Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 3370–3374, May 1985
Genetics

Isolation of a gene enhancer within an amplified inverted duplication after "expression selection"

(gene amplification/gene expression/DNA transfection/polyoma virus transformation)

MIRIAM FORD, BRENDAN DAVIES, MOIRA GRIFFITHS, JOANNA WILSON, AND MIKE FRIED

Imperial Cancer Research Fund Laboratories, P. O. Box 123, Lincoln's Inn Fields, London, WC2A 3PX England

Communicated by Walter F. Bodmer, January 17, 1985

ABSTRACT **We have attempted to isolate and identify cellular expression sequences from F9 teratocarcinoma DNA by utilizing their ability to reactivate a selectable gene devoid of its own expression sequences (expression selection). Restriction nuclease-digested F9 cellular DNA was ligated to a polyoma virus (Py) DNA fragment which contains an intact transforming region but is incapable of inducing transformation because it lacks the viral 5′ enhancer sequence. The ligation mixture was used to transfect Rat-1 cells and a transformed cell line, 3B, was isolated. The 3B cell line contained a single type of Py DNA insert, which was molecularly cloned as an 18-kilobase *Bgl* II fragment. A weak cellular enhancer was identified in a 4.7-kilobase *Bam*HI fragment upstream from the Py sequences. Both the Py DNA and the enhancer sequences were found to be present in an inverted duplication in the 3B clone. The presence of this structure in 3B genomic DNA was confirmed by the analysis of selectively isolated inverted duplicated sequences, and the structure was found to be at least 22 kilobases long. In the 3B cell line, the inverted duplicated sequences containing the Py and enhancer sequences are quite stable and are amplified 20- to 40-fold. The strongly transformed phenotype of the 3B cells may be a result of this amplification. The formation of inverted duplications as a part of the amplification mechanism as well as a general strategy for the cloning of inverted duplicated (amplified) sequences is discussed.**

In mammalian cells a number of sequence elements have been shown to be important for gene expression. Several cellular expression sequences have been identified by means of their association with a cloned gene (1). An alternative approach has been to isolate expression sequences by their ability to confer expression on a selectable gene devoid of its own expression sequences (2). This "expression-selection" system involves the transfection of cellular DNA ligated to the inactive test gene into cells and selection for cells containing a reactivated test gene. The cellular expression sequences are then molecularly cloned in conjunction with the test gene. One expression sequence previously isolated in this manner has many of the properties of an enhancer element (3). We now have used the expression-selection procedure to isolate expression sequences from F9 teratocarcinoma cellular DNA. Here we report the characterization of DNA from a cell line containing an activated test gene and a cellular enhancer. A large stretch of the cellular DNA containing the test gene and cellular enhancer is found as an inverted duplication, at least 22 kilobases (kb) long, and is also amplified.

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.

MATERIAL AND METHODS

Transformed cells were isolated after transfection of Rat-1 cells (2) with the polyoma virus (Py) 4.1-kb *Bam*HI–*Eco*RV fragment ligated to restriction endonuclease-digested mouse F9 teratocarcinoma DNA. Total cellular DNA from the different transformed cell lines was analyzed by blot hybridization with a Py probe (2). Py tumor (T) antigens were detected after immunoprecipitation and gel electrophoresis (4). After fractionation of *Bgl* II-cleaved DNA in a 10–40% sucrose gradient (5), the fraction containing the 18-kb Py-probe positive fragment was inserted into the *Bam*HI site of λEMBL-4 and cloned (6), and positive plaques were isolated as described (2). Inverted duplications were isolated by a modification of the procedure of Ish-Horowicz and Pinchin (7). Total cellular DNA was denatured with 50 mM NaOH for 60 min at room temperature and then neutralized with 1 M HCl and 1 M Tris Cl (pH 7.8) on ice. The DNA was digested with nuclease S1 (50 units/μg of DNA) in 0.2 M NaCl/1 mM $ZnSO_4$/0.05 M sodium acetate, pH 4.5/0.5% glycerol for 30 min at 14°C before the reaction was terminated with 10 mM EDTA and by phenol extraction. The S1-resistant DNA was precipitated and analyzed with various restriction enzymes. The degree of gene amplification was estimated by densitometric scanning of an autoradiogram from a Southern blot of restriction nuclease-digested 7AXT (4) and 3B DNA and normalized for DNA concentration (8).

