JOURNAL OF VIROLOGY, June 1988, p. 2186–2190
0022-538X/88/062186-05$02.00/0

Vol. 62, No. 6

Identification of Three Sequence-Specific DNA-Binding Proteins Which Interact with the Rous Sarcoma Virus Enhancer and Upstream Promoter Elements

G. H. GOODWIN

Institute of Cancer Research, Chester Beatty Laboratories, Fulham Road, London SW3 6JB, United Kingdom

Received 17 November 1987/Accepted 29 February 1988

Three avian nuclear proteins which bind to the Rous sarcoma virus long terminal repeat have been detected. Two of the proteins bind to sequences within the enhancer, and the third protein binds to a sequence spanning the enhancer and an upstream promoter region.

The long terminal repeat (LTR) of the Rous sarcoma virus (RSV) contains multiple DNA regulatory elements typical of eucaryotic promoters and enhancers. These are contained within the U3 region of the LTR (3, 11, 12, 14). Upstream of the TATA box there is a promoter region between bases −51 and −138 and an enhancer element between bases −139 and −229 (see Fig. 1). In chickens, the virus predominantly causes tumors arising from mesenchymal tissues, and in transgenic mice, the LTR of the virus is most active in mesenchymal tissues (16, 21). However, the virus can transform avian erythroid cells (17), and the enhancer-promoter elements are active in a wide variety of cells (3, 11, 12, 14), including erythrocytes (4) and an avian erythroid precursor cell line (HD3) which is transformed by the avian erythroblastosis virus (2) (C. A. Wright and G. H. Goodwin, unpublished data). To characterize the *trans*-acting factors required for activating the enhancer and promoter elements in erythroid cells, nuclei from HD3 cells, 14-day embryonic chicken erythrocytes, and adult chicken erythrocytes were extracted with 0.3 M NaCl as described previously (18), and the extracts were analyzed for sequence-specific binding proteins by using gel retardation and footprinting methods.

The *Eco*RI LTR fragment from the duplicated LTR present in the complete proviral recombinant pSRA2 (22) was subcloned into the plasmid pAT153 (Fig. 1). This plasmid was used to isolate end-labeled fragments containing the enhancer and upstream promoter regions. The plasmid was cut with *Hin*dIII, 5′-end labeled with polynucleotide kinase or 3′-end labeled with reverse transcriptase, and recut with *Hin*fI. The labeled fragment was isolated by polyacrylamide gel electrophoresis and purified, using Schleicher and Schuell Elutip columns (18). The *Eco*RI-*Pvu*I fragment, labeled at the *Eco*RI site, was similarly isolated. Gel retardation analyses (5, 7) were carried out by mixing the labeled DNA (2 ng) with nuclear protein (5 to 20 μg) and poly(dI-dC) · poly(dI-dC) (9 μg) in 20 μl of 60 mM NaCl–10 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid (HEPES) (pH 7.9)–5 mM $MgCl_2$–1 mM dithiothreitol–10% glycerol. pBR322 DNA (1 to 3 μg) was also added in some assays as indicated. The protein-DNA complexes were then electrophoretically separated on 5% polyacrylamide gels containing 10 mM Tris, 10 mM borate, 0.2 mM EDTA (pH 8.3) and detected by autoradiography. Methylation interference binding studies (20) were carried out by partially methylating the G bases of end-labeled DNA fragments with dimethyl sulfate as described by Maxam and Gilbert (13). The DNA was then incubated with nuclear protein in the same way as it was for the gel retardation analyses, except that the quantities of protein and DNA used were about tenfold higher. The complexes were separated on 5% polyacrylamide gels, and the radioactive bands were detected by autoradiography of the wet gels. The DNA was purified from the protein-bound and protein-free bands, as described above, and cleaved at methylated G bases with piperidine (13), and the DNA fragments were resolved on a sequencing gel. Footprint analyses were carried out as described previously (18) by mixing DNA (2 ng) and nuclear protein (0.5 to 250 μg) in 20 or 100 μl of buffer (20 mM HEPES [pH 7.9], 5 mM $MgCl_2$, 0.1 mM EDTA, 1 mM dithiothreitol) containing 50 mM NaCl, 20% glycerol, 3 to 5 μg of poly(dI-dC) · poly(dI-dC), and 2 mM $CaCl_2$. The samples were digested with DNaseI, and the extracted DNA was analyzed on a sequencing gel.

