Efficient expression of *Escherichia coli* galactokinase gene in mammalian cells

(pBR322–simian virus 40 fusion plasmid/splice deletion/genetic complementation)

**Daniel Schümpelri**, Bruce H. Howard†, and Martin Rosenberg*

*Laboratory of Biochemistry and †Laboratory of Molecular Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205

Communicated by Igor B. David, September 10, 1981

**ABSTRACT** The *Escherichia coli* galactokinase gene (galK) was inserted into a modified early region transcription unit of simian virus 40 (SV40) contained on a bacterial plasmid. Introduction of this pSVK vector into monkey, mouse, and hamster cell lines by transfection resulted in efficient expression of the bacterial galK gene. This expression was shown to be dependent upon fusion of the galK gene to the early promoter of SV40 and did not appear to require SV40 splice signals. Moreover, expression in these cells could be obtained either transiently, 24–72 hr after transfection, or continuously, after stable transformation. In particular, pSVK-dependent galK expression was obtained in a hamster cell line genetically deficient in galactokinase activity. Expression of the bacterial enzyme was shown to complement the galactosemic defect of these cells, thereby allowing their selective survival and growth on galactose as the only carbon source. The ability to readily assay, select for, and potentially select against galK expression from pSVK and its derivatives should prove extremely useful in studying eukaryotic gene regulatory signals.

The universality of the genetic code implies that protein coding sequences should be faithfully translated in heterologous systems. We previously demonstrated that certain prokaryotic mRNAs could be translated at high efficiency in both wheatgerm and reticulocyte cell-free translation systems, provided their 5' ends were modified with a 7-methylguanosine cap structure (1). Apparently, except for the cap structure, the prokaryotic transcripts contained all the information necessary for their efficient expression within a eukaryotic cellular environment. These results suggested that prokaryotic genes that were properly introduced into eukaryotic cells might be efficiently expressed in vivo.

However, for a bacterial gene to be expressed in a eukaryotic cell requires that the gene be fitted with all the regulatory information that normally directs the complex steps of transcription and mRNA maturation of eukaryotic genes and that it be efficiently introduced into the cell nucleus. Recent efforts by several laboratories have focused on the development of recombinant vectors that allow new genetic material to be introduced into eukaryotic cells. Defective simian virus 40 (SV40) or other recombinant viruses that can be propagated in certain permissive cell lines or stably integrated into others have been used successfully (reviewed in ref. 2). Derived from these and of particular interest is a recently developed class of vectors that can be propagated in *Escherichia coli* and then can be used to introduce new genes into mammalian cells. These plasmid vectors have been used to obtain expression of inserted genes (including two bacterial genes) in the absence of any ongoing virus infection (3–5).

In this work, we have inserted the *E. coli* galactokinase gene (galK) into one of these plasmid vectors in such a way that transcription of the gene is controlled by the SV40 early promoter. When introduced into cells, the vector directs the efficient synthesis of *E. coli* galK enzyme. Moreover, we show that cells can be transformed to stably express galK and that this expression can be used to select for complementing cells that are genetically deficient in their own galactokinase gene function. The potential use of this vector system for studying eukaryotic gene regulation signals will be discussed.

**METHODS**

Plasmid Construction and DNA Preparation. The design and construction of various recombinant plasmids is described in detail in the text and in the legend to Fig. 1. Digestions with BAL 31 (a gift of H. Gray) were carried out as described in 200 mM NaCl buffer (6). DNA from appropriate time points (as judged by restriction endonuclease analysis) was purified by phenol extraction and repeated ethanol precipitation before it was used for ligation. HindIII linkers (Collaborative Research, Waltham, MA) were activated with phage T4 polynucleotide kinase (P-L Biochemicals, Milwaukee, WI) and ATP (0.2 mM final concentration) as described (7) and used at a linker-to-plasmid ratio of 50:1 (mol/mol). DNAs were ligated overnight at 14°C with T4 DNA ligase (P-L Biochemicals) and then used to transform appropriate bacterial hosts. The desired plasmid constructions were identified by a rapid DNA isolation and screening procedure and then amplified and purified by a cleared lysate technique (8) followed by digestion with DNase-free RNase A (0.5 µg/ml), repeated extraction with phenol/chloroform/isoamyl alcohol (25:24:1 vol/vol), and gel filtration on Sepharose 4B (Pharmacia).

DNA Sequence Analysis. DNA sequences were determined by the chemical methods of Maxam and Gilbert (9), or by partial snake venom phosphodiesterase digestion (10, 11).

Cell Culture. African green monkey CV-1 cells, mouse NIH 3T3 cells, and Chinese hamster V6 and R1610 cells [galactokinase-deficient, hgpt- (hpt 13); hgpt is the gene for hypoxanthine (guanine) phosphoribosyltransferase; a kind gift of J.-P. Thirion] were propagated in Dulbecco's modified minimal essential medium supplemented with 5% fetal bovine serum (KC Biological, Lenexa, KA). About 2.5 x 10⁶ cells were transfected with 10 µg (unless specified otherwise) of DNA according to Wigler et al. (12). Selections for expression of the *E. coli* hpt gene [the gene for xanthine (guanine) phosphoribosyltransferase] were carried out as described by Mulligan and Berg (13).

R1610 cells were selected for galactokinase expression by incubation from the third day after transfection in glucose-free modified minimal essential medium supplemented with 10% FCS.

Abbreviations: SV40, simian virus 40; t antigen, small tumor antigen.
dialyzed fetal bovine serum and 1 g of D(+) galactose per liter (14). Individual clones were isolated as described (15).