RESULTS

Isolation of the Transformed Cell Line 3B and Cloning of the Py Insert. The test gene used for reactivation was the Py transforming region. The Py fragment used extended from the *Pvu* II site at nucleotide (nt) 5265 (converted to a *Bam*HI site by linker addition) to the *Eco*RV site at nt 4106 (Fig. 1*A*). This 4.1-kb Py fragment contains the origin of DNA replication, the complete Py early region (transforming region) with its associated "TATA" box and cap site, and part of the Py late region (9) but lacks the Py enhancer region (nt 5021–5265) (10) and thus is incapable of DNA replication and greatly inhibited in its transforming ability. Total cellular DNA from mouse F9 teratocarcinoma cells was cleaved with *Bam*HI, *Bcl* I, *Bgl* II, or *Mbo* I and was ligated to the Py fragment. The single-stranded ends generated by these enzymes (G-A-T-C-) are compatible with ligation with the *Bam*HI site 5′ to the Py early region. The different ligation mixtures were individually used to transfect Rat-1 cells. Total cellular DNA, isolated from transformed cell colonies arising after transfection, was assessed for the number of Py inserts (Fig. 1*B*). Cleavage of cellular DNA containing a single Py insert with *Eco*RI (cleaves the 4.1-kb fragment once) should generate two

Abbreviations: Py, polyoma virus; T antigen, tumor antigen; kb, kilobase(s); nt, nucleotide(s).

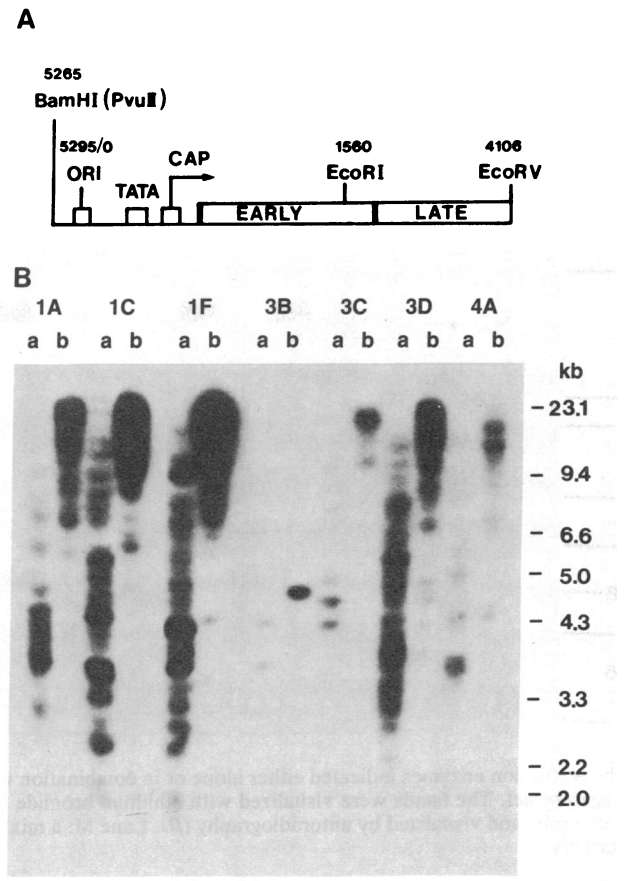

FIG. 1. *(A)* The structure of the 4.1-kb Py fragment with the location of the *Bam*HI, *Eco*RI, and *Eco*RV cleavage sites indicated. Numbers are the nt positions in the Py genome. The Py origin of replication (ORI), TATA box, cap site, and early and late regions are represented. *(B)* Restriction analysis of the integrated Py sequences in transformed rat cell lines. Cellular DNA from transformed cell lines, arising after transfection with F9 mouse DNA ligated to the *Bam*HI site of the Py fragment shown in *A*, was cleaved with either *Eco*RI (lanes a) or *Eco*RV (lanes b) and fractionated in an 0.8% agarose gel. The bands containing Py specific sequences were detected by Southern blot analysis with a Py probe and visualized by autoradiography. The notation for each cell line is shown above the corresponding pair of lanes. Size markers at right show positions of fragments in a mixture of *Hin*dIII-digested λ DNA and *Eco*RI/*Hin*dIII-digested Py DNA.

positive fragments and cleavage with *Eco*RV (does not cleave the Py 4.1-kb fragment) should generate one positive fragment. Many of the cell lines appeared to contain multiple Py inserts (many positive fragments), but one line, 3B, appeared to contain a single Py insert. Py gene expression was shown to be reactivated in the transformed cell lines by the production of Py T antigens (Fig. 2) and Py-specific early region mRNAs initiated at the normal Py cap site (data not shown). An 18-kb *Bgl* II fragment from 3B genomic DNA was found to contain the Py sequences and a large amount of 5′ flanking cellular sequences and was therefore chosen for cloning in the *Bam*HI site of λEMBL-4. From a total of 70,000 plaques screened, 7 positive plaques were isolated. Three of these were analyzed and found to contain the same insert. One, λ3B, was studied further.