Gel retardation analysis of the HD3 0.3 M NaCl nuclear extract gave three retarded bands, BI, BII, and BIII, due to specific binding of proteins to the *Hin*fI-*Hin*dIII DNA fragment (Fig. 2A, lane 5). Addition of 100-fold excess of the unlabeled competitor fragment *Eco*RI-*Sph*I (Fig. 1) resulted in a depletion of BI and BIII (Fig. 2A, lane 2), while the addition of fragment *Sph*I-*Eco*RI caused depletion of BII (Fig. 2A, lane 3). Addition of the *Eco*RI-*Eco*RI fragment (Fig. 2A, lane 4), but not plasmid pAT153 digested with *Hin*fI (lane 1), caused reduction of all three bands. The same results were obtained by using the *Eco*RI-*Pvu*I-labeled fragment (data not shown). These results are consistent with the interpretation that two proteins (designated I and III in this report) bind to the LTR between the *Hin*fI and *Sph*I sites, giving bands BI and BIII, and a third protein (designated II) binds between the *Sph*I and *Pvu*I sites, giving band BII. The polyomavirus enhancer (1) also appears to have sequences which bind these three proteins, since the 134-base-pair *Pvu*II fragment of this enhancer was able to compete with the RSV LTR for the binding of proteins I and II and partially competed for the binding of protein III (Fig. 2B, lane 2 and Fig. 2C, lane 2). A DNA fragment containing the consensus-binding sequence for the NF1/TGGCA-binding protein does not compete for the binding of these proteins to the RSV LTR (Fig. 2B, lane 1).

Analysis of the HD3 nuclear extract by footprinting (Fig. 3A) revealed two regions of protection, FI and FIII, within

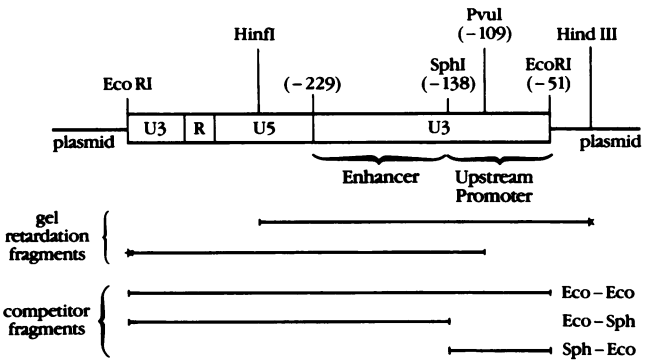

FIG. 1. Structure of the subcloned RSV LTR in plasmid pAT153. The 0.32-kilobase *Eco*RI fragment of the duplicated LTR in plasmid pSRA2 (which contains the complete Schmidt-Ruppin A2 RSV DNA [22]) subcloned into pAT153 is shown as a box containing the U3, R, and U5 sequences. (This plasmid was a generous gift from J. Wyke.) The U3 region contains the enhancer and promoter element which are upstream of a TATA box (at −30 base pairs) in the complete proviral DNA. The numbers in brackets refer to base-pair distances upstream from the start of transcription of the proviral DNA. Also shown are the DNA fragments which were ^{32}P-labeled for gel retardation and footprinting analyses and DNA fragments that were used for competition experiments.

the RSV enhancer region (see Fig. 5). Rather surprisingly, no footprint indicative of protein II binding was detected between the *Sph*I and *Pvu*I sites with the HD3 extract. However, gel retardation analysis of 0.3 M NaCl nuclear extracts from embryonic and adult chicken erythrocytes gave a strong BII band with both extracts, but bands BI and BIII were absent (data not shown). The DNA-binding activity of protein II in the chicken erythrocyte extract was further purified by DNA-cellulose chromatography carried out as described previously (18), the activity giving the BII gel retardation complex eluting with 0.25 M $(NH_4)_2SO_4$ (the C4 fraction). When this fraction was used for footprinting on the LTR, an FII footprint was seen on both strands around the *Sph*I site (Fig. 3B and see Fig. 5).

