-236C, and Sp1-168C, but not Sp1-778C and Sp1-516C, could artificially inhibit binding of Sp1 to DNA because of an unnatural conformation that is adopted in the absence of the remainder of the protein. We feel that this explanation is unlikely, however, because Sp1-327C, which binds with low affinity, activates in vitro transcription as well as Sp1-778C and Sp1-516C, which bind with high affinity. Hence, the region of Sp1-327C that is NH_2-terminal to the Zn fingers is probably in the native conformation.

It might be predicted, on the basis of the finding that the larger bacterially synthesized Sp1 proteins bind to DNA with high affinity and the smaller Sp1 proteins bind with low affinity, that full-length HeLa Sp1 would elute from the DNA resin at high salt. In contrast to this expectation, HeLa Sp1 contains both low and high salt-binding forms. In addition, unlike the remainder of the *E. coli*–synthesized Sp1 proteins, Sp1-447C elutes from the resins at both low and high salt concentrations. Although many different hypotheses could be presented to rationalize these observations, we suggest that Sp1-447C, which is intermediate in size between Sp1-516C (high affinity) and Sp1-327C (low affinity), is partially defective in the ability to increase the strength of DNA binding. Then, in the case of HeLa Sp1, there might be a mechanism, perhaps involving modification of the region between 516 and 327 amino acid residues from the COOH-terminus, by which the protein could be converted to either low- or high-affinity forms. Further studies are necessary, however, to clarify the importance of these phenomena with regard to the function of Sp1 in the cell.

As summarized in Fig. 4, we have used deletion mutagenesis to resolve three distinct biochemical activities of Sp1. Progressive deletion of Sp1 from the NH_2-terminus first affects the high-affinity binding activity and then the transcriptional activity of Sp1. In contrast, the sequence-specific binding activity of Sp1, which requires only the Zn finger motifs (*7, 13*), is unaffected in any of the mutants. The deletion analysis presented here suggests that the NH_2-terminal 451 amino acids of Sp1 are completely dispensable for Sp1 transcriptional activity. Alternatively, it is possible that Sp1 contains multiple redundant activation domains and that the activity of the NH_2-terminal portion of the protein is masked in our in vitro assay by the presence of the more COOH-terminal activation domains. In fact, recent experiments in which Sp1 mutants have been assayed in vivo indicate the presence of additional activation domains in the NH_2-terminal portion of the protein (*13*).

Studies on sequence-specific transcription factors from yeast suggest that negatively charged amino acid residues are an important feature of RNA polymerase II transcriptional activation (*20*). However, the transcriptional activation region of Sp1 that maps between 327 and 236 residues from the COOH-terminus is devoid of acidic amino acid residues. It is also unlikely that the *E. coli*–synthesized Sp1 proteins used in the in vitro transcription assays were phosphorylated. Hence, interpretation of the data leads to the hypothesis that Sp1 contains a region that stimulates RNA polymerase II transcription by a mechanism that does not involve negatively charged amino acid residues and may represent a novel type of activation domain.