Structure of the Cellular DNA Insert in the λ3B Clone. Apparent inconsistencies were observed in the restriction-enzyme and blot–hybridization analysis of the 3B cellular DNA insert in λ3B clone (Fig. 3). One was the size of the insert cloned. Since *Bgl* II-cleaved 3B cellular DNA (size-selected to be 18-kb) was ligated to the *Bam*HI cleavage site

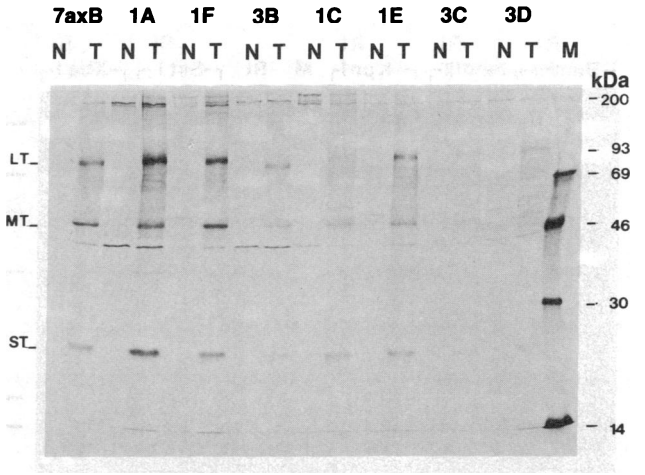

FIG. 2. T antigens in Py-transformed rat cell lines. Cultures of nearly confluent transformed cells were labeled for 4 hr at 37°C with [^{35}S]methionine as described (4). The extracts were incubated with normal serum (N) or anti-tumor serum (T), precipitated with formalin-fixed *Staphylococcus aureus* bacteria, and fractionated by $NaDodSO_4$/10% PAGE. The autoradiogram of the gel is shown. Cell line designations are above the corresponding lane pairs. 7axB is a Py-transformed rat line producing normal-sized T antigens (4). Size markers are in lane M. Positions of large (LT), middle-sized (MT), and small (ST) T antigens are shown at left.

of λEMBL-4, both the *Bam*HI and *Bgl* II sites were lost in the λ3B. The polylinker in λEMBL-4 contains an *Eco*RI site just distal to the *Bam*HI site so that digestion with *Eco*RI cleaves the insert from the λ arms and reveals the number of *Eco*RI sites in the insert. Such cleavage of λ3B generated six *Eco*RI fragments (0.45, 1.1, 1.3, 1.5, 1.9, and 3.5 kb) in addition to the 9- and 19-kb λ arms, indicating a size of about 9.8 kb for the insert (Fig. 3*A*). Different size estimates for the insert were obtained by analysis with different restriction enzymes: with *Bam*HI, 13.4 kb; *Xba* I, 10.6 kb; *Kpn* I, 18.0 kb; *Hin*dIII, 15.9 kb; and *Sst* I, 17.7 kb (Fig. 3*A*). A second inconsistency was the molar yield of the fragments generated from the insert with different restriction enzymes. The *Eco*RI 1.3-kb fragment appeared to be in half the molar yield of the other five *Eco*RI fragments (Fig. 3*A*). Similar fragments in apparent half molar yield were observed in digests with other enzymes alone or in combination with *Eco*RI [e.g., the 0.9-kb fragment for *Xba* I and the 0.9- and 0.65-kb fragments for *Bam*HI (Fig. 3*A*)]. A third problem involved the placement of the Py sequences within the insert. Blot analysis of *Eco*RI-cleaved λ3B with a Py probe showed that only the 3.5- and 1.5-kb fragments contained Py sequences (Fig. 3*B*). A restriction analysis of λ3B DNA either alone or in combination with *Eco*RI, using enzymes that cleave Py DNA on both sides of the single viral *Eco*RI site (*Pvu* II, *Pst* I, *Hpa* II), generated bona fide Py fragments (data not shown). These results showed that the 3.5- and 1.5-kb fragments are separated by the Py *Eco*RI cleavage site. This indicated that there was only one block of Py sequences in the insert, in apparent contradiction to the results obtained by cleavage with *Sst* I, *Bam*HI, *Hin*dIII, *Kpn* I, and *Xba* I for which fragments from both λ arms are found joined to Py sequences (Fig. 3*B*). Treatment of these enzyme digests with *Eco*RI separates the Py sequence of the insert from the λ arm sequences (Fig. 3*B*). These results indicated that Py sequences exist at both ends of the 3B insert. From the above results and from analysis with a number of other restriction enzymes, all the inconsistencies were reconciled in the map of the 3B insert presented in Fig. 4. The 18-kb *Bgl* II insert contains an inverted duplication of approximately 8.5 kb separated by about 1 kb of noninverted sequence. This accounts for the