Methylation interference gel retardations were also carried out to determine the binding sites of proteins I, II, and III from HD3 cells. The HD3 nuclear extract was first further fractionated by MonoQ (Pharmacia) ion-exchange fast-protein liquid chromatography (FPLC) by loading 2 mg of protein onto the column in the buffer described above containing 100 mM NaCl and 0.025% *n*-octyl-D-glucopyranoside. The proteins were eluted with a salt gradient; protein III eluted in the excluded peak, and protein I eluted with approximately 0.25 M NaCl. Protein II activity was not recovered from this column. Fractions containing the activities of proteins I and III and the total HD3 extract were incubated with the partially methylated *Hin*fI-*Hin*dIII LTR fragment labeled at the *Hin*dIII site on either strand. After electrophoretic separation, bands BI, BII, and BIII and the unbound fragments were eluted from the gel and cleaved at methylated G bases with piperidine. Equal amounts of radioactivity from retarded and unretarded bands were loaded onto a sequencing gel (Fig. 4). The results are summarized in Fig. 5. For the BI band, cleavage at three Gs on each strand within the FI footprint region was reduced (Fig. 4A, lanes 3 and 6, and Fig. 4B, lanes 2 and 4), indicating that methylation of these Gs inhibits protein I binding. For the BII band, cleavage at three Gs on the top strand on either side of the *Sph*I site was reduced (Fig. 4A, lanes 2 and 5). All three Gs were within the FII footprint obtained with the chicken erythrocyte protein II. The Gs on the bottom strand were unaffected (data not shown). For BIII, cleavage at one G on each strand in the FIII footprint region was reduced (Fig. 4A, lanes 4 and 7, and Fig. 4B, lanes 1 and 3). These results, therefore, confirm that in erythroid cells there are three proteins (I, II, and III) which bind to the footprint regions FI, FII, and FIII, respectively. Proteins I and III bind to the enhancer region, and protein III binds to a region spanning the junction of the upstream promoter and the enhancer. Gel retardation analyses indicated that these three factors are present in the undifferentiated HD3 cells but only protein II is present in 14-day embryonic chicken erythrocytes and adult chicken erythrocytes, suggesting that factors I and III are developmentally regulated. The three proteins are present in chicken embryo fibroblasts (data not shown).

Protein I bound to a sequence within the enhancer containing the imperfect inverted repeat TGGTAAC. . .GT TAGCA. The symmetry of the binding site is partly reflected in the Gs found to be important for binding in the methylation interference analysis. The polyomavirus enhancer competed for the binding of this protein (Fig. 2C, lane 2),