REFERENCES AND NOTES

1. R. M. Evans, *Science* **240**, 889 (1988); N. C. Jones, P. W. J. Rigby, E. B. Ziff, *Genes Dev.* **2**, 267 (1988); T. Maniatis, S. Goodbourn, J. A. Fischer, *Science* **236**, 1237 (1987); S. McKnight and R. Tjian, *Cell* **46**, 785 (1986); M. Ptashne, *Nature* **322**, 697 (1986).
2. J. T. Kadonaga *et al.*, *Trends Biochem.* **11**, 20 (1986); J. T. Kadonaga, A. J. Courey, J. Ladika, T. Tjian, in *The Control of Human Retrovirus Gene Expression*, B. R. Franza, Jr., B. R. Cullen, F. Wong-Staal, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1988), pp. 239–250.
3. W. S. Dynan and R. Tjian, *Cell* **32**, 669 (1983).
4. ———, *ibid.* **35**, 79 (1983).
5. J. T. Kadonaga and R. Tjian, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5889 (1986).
6. M. R. Briggs, J. T. Kadonaga, S. P. Bell, R. Tjian, *Science* **234**, 47 (1986).
7. J. T. Kadonaga, K. C. Carner, F. R. Masiarz, R. Tjian, *Cell* **51**, 1079 (1987).
8. J. T. Kadonaga *et al.*, unpublished data.
9. Sp1-E7 has 504 nucleotides upstream of the original Sp1-1 isolate, whereas all of the other cDNA clones contain less than 250 nucleotides upstream of Sp1-1. We also isolated and sequenced portions of Sp1 genomic DNA (*8*). These experiments revealed identity between Sp1-E7 and Sp1 genomic DNA for 261 nucleotides immediately upstream of the Sp1-1 cDNA, and thus, this segment of the Sp1-E7 clone appears to be authentic. In contrast, the 244 nucleotide segment of Sp1-E7 that is upstream of the 261 nucleotide cDNA extension does not hybridize to the major Sp1 mRNA species detected by ribonuclease (RNase) protection assays (B. Lüscher and R. Tjian, unpublished data).
10. M. Kozak, *Nucleic Acids Res.* **12**, 857 (1984).
11. The Zn fingers probe comprises the Sp1 cDNA sequence from the Bam HI restriction site at nucleotide 1847 to nucleotide 2117 and contains the coding sequence of the three Zn(II) finger motifs. The no fingers probe comprises the cDNA sequence from the Apa I restriction site at nucleotide 1362 to the Bam HI restriction site at nucleotide 1847 and contains the coding sequence of the region between 327 and 168 amino acid residues from the COOH-terminus of Sp1. The high-stringency conditions are as follows. Hybridization: 0.2*M* sodium phosphate, *p*H 6.5, 10 m*M* EDTA, 15% (v/v) formamide, 7% (w/v) SDS, and 1% (w/v) bovine serum albumin; 65°C. The solution used for washing the filter: 0.1× SST, 0.1% (w/v) SDS; 65°C. The low-stringency conditions are as follows. Hybridization: 0.5*M* sodium phosphate, *p*H 6.5, 10 m*M* EDTA, 7% (w/v) SDS; and 1% (w/v) bovine serum albumin; 65°C. The solution used for washing the filter: 5× SST and 0.1% (w/v) SDS; 65°C 20× SST is 0.3*M* tris-HCl, *p*H 7.5, 3*M* NaCl, 50 m*M* EDTA.
12. J. M. Berg, *Science* **232**, 485 (1986); A. Klug and D. Rhodes, *Trends Biochem.* **12**, 464 (1987); R. M. Evans and S. M. Hollenberg, *Cell* **52**, 1 (1988).
13. A. J. Courey and R. Tjian, *Cell*, in press; and unpublished data.
14. Sp1-778C contains the NH_2-terminal 34 amino acid residues of the *lacZ'* gene encoded by pBluescript SK(+) (Stratagene) fused to the COOH-terminal amino acid residues of Sp1. Sp1-516C contains the NH_2-terminal 11 amino acid residues of the *lacZ'* gene encoded by pUC118 fused to the COOH-terminal 516 amino acid residues of Sp1. Sp1-327C contains the NH_2-terminal 17 amino acid residues of the *lacZ'* gene encoded by pUC118 fused to the COOH-terminal 327 amino acid residues of Sp1. Sp1-236C contains the NH_2-terminal 17 amino acid residues of the *lacZ'* gene encoded by pUC118 fused to the COOH-terminal 236 amino acid residues of Sp1. Sp1-168C contains the NH_2-terminal 11 amino acid residues of the *lacZ'* gene encoded by pUC118 fused to the COOH-terminal 168 amino acid residues of Sp1. The plasmid vector that was used to express Sp1 in *E. coli* is pKK233-3 [E. Amann, J. Brosius, M. Ptashne, *Gene* **25**, 167 (1983).] The identity of each construction that expresses Sp1 in *E. coli* was confirmed by DNA sequencing.
15. Sp1 proteins were expressed in *E. coli* and purified to greater than 50% homogeneity as follows. Crude cell extracts were prepared as described by Kadonaga *et al.* (*7*), except that the 1 hour incubation of the lysed cells in 4*M* urea at 4°C was replaced by a 90-min incubation in 6*M* urea at room temperature. The purity of Sp1 in these extracts was less than 1%. After brief centrifugation to remove insoluble material, the proteins were subjected to sequence-specific DNA affinity chromatography (*5*). A crude extract (from 500 ml culture) in buffer Z^e [25 m*M* Hepes (K^+), *p*H 7.5, 20% (v/v) glycerol, 0.1% (v/v) Nonidet P-40, 1 m*M* dithiothreitol (DTT), and 10 μ*M* $ZnSO_4$] containing 0.15*M* KCl was combined with 0.5 ml of poly[d(I-C)] [in 10 m*M* tris-HCl, *p*H 7.8, 100 m*M* NaCl, and 0.1 m*M* EDTA]; optical density (260 nm) = 10; and 50 μg of sonicated calf thymus DNA, incubated for 10 min, and then applied to a 1-ml DNA affinity resin. The resin was washed four times with 3 ml of the same buffer. Protein was eluted by successive addition of 1-ml portions of buffer Z^e containing 0.2*M*, 0.3*M*, 0.4*M*, 0.4*M*, 0.45*M*, 0.45*M*, 0.5*M*, 0.5*M*, 0.6*M*, 0.7*M*, 0.8*M*, 0.9*M*, 1.0*M*, and 1.0*M* KCl, and 1-ml fractions were collected. The identity of each of the purified *E. coli*–synthesized Sp1 proteins was confirmed by immunoblot analysis (*8, 16*). Purified preparations of Sp1-778C and Sp1-516C contained additional proteins, which were recognized by Sp1 antibodies, that were probably derived from translation initiation at Met codons located 618 and 447 amino acid residues from the COOH-terminus of Sp1. The apparent molecular mass of these proteins, as determined by SDS-gel electrophoresis (*8*), is consistent with the expected mobility of Sp1-618C and Sp1-447C polypeptides. In addition, both of the putative Sp1-618C and Sp1-447C proteins are synthesized in *E. coli* in the absence of an upstream translation initiation codon.
16. S. P. Jackson and R. Tjian, *Cell* **55**, 125 (1988).
17. Sp1-516C elutes from the DNA affinity resin at the same KCl concentration as Sp1-778C (Fig. 3A), whereas Sp1-236C and Sp1-168C elute from the affinity resin at the same KCl concentration as Sp1-327C (Fig. 3B) (*8*).
18. We are interchangeably using the terms "low and