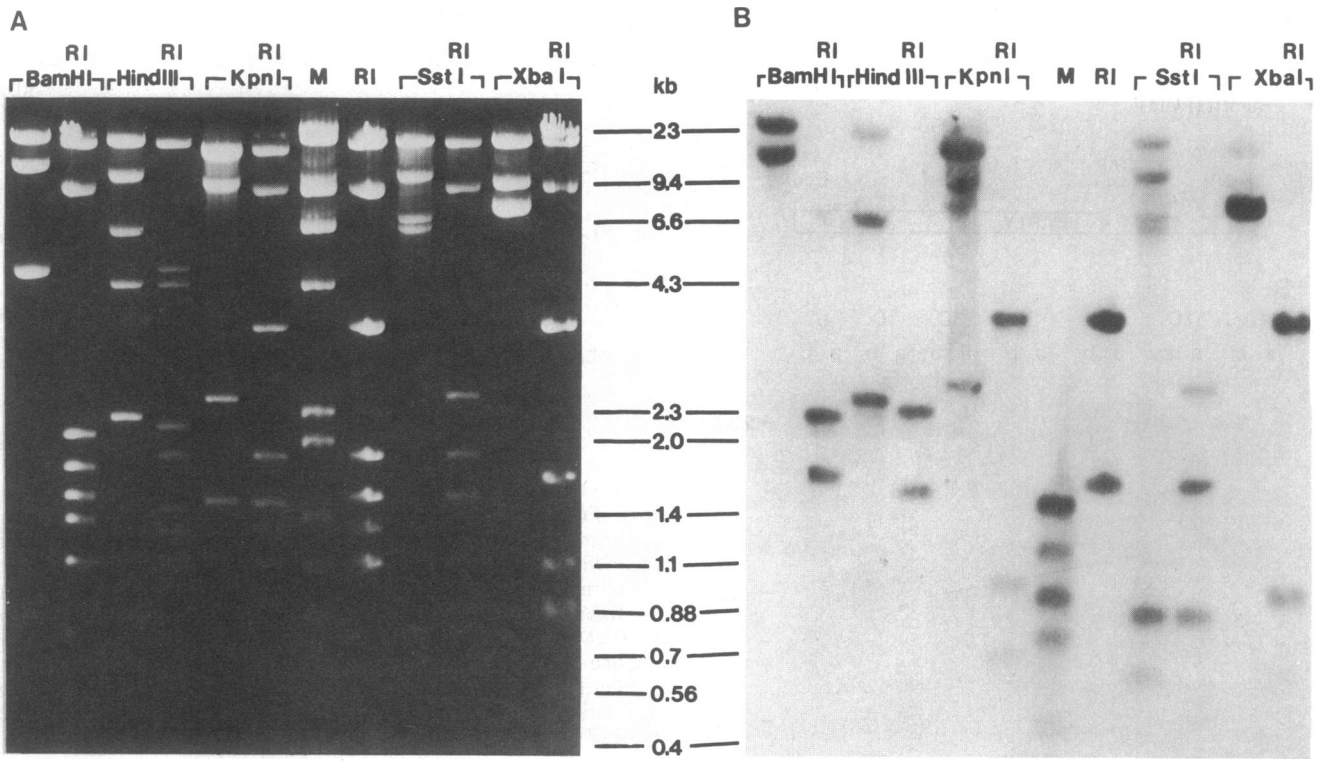

FIG. 3. Restriction nuclease analysis of λ3B. DNA was digested with the restriction enzymes indicated either alone or in combination with *Eco*RI (lanes RI), and 3 μg of each digest was electrophoresed in a 0.6% agarose gel. The bands were visualized with ethidium bromide *(A)*. The Py-specific sequences were detected by Southern blot analysis with a Py probe and visualized by autoradiography *(B)*. Lane M: a mixture of *Hin*dIII-digested λ DNA and *Hpa* II-digested Py DNA, used as size markers.