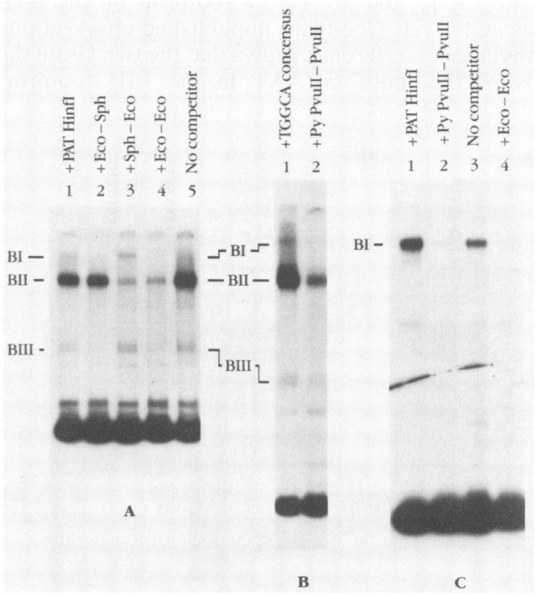

FIG. 2. Gel retardation analyses of the binding of HD3 proteins to the LTR. The 0.25-kilobase (kb) *Hin*fI-*Hin*dIII labeled fragment (2 ng) was incubated with 0.3 M NaCl-extracted HD3 nuclear protein (15 μg), poly(dI-dC) · poly(dI-dC) (9 μg), and pBR322 (3 μg) together with various unlabeled competitor DNA fragments (see Fig. 1) before electrophoresis. (A) Lanes: 1, 250 ng of plasmid pAT153 (PAT) digested with *Hin*fI; 2, 250 ng of the 0.23-kb *Eco*RI-*Sph*I LTR fragment; 3, 150 ng of the 0.09-kb *Sph*I-EcoRI LTR fragment; 4, 250 ng of the 0.32-kb EcoRI-EcoRI LTR fragment; 5, no additional competitor DNA. BI, BII, and BIII are retarded bands due to specific binding of proteins to the *Hin*fI-*Hin*dIII DNA fragment. (B) Lanes: 1, 140 ng of the 0.32-kb cloned fragment containing the consensus sequence TGGCAGCTTGC CAA which binds the TGGCA-binding protein NF1. 2, 150 ng of the 0.13-kb *Pvu*II fragment of the polyomavirus (Py) enhancer. (C) The labeled *Hin*fI-*Hin*dIII fragment (2 ng) was incubated with 3 μl of protein I partially purified by FPLC MonoQ ion-exchange chromatography together with poly(dI-dC) · poly(dI-dC) (9 μg), pBR322 (1 μg), and various competitor fragments before electrophoresis. Lanes: 1, 300 ng of plasmid pAT153 (PAT) digested with *Hin*fI; 2, 150 ng of the 0.13-kb *Pvu*II polyomavirus (Py) enhancer fragment; 3, no additional competitor DNA; 4, 125 ng of the 0.32-kb EcoRI-EcoRI LTR fragment.

although it does not have this inverted repeat sequence (6). However, it does have the inverted repeat AGTTGCTAGG CAACT, which is known to be necessary for the binding of a nuclear factor (15). Since the protein I-binding site in the RSV enhancer contains the sequence GTTGCTA, these two factors could be related and may explain the competition results.

Protein II bound to a sequence containing the CAAT-box motif CCAAT (on the bottom strand), and the two Gs of this motif (on the top strand) are involved in the binding of this protein. Since the polyomavirus *Pvu*II enhancer fragment has a CCAAT motif (6), this probably explains the ability of the polyomavirus enhancer to compete with the RSV LTR for protein II binding (Fig. 2B, lane 2). Protein II may therefore be a CAAT-box-binding protein, but it is not the same as the previously described CAAT-box factor CTF which has been shown to be the same as the NF1/TGGCA-binding factor (10), since the gel retardation analysis using an unlabeled competitor fragment containing the NF1/TGGCA-consensus sequence did not cause reduction of the BII band (Fig. 2B, lane 1). Also, the TGGCA-binding protein, which was partially purified from erythrocytes by DNA-affinity chromatography, did not give a footprint on the LTR (R. Nicolas and G. H. Goodwin, unpublished observations).

Methylation interference shows that the two Gs on the two strands of the motif GCAAT in the FIII footprint are