high affinity" and "low and high salt eluting" because under the conditions at which Sp1-327C, Sp1-236C, and Sp1-168C elute from the column, the Sp1-778C and Sp1-516C proteins remain bound to the resin. In other words, the high salt-eluting proteins have a higher affinity to DNA than the low salt-eluting proteins at salt concentrations at which the low salt forms are eluted from the resin. These experiments do not directly address, however, the relative affinity of the larger versus the smaller Sp1 species at low salt concentrations where both species bind.

19. A. D. Johnson, B. J. Meyer, M. Ptashne, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 5061 (1979); M. Ptashne *et al.*, *Cell* **19**, 1 (1980).
20. I. A. Hope and K. Struhl, *Cell* **46**, 885 (1986); G. Gill and M. Ptashne, *ibid.* **51**, 121 (1987); E. Giniger and M. Ptashne, *Nature* **330**, 670 (1987); J. Ma and M. Ptashne, *Cell* **48**, 847 (1987); *ibid.* **51**, 113 (1987).

21. We thank W. Lim for construction of the Sp1 cDNA extension libraries, S. P. Jackson for the Sp1 antibodies, and J. Brosius for the plasmid vector. We are grateful to K. Perkins, K. Arndt, and S. Jackson for improving the quality of this manuscript. We also thank K. Ronan for assistance in the preparation of the figures. J.T.K. is a Lucille P. Markey Scholar in Biomedical Science. A.J.C. is supported by a postdoctoral fellowship from the American Cancer Society. Supported in part by a research grant from the NCI to R.T.

15 July 1988; accepted 5 October 1988

The *elav* Gene Product of *Drosophila*, Required in Neurons, Has Three RNP Consensus Motifs

STEVEN ROBINOW, ANA REGINA CAMPOS, KWOK-MING YAO, KALPANA WHITE

A sequence of developmental events transforms neurons from their immature state to their mature, terminally differentiated state. The *elav* locus is one of the first examples of a gene that is expressed in neurons early during this developmental sequence. This gene has been shown to be required for the proper development of young neurons and for the maintenance of mature neurons. DNA sequence data presented in this report suggest that the *elav* gene product is an RNA binding protein, based on the presence of RNP (ribonucleotide) consensus sequences. This leads to the proposal that this protein is involved in the RNA metabolism of neurons.

AS A NEWLY BORN NEURON MATURES, it must alter its morphology and physiology in order to become a functional component of a nervous system. To investigate the genetic basis of these events of neuronal maturation we studied the locus *embryonic lethal, abnormal visual system (elav)* of the fruit fly *Drosophila melanogaster*. Genetic and molecular data concerning this locus suggest that the *elav*-encoded function is required in all neurons from their birth, throughout their maturation, and during their maintenance (*1–5*). Transcripts from this locus are expressed in young neurons but not in neuroblasts, the neuronal progenitor cells (*5*). Thus, transcription of *elav* is one of the first demonstrated molecular events in the process of neuronal maturation.