different molar yield of some of the fragments (1 or 2 molar yield) and differences in the estimated size of the insert with different restriction enzymes. The Py sequences were found to be within the inverted duplication close to the ends of the 3B insert near the λ arm sequences. Approximately 1.3 kb of sequence from the *Eco*RV site of the 4.1-kb Py vector has been lost in 3B DNA (compare with map of Py fragment, Fig. 1*A*). This deletion leaves the coding regions for the Py middle-sized and small T antigens intact but results in the loss of about 40 codons for the COOH-terminal end of the large T antigen (DNA replication protein). This explains the smaller-sized large T antigen found in 3B cells (Fig. 2).

Analysis of the Inverted Duplication in 3B Genomic DNA. It was surprising to find an inverted duplication in the cellular DNA insert in λ3B. A restriction analysis of 3B genomic DNA generated similar positive fragments to those generated from the λ3B insert, showing that this structure was not an artifact of cloning (compare Figs. 3*B* and 5*B*: e.g., *Sst* I-generated fragments of 0.9, 6.3, and 7.2 kb). To confirm both the presence and extent of the inverted duplication, 3B genomic DNA was treated with the single-strand-specific nuclease S1 after denaturation and rapid renaturation. The outline of this procedure is shown in Fig. 5*A*. Upon rapid renaturation of denatured DNA that contains an inverted duplication, intrastrand duplexes will form with high efficiency. These intrastrand duplexes will be resistant to the

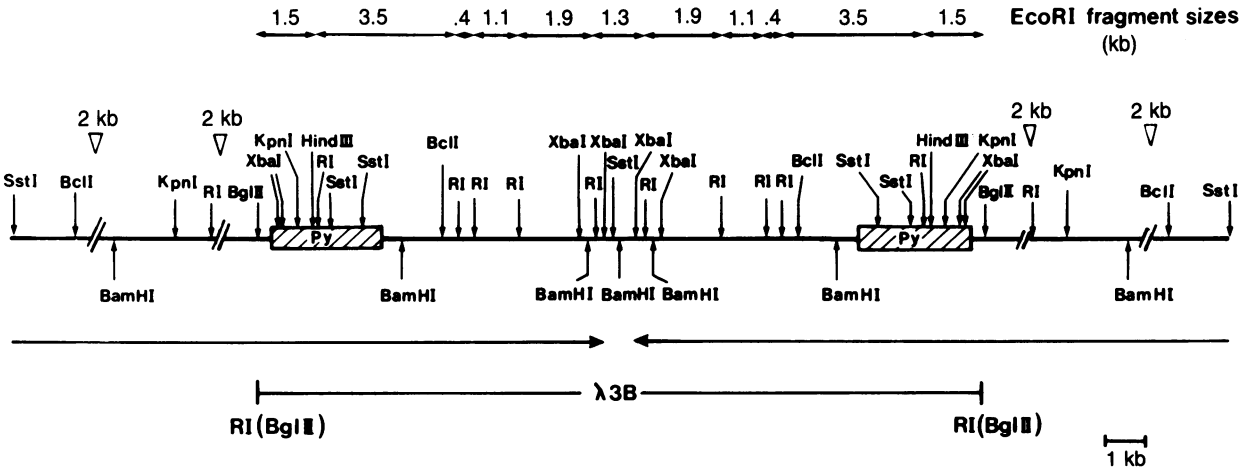

FIG. 4. The restriction map of 3B genomic DNA containing the Py sequences. The hatched boxes indicate the 2.8 kb of Py sequence present in 3B DNA. Arrows below the map indicate the inverted duplicated sequences, which are at least 22 kb long. The 18-kb insert of 3B genomic DNA in λ3B is indicated. Arrowheads indicate 2-kb interruptions in the scale of the map. RI, *Eco*RI site.