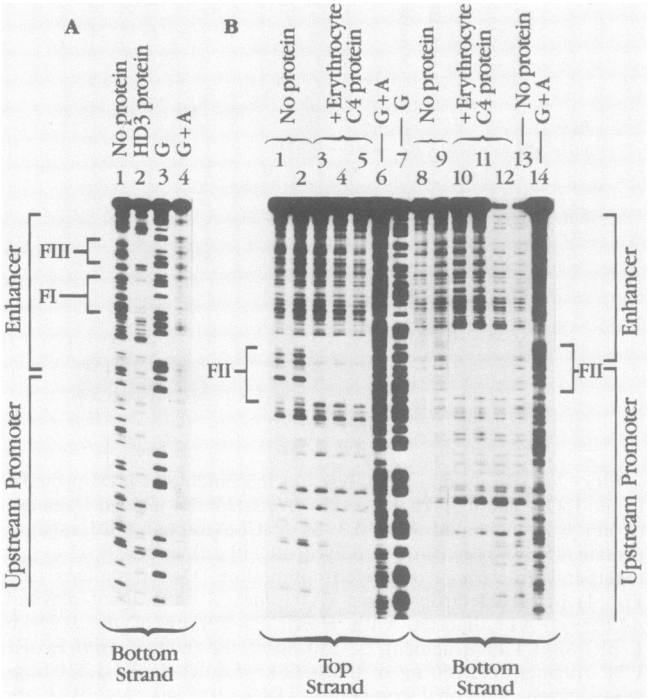

FIG. 3. Footprint analyses of erythroid nuclear proteins. (A) The 0.25-kilobase *Hin*fI-*Hin*dIII fragment which was 5′-end labeled at the *Hin*dIII site was incubated with no protein (lane 1) or 250 μg of 0.3 M NaCl-extracted protein from HD3 nuclei (lane 2) together with 3 μg of poly(dI-dC) · poly(dI-dC), digested with DNaseI, and analyzed on a sequencing gel. Lanes 3 and 4 show G and G + A cleavage reactions. (B) The *Hin*fI-*Hin*dIII fragment 5′- and 3′-end labeled at the *Hin*dIII site was incubated with 5 μg of poly(dI-dC) · poly(dI-dC) and no protein (lanes 1, 2, 8, 9, and 13) or with DNA-cellulose-purified chicken erythrocyte C4 protein fraction (lanes 3 and 12, 0.5 μg of protein; lanes 4 and 10, 1 μg of protein; lanes 5 and 11, 2 μg of protein). Lanes 6, 7, and 14 show G and G + A cleavage reactions.

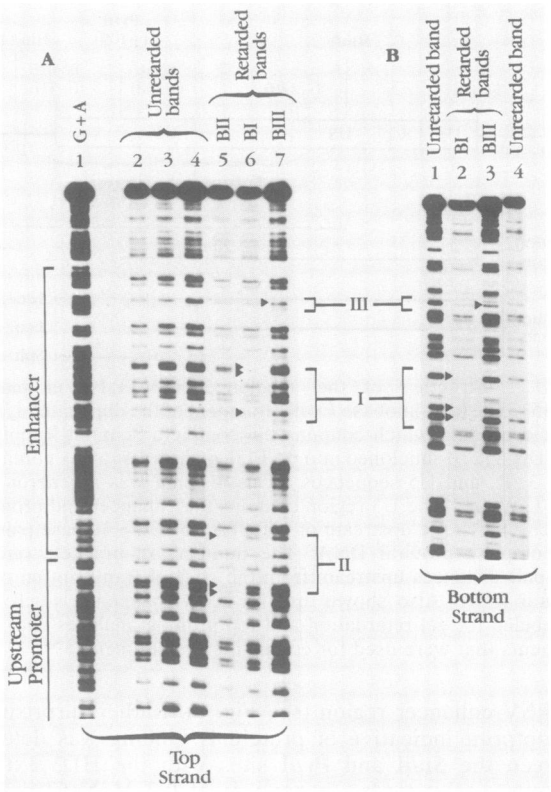

FIG. 4. Methylation interference analysis of the binding of proteins I, II, and III. The *Hin*fI-*Hin*dIII LTR fragment labeled at the *Hin*dIII site on either strand was methylated with dimethyl sulfate and used in scaled-up gel retardation analyses with FPLC MonoQ-purified proteins I and III and the total HD3 0.3 M NaCl nuclear extract. After autoradiographic exposure of the electrophoretic gel, the retarded bands I, II, and III, together with the unretarded counterparts, were excised from the gel. The DNA was purified, cleaved with piperidine, and analyzed on sequencing gels. (A) Top strand. Lanes: 1, G + A cleavage reaction; 2 and 5, DNA from unretarded band and retarded BII band, respectively, produced by incubation with HD3 nuclear extract; 3 and 6, DNA from unretarded band and retarded BI band, respectively, produced by incubation with protein I; 4 and 7, DNA from unretarded band and retarded BIII band, respectively, produced by incubation with protein III. (B) Bottom strand. Lanes: 1 and 3, DNA from unretarded band and retarded BIII band, respectively, produced by incubation with protein III; 4 and 2, DNA from unretarded band and retarded BI band, respectively, produced by incubation with protein I. Arrowheads indicate Gs exhibiting reduced cleavage.

important for the binding of protein III. This protein may therefore be related to the CAAT-box-binding protein described by Graves et al. (8). This factor binds to the herpes simplex virus thymidine kinase CCAAT box with greater affinity when the sequence is mutated to GCAAT, and it also appears to bind to the polyomavirus, simian virus 40, and murine sarcoma virus enhancer core sequences (9). The relationship between CAAT-box-binding protein and protein III is not clear, since the polyomavirus enhancer does not compete strongly with the RSV enhancer for protein III binding (Fig. 2B, lane 2). Also, while this article was being prepared, Sealey and Chalkley (19) reported the presence of two factors in quail fibroblasts which appear to be similar to the proteins II and III described here, and they found that the murine sarcoma virus and simian virus 40 enhancers did not compete with the RSV enhancer for binding of protein III.