To gain insight into the molecular nature of the product or products of this locus, we determined the nucleotide sequence of a 2.5-kb embryonic cDNA clone (*elav* cDNA-1) (*3*) and of a 9.2-kb genomic DNA fragment. An analysis of these data reveal an open reading frame (ORF) that can code for a 50.76-kD protein. We present molecular data that show that this ORF encodes a product of the *elav* locus. An analysis of the deduced amino acid sequence of this ORF reveals three repeats of two previously defined consensus sequences; an octapeptide

Department of Biology, Brandeis University, Waltham, MA 02254-9110.

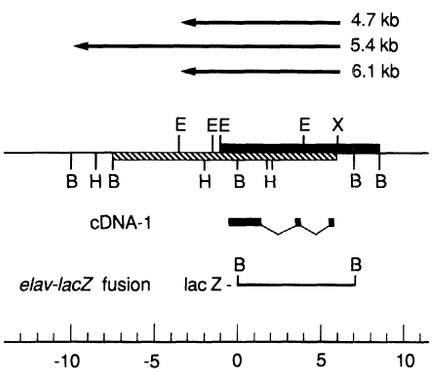

Fig. 1. Genomic organization of the *elav* locus. Above the genomic DNA restriction map, each embryonic transcript is drawn such that the underlying genomic region detects that transcript by RNA analysis (*3*). The maps of the 5.4-kb and 6.1-kb transcripts have been revised from the map shown in a previous paper (*3*). The black bar on the genomic map represents the 9.2-kb of genomic DNA sequenced. The hatched region of the genomic map indicates a fragment that contains sufficient information to provide the *elav* function required for viability (*3*). The splicing pattern of cDNA-1 is shown below the genomic map. Also diagrammed is an *elav-lacZ* gene fusion that has been constructed such that the *lacZ* gene is fused at the Bam HI site (genomic position −1) in frame with the 1449-nucleotide ORF (*16*). Transcription from this construct is driven from the *elav* promoter. The Elav-LacZ fusion protein is predicted to have a molecular size of 161 kD on the basis of an analysis of the DNA sequence of the fusion gene. The genomic DNA and cDNA-1 were subcloned into appropriate vectors [Bluescript (Stratagene) and pGEM1 (Promega)] and sequenced by established techniques (*23*). In all cases, the sequence was determined for both strands.

```
  1  MDFIMANTGA GGGVDTQAQL MQSAAAAAAV AATNAAAAPV QNAAAVAAAA QLQQQQVQQA
 61  ILQVQQQQTQ QAVAAAAAAV TQQLQQQQQA VVAQQAVVQQ QQQQAAAVVQ QAAVQQAVVP
121  QPQQAQPNTN GNAGSGSQNG SNGSTETRTN LIVNYLPQTM TEDEIRSLFS SVGEIESVKL
                                       ******
181  IRDKSQVYID PLNPQAPSKG QSLGYGFVNY VRPQDAEQAV NVLNGLRLQN KTIKVSFARP
                                  ••••••••
241  SSDAIKGANL YVSGLPKTMT QQELEAIFAP FGAIITSRIL QNAGNDTQTK GVGFIRFDKR
               * *****                                    • •••••••
301  EEATRAIIAL NGTTPSSCTD PIVVKFSNTP GSTSKIIQPQ LPAFLNPQLV RRIGGAMHTP
361  VNKGLARFSP MAGDMLDVML PNGLGAAAAA ATTLASGPGG AYPIFIYNLA PETEEAALWQ
                                                  ******
421  LFGPFGAVQS VKIVKDPTTN QCKGYGFVSM TNYDEAAMAI RALNGYTMGN RVLQVSFKTN
                                  ••••••••
481  KAK°
```

Fig. 2. Conceptual translation of the 1449-nucleotide ORF. A single letter representation of the 483–amino acid protein is shown. A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; and Y, tyrosine. This protein has a predicted molecular size of 50,760 daltons. Stretches of alanines (A) and glutamines (Q) are found between amino acids 24 and 126. Seventy-three percent of these residues are either alanine or glutamine. •••••• Underlie the residues of the RNP1 motif. ****** Underlie the residues of the RNP2 motif. ° Indicates a stop codon that terminates the ORF.