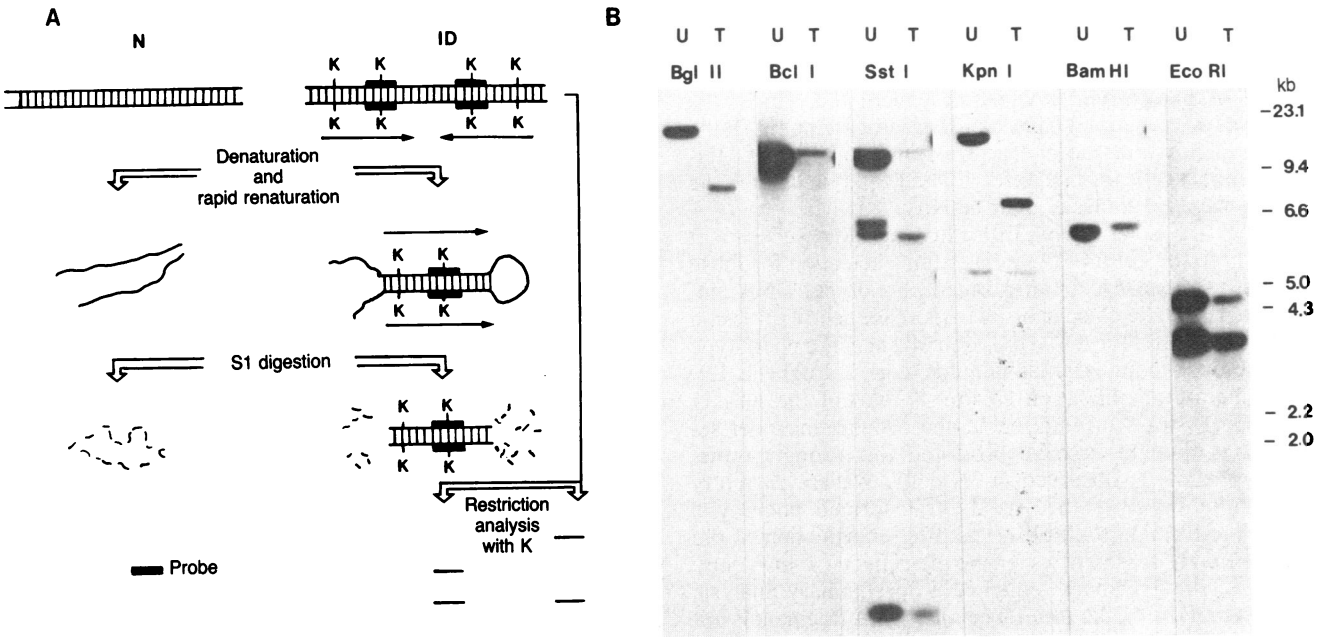

FIG. 5. *(A)* A diagrammatic representation of cleavage with a restriction enzyme (K) of noninverted DNA (N) and inverted duplicated DNA (ID) after denaturation and rapid renaturation followed by nuclease S1 digestion as described in the text. *(B)* Untreated (U) or treated (T) (as described in *A*) 3B genomic DNA was cleaved with the restriction enzymes indicated, fractionated, and analyzed as described in the legend to Fig. 1*B*.

action of nuclease S1. However, unrenatured DNA as well as single-stranded regions (noninverted DNA) of the molecules that contain the intrastrand duplex will be hydrolyzed by S1. A restriction analysis can be performed on the S1-resistant duplex DNA and compared to untreated genomic double-stranded DNA. Restriction fragments that lie between cleavage sites within the inverted duplicated sequences will be the same size whether generated from S1-treated or untreated genomic DNA. Restriction fragments that lie between a cleavage site within the inverted sequences and a site within the noninverted sequences will be shorter when generated from the S1-treated DNA than when produced from the untreated genomic DNA (Fig. 5*A*). These shorter fragments will extend from the restriction site within the inverted duplication to the end of the inverted duplicated sequences—this latter end being formed as a result of the S1 hydrolysis (see Fig. 5*A*). Thus, this technique will detect inverted duplications and measure the extent of the inverted duplication relative to a restriction enzyme cleavage site (Fig. 5*A*). A Southern blot analysis of 3B genomic DNA before and after treatment with nuclease S1 is presented in Fig. 5*B*. The Py-specific fragments generated from cleavage sites within the inverted duplication (see Fig. 4) by *Eco*RI (3.5 and 4.5 kb), *Bam*HI (6.0 kb), *Sst* I (12 and 0.8 kb), *Kpn* I (5 kb), and *Bcl* I (13 kb) are not altered by S1 treatment (Fig. 5*B*). On the other hand, those Py-sequence-containing fragments that extend from within the inverted duplicated sequences into the noninverted sequences (see Fig. 4) are altered by S1 treatment. Thus the 14-kb *Kpn* I fragment is reduced to 7 kb, the 18-kb *Bgl* II fragment to 8.5 kb, and the 6.3- and 7.2-kb *Sst* I fragments to 6.2 kb (Fig. 5*B*). These results confirmed that the structure of the inverted duplication was as described for the λ3B insert. In addition, the Py DNA-containing genomic 15-kb *Sst* I fragment remained unaltered after nuclease S1 treatment, indicating that the inverted duplication extends for at least 13.5 kb outside the sequences cloned in λ3B. Therefore, the size of the inverted duplication is at least 22 kb (Fig. 4).