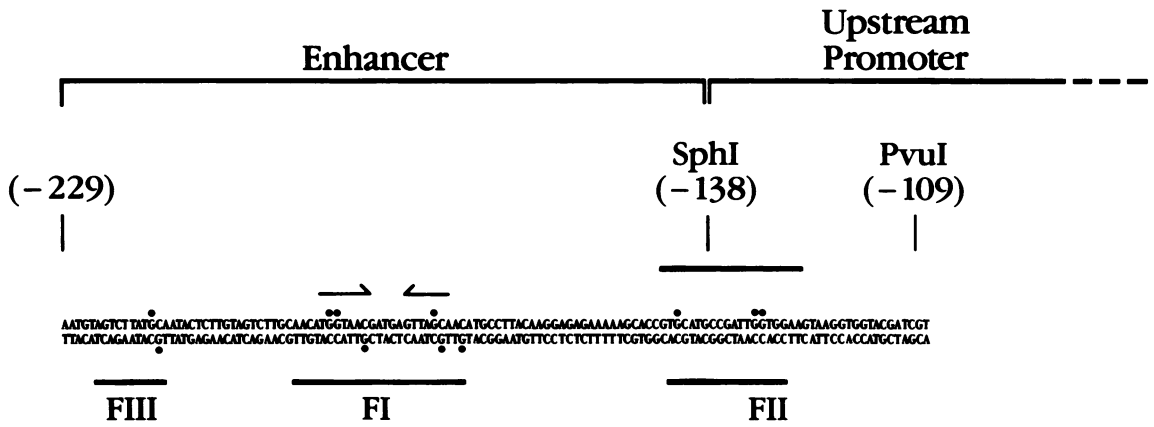

FIG. 5. Sequence of the enhancer element and part of the upstream promoter element showing the binding sites for proteins I, II, and III. The thick horizontal lines show the footprints FI, FII, and FIII shown in Fig. 3. (Note that the exact boundaries of the footprints are not definite due to the presence of sequences which are poorly cut by DNaseI in the no-protein controls.) The filled circles indicate the Gs at which methylation inhibits protein bindings, as determined by the gel retardation analysis of Fig. 4. Note that FIII has the GCAAT motif on the top strand, FII has the CCAAT motif on the bottom strand, and FI has the inverted repeat shown by the arrows.

Proteins I and III bind to LTR regions that have been shown to be important for enhancement (3, 11, 12, 14). Thus, for example, deletion of 48 base pairs from the 5′ end of the enhancer element (which removes the protein III site and part of the protein I site) results in impaired enhancer activity (11). The role of the sequences bound by protein II is not so clear. Deletion of the four central nucleotides (CATG) of the *Sph*I site adjacent to the putative CCAAT motif has little effect on transcription (14), and insertion of various lengths of DNA into the *Sph*I site has minimal effect unless the inserts are large (3, 14). These results suggest that sequences just upstream of the CCAAT motif are not important but that the spacing between the upstream promoter region and the enhancer may be important. More extensive deletions around the *Sph*I site (e.g., deletion of the *Sph*I-*Pvu*I fragment [3], which deleted the CCAAT motif) reduced transcriptional activity substantially, but this also could be due to altered spacing between the enhancer and promoter. More detailed mutational analysis of this region is required to ascertain the importance of the protein II-binding site.

This research was supported by a grant from the Medical Research Council and the Cancer Research Campaign. The author thanks J. A. Wyke, A. Sippel, P. Rigby, and P. Gruss for plasmids containing the RSV LTR, the TGGCA-binding sequence, and the polyomavirus enhancer and R. Nicolas for the erythrocyte C4 protein preparation.