Enhancing Activity of 3B Cellular DNA Sequences. Various fragments of 3B cellular DNA were tested for their ability to reactivate the transforming activity of the 4.1-kb Py fragment (Fig. 1) after transfection of Rat-1 cells. A plasmid containing the Py enhancer 5′ to the 4.1-kb fragment induces large transformed colonies (about one colony per ng of DNA) 2 weeks after transfection. The cellular sequences present in the 3.5- or 1.5-kb *Eco*RI fragments (see Fig. 4) were inactive in this assay. Twenty to fifty percent of the transforming activity was restored by the 4.7-kb *Bam*HI fragment which contains the 1.1- and 0.4-kb *Eco*RI fragments and part of the 1.9- and 3.5-kb *Eco*RI fragments (see Fig. 4). This activity was independent of orientation when the fragment was placed 5′ to the Py sequences. However, the colonies were, in general, much smaller and took about a week longer to appear. This result indicated that the enhancer from 3B cellular DNA was only weakly expressed in Rat-1 cells.

Gene Amplification in the Transformed Cell Line 3B. It was noted that the intensity of the signal from the Py sequences of 3B cellular genomic DNA was very strong (Fig. 1*B*) and that a higher-than-expected number of positive λ clones were isolated. This suggested that these sequences were amplified in 3B DNA. The degree of amplification was determined by comparing the intensity of the signal from restriction fragments containing only Py sequences in 3B cellular DNA with the DNA from the well-characterized Py-transformed rat cell line 7AXT, which is known to contain a single copy of Py sequences (4). When the intensities of the Py fragments (*Sst* I, *Pvu* II, and *Msp* I) common to the viral inserts in both cellular DNAs were normalized to the amount of DNA loaded onto the gel, it was found that the 3B sequences were amplified 20- to 40-fold in relation to those in the single-copy 7AXT DNA (data not shown).

DISCUSSION

Using the expression-selection procedure (2), we have isolated DNA sequences with enhancer-like activity from the 3B cell line. The Py and cellular enhancer sequences are contained within a large inverted duplication that is amplified 20- to 40-fold in 3B cellular DNA. The cellular sequences with enhancer-like activity are located in a 4.7-kb *Bam*HI fragment about 500 base pairs upstream from the Py sequences in 3B DNA. These sequences have been shown to weakly

activate Py gene expression in an orientation-independent manner in rat cells. We do not know whether these enhancer sequences are intrinsically weak or are reduced in activity because of the cell type in which they were assayed. As the sequences were derived from mouse teratocarcinoma DNA, they may function more efficiently in mouse cells, teratocarcinoma cells, or in a differentiated cell type than in the rat cells we used to assay their activity.

During the isolation of 3B cells and their subsequent growth in tissue culture, there has been a selection for those cells with the strongest transformed phenotype. DNA amplification can take place in order to increase gene expression to a level required for the expression of a specific phenotype or in response to an external stimulus (e.g., a drug) (11, 12). The moderate amplification (20- to 40-fold) of the weakly expressed Py-transforming region and cellular enhancer sequences is thought to have produced the strongly transformed phenotype observed in 3B cells. We are currently investigating whether the weakly transformed phenotype (small, late-forming foci) observed after transfection of the 4.7-kb *Bam*HI fragment (see above) reflects a low copy number of this enhancer sequence. A preliminary analysis indicates that the 4.7-kb *Bam*HI cellular DNA fragment is not amplified in F9 teratocarcinoma cellular DNA. It has been shown that integrated Py sequences can become amplified in the presence of a functional Py origin of DNA replication and Py large T antigen (DNA replication protein) (13). The Py insert in 3B DNA has lost approximately 120 base pairs (40 amino acids) of the large T-antigen coding region and is inactive for DNA replication (unpublished results). Thus, we think that the observed amplification is not Py-induced.

As the Py vector was ligated to a restriction digest of total cellular DNA fragments, most if not all of the Py sequences were joined to different cellular sequences. The Py and cellular DNA sequences detected in the inverted duplication are, therefore, most likely to have arisen from a single species of the transfected DNA and not as a result of recombination between two identical hybrid DNA molecules. We have not determined whether the inverted duplication arose before, during, or after integration into the cell chromosome.

The choice of the 18-kb Py-sequence-containing *Bgl* II fragment for molecular cloning was fortunate for the detection of the inverted duplication containing the Py and enhancer sequences, as this fragment spans one set of ends of the inverted duplication. The cloning of other Py-sequence-containing restriction fragments, which lie entirely within the inverted duplication (e.g., *Bam*HI, *Eco*RI, *Bcl* I; see Fig. 4), would not have allowed its detection. Upon denaturation and rapid renaturation, inverted duplicated sequences form intrastrand duplexes (Fig. 5*A*). Nuclease S1 digestion of DNA treated in this fashion results in the hydrolysis of single-stranded DNA. This procedure has been used with 3B genomic DNA (Fig. 5*B*) to confirm the position of one set of ends of the inverted duplication, as mapped by restriction analysis of the cloned 3B insert, and also to show that the inverted duplication extends for at least 22 kb within 3B genomic DNA. It is interesting to note that this large inverted duplication seems to be quite stable in 3B chromosomal DNA, as no differences have been detected on continued passage of 3B cells in culture for 6 months.

In the sequences studied here, both an amplification event and an inverted duplication are found. The question arises as to whether inverted duplications can occur as part of the DNA amplification mechanism. By definition, the formation of an inverted duplication is a form of amplification, as one copy of a sequence is doubled. In the 3B cells, there are 20–40 copies of the inverted duplicated sequences; however, this may reflect later amplification events, and the primary event may have involved the formation of the inverted duplication. Preliminary analysis of the occurrence of inverted duplications in different gene amplification systems suggests that inverted duplications can be associated with the amplification of other genes. If the formation of inverted duplications occurs during gene amplification, such sequences can be enriched by denaturation, rapid renaturation, and nuclease S1 treatment and be selectively cloned.

We are indebted to Dr. D. Ish-Horowicz for useful discussions and suggestions, C. Thacker for his excellent technical assistance, and Drs. G. Stark, P. Goodfellow, C. Dickson, and E. Giulotto for their comments during the preparation of this manuscript.

1. Gluzman, Y. & Shenk, T., eds. (1983) *Enhancers and Eukaryotic Gene Expression* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
2. Fried, M., Griffiths, M., Davies, B., Bjursell, G., LaMantia, G. & Lania, L. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 2117–2121.
3. Fried, M. & Williams, T. (1983) in *Enhancers and Eukaryotic Gene Expression,* eds. Gluzman, Y. & Shenk, T. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 118–120.
4. Lania, L., Hayday, A. & Fried, M. (1981) *J. Virol.* **39,** 422–431.
5. Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, D. K. & Efstratiadis, A. (1978) *Cell* **15,** 687–701.
6. Frischauf, A. M., Lehrach, H., Poustka, A. & Murray, N. (1983) *J. Mol. Biol.* **170,** 827–842.
7. Ish-Horowicz, D. & Pinchin, S. M. (1980) *J. Mol. Biol.* **142,** 231–245.
8. Burton, K. (1956) *Biochem. J.* **62,** 315–323.
9. Tooze, J., ed. (1981) *DNA Tumor Viruses* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
10. De Villiers, J. & Schaffner, W. (1981) *Nucleic Acids Res.* **9,** 6251–6264.
11. Stark, G. R. & Wahl, G. M. (1984) *Annu. Rev. Biochem.* **53,** 447–491.
12. Schimke, R. T. (1984) *Cell* **37,** 705–713.
13. Colantuoni, V., Dailey, L. & Basilico, C. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 3850–3854.