LITERATURE CITED

1. **Amati, P.** 1985. Polyoma regulatory region: a potential probe for mouse cell differentiation. Cell **43**:561–562.
2. **Beug, H., G. Doederlein, C. Freudenstein, and T. Graf.** 1982. Erythroblast cell lines transformed by a temperature sensitive mutant of avian erythroblastosis virus: a model system to study erythroid differentiation in vitro. J. Cell. Physiol. **1**(Suppl.): 195–200.
3. **Cullen, B. R., K. Raymond, and G. Ju.** 1985. Functional analysis of the transcription control region located within the avian retroviral long terminal repeat. Mol. Cell. Biol. **5**:438–447.
4. **Emerson, B. M., J. M. Nickol, P. O. Jackson, and G. Felsenfeld.** 1987. Analysis of the tissue-specific enhancer at the 3′ end of the chicken β-globin gene. Proc. Natl. Acad. Sci. USA **84**:4786–4790.
5. **Friend, M., and D. M. Crothers.** 1981. Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. Nucleic Acids Res. **9**:6505–6525.
6. **Fujimura, F. K.** 1986. Nuclear activity from F9 embryonal carcinoma cells binding specifically to the enhancers of wild-type polyoma virus and PyEC mutant DNAs. Nucleic Acids Res. **14**:2845–2866.
7. **Garner, M. M., and A. Revzin.** 1981. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the E. coli lactose operon regulatory sequence. Nucleic Acids Res. **9**:3047–3060.
8. **Graves, B. J., P. F. Johnson, and S. L. McKnight.** 1986. Homologous recognition of a promoter domain common to the MSV LTR and HSV tk gene. Cell **44**:565–576.
9. **Johnson, P. F., W. H. Landschulz, B. J. Graves, and S. L. McKnight.** 1987. Identification of a rat liver nuclear protein that binds to the enhancer core element of three animal viruses. Genes Devel. **1**:133–146.
10. **Jones, V. A., J. T. Kadonaga, P. J. Rosenfeld, T. J. Kelly, and R. Tjian.** 1987. A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. Cell **48**:79–89.
11. **Laimins, L. A., P. Tsichlis, and G. Khoury.** 1984. Multiple enhancer domains in the 3′ terminus of the Prague strain of Rous sarcoma virus. Nucleic Acids Res. **12**:6427–6442.
12. **Luciw, P. A., J. M. Bishop, H. E. Varmus, and M. R. Capecchi.** 1983. Location and function of retroviral and SV40 sequences that enhance biochemical transformation after microinjection of DNA. Cell **33**:705–716.
13. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
14. **Norton, P. A., and J. M. Coffin.** 1987. Characterization of Rous sarcoma virus sequences essential for viral gene expression. J. Virol. **61**:1171–1179.
15. **Ostapchuk, P., J. F. X. Diffley, J. T. Bruder, B. Stillman, A. J. Levine, and P. Hearing.** 1986. Interaction of a nuclear factor with the polyomavirus enhancer region. Proc. Natl. Acad. Sci. USA **83**:8550–8554.
16. **Overbeek, P. A., S.-P. Lai, K. R. Van Quill, and H. Westphal.** 1986. Tissue-specific expression in transgenic mice of a fused gene containing RSV terminal sequences. Science **231**:1574–1577.
17. **Palmieri, S.** 1985. Transformation of erythroid cells by Rous sarcoma virus. Virology **140**:269–280.
18. **Plumb, M. A., R. H. Nicolas, C. A. Wright, and G. H. Goodwin.** 1985. Multiple sequence-specific DNA binding activities are eluted from chicken nuclei at low ionic strengths. Nucleic Acids

Res. **13**:4047–4065.
19. **Sealey, L., and R. Chalkley.** 1987. At least two nuclear proteins bind specifically to the Rous sarcoma virus long terminal repeat enhancer. Mol. Cell. Biol. **7**:787–798.
20. **Sienbenlist, U., and W. Gilbert.** 1980. Contacts between E. coli RNA polymerase and an early promoter of phage T7. Proc. Natl. Acad. Sci. USA **77**:122–126.
21. **Swain, J. L., T. A. Stewart, and P. Leder.** 1987. Parental legacy determines methylation and expression of an autosomal transgene: a molecular mechanism for parental imprinting. Cell **50**: 719–727.
22. **Swanstrom, R., W. J. DeLorbe, J. M. Bishop, and H. E. Varmus.** 1981. Nucleotide sequence of cloned unintegrated avian sarcoma virus DNA: viral DNA contains direct and inverted repeats similar to those in transposable elements. Proc. Natl. Acad. Sci. USA **78**:124–128.