**RESULTS**

**Design and Construction of the pSVK Vector.** The procedures used to insert the *E. coli* galK gene into a vector designed to express protein coding sequences in eukaryotic cells are outlined in Fig. 1. The entire *E. coli* galK gene is carried on the plasmid vector pDS26 (see legend to Fig. 1) and is flanked at its 5' side by unique HindIII and Sma I sites and at its 3' side by unique Hpa I and BamHI sites. These sites were specifically engineered into the vector so that the galK gene could be readily obtained on a variety of DNA fragments and introduced by defined in vitro fusions into any genetic system. Because eukaryotic ribosomes are thought to utilize predominantly the cap-proximal AUG triplet to initiate translation (19, 20), we wanted to ensure that the authentic galK initiation codon would be the first AUG on the mRNA produced in mammalian cells. On pDS26, there are three ATG triplets between the HindIII site and the galK translation start codon (21). We therefore shortened the leader region preceding galK by first linearizing pDS26 with HindIII and subsequently digesting the ends with the double strand-specific exonuclease BAL 31. HindIII linker molecules were then inserted into the various deletion sites. Individual shortened plasmids were characterized by restriction and DNA sequence analysis, and some were found to have deleted the three ATG triplets upstream of the galK gene. One of these plasmids with deletions had the HindIII linker positioned 35 base pairs upstream of the galK translation start site (pDSAK, Fig. 1).

The galK coding sequence, carried on pDSAK, was then introduced into pSV2, a vector that carries certain regulatory signals thought important for gene expression in a eukaryotic background (3, 5). These elements are: (i) the SV40 early promoter region, on a fragment carrying the whole viral replication origin, to which the coding sequence can be attached so that its first ATG triplet will also be the first AUG on the transcript; (ii) a segment carrying the SV40 t antigen intervening sequence, which is positioned distal to the inserted coding sequence; and (iii) the SV40 early transcript polyadenylation site located downstream from the t antigen splice region. The vector also contains a segment of pBR322 with the bacterial replication origin and the β-lactamase gene conferring ampicillin resistance, thus allowing plasmid constructions and amplification of DNA to be carried out in *E. coli*. pSV2, originally constructed with a rabbit β-globin cDNA insert, was cut with HindIII and Bgl II to remove the β-globin segment and thereby produce its vector form. We then inserted into this vector the HindIII/BamHI fragment from pDSAK that carries the *E. coli* galK gene in the proper orientation to form the prototype pSVK plasmid. An important consequence of this cloning scheme is that the arrangement of restriction sites on the pSVK construction allows any of its functional elements to be easily removed or substituted for (Fig. 1). In addition, the vector has unique BamHI, EcoRI, and Pvu II sites into which additional sequences could be inserted without disturbing the putative galK transcription unit.

**pSVK Directs *E. coli* galK Expression.** We tested if pSVK DNA, brought into mammalian cells by transfection, was capable of expressing the bacterial galK gene to produce an enzymatically active product. For these initial experiments, we used the CV-1 African green monkey cell line, whose ability to support transcription from the SV40 early promoter has been well documented (for review see ref. 22). Although CV-1 cells are permissive for SV40 growth, the origin region present on pSVK will not support replication because no SV40 large tumor (T) antigen is available.

CV-1 cells were transfected with pSVK DNA and grown for two days. The *E. coli* galK enzyme can be readily separated from the monkey enzyme by using a starch gel electrophoresis/assay system (ref. 17, Fig. 2A, lanes 6 and 1, respectively). This endogenous galactokinase activity provides a useful internal marker for the number of cells analyzed. Extracts from transfected cells were thus analyzed and found to contain a new galactokinase activity (Fig. 2A, lane 4). This enzymatic activity was dependent entirely on the presence of the pSVK vector and was lost when the cells were maintained in culture for another week. Identical transfection with other plasmid or phage vectors carrying the galK gene but without eukaryotic regulatory sequences did not result in detectable synthesis of this enzyme (Fig. 2A, lanes 2 and 3). More importantly, a pSVK derivative from which the entire SV40 fragment upstream of the galK gene carrying the SV40 early promoter had been deleted also did not produce any detectable *E. coli* galK enzyme (Fig. 2B, lane 9).
Thus, we have obtained transient expression of the bacterial galK gene in monkey cells, and this expression seems to depend on the fusion of the gene to a eukaryotic promoter signal. By comparing the relative intensities of galactokinase bands produced in pSVK-transfected cells with the ones obtained from assayed bacterial markers we estimate that about 0.1–1.0 unit of enzyme per $10^6$ cells is produced (1 unit = 1 nmol/min). If the specific activity of the E. coli galK enzyme produced in monkey cells is similar to the one of the purified bacterial enzyme (23), this activity, averaged over all the cells, corresponds to about $2 \times 10^4$ to $2 \times 10^5$ protein molecules per cell. Presumably, only some fraction of the transfected cells actually expressed the E. coli galK enzyme, and expression in individual cells was therefore even more efficient.

To test whether pSVK would also direct the synthesis of E. coli galK enzyme in cells that are not permissive for SV40, we used NIH 3T3 mouse fibroblasts. In this case, we cotransfected the cells with pSVK and another pSV2-like vector carrying a dominant selectable marker, the E. coli zgtP gene (ref. 3; plasmid pSV2-gpt). Beginning 2 days after transfection, the cells were selected for the expression of the zgtP gene (13), pooled, and then analyzed for their ability to express both the zgtP and the cotransfected galK gene. As expected, the resistant population contained E. coli xanthine (guanine) phosphorysityltransferase activity (data not shown). In addition, these cells also contained the bacterial galactokinase activity (Fig. 2C, lane 11). This activity was more than 10-fold higher than when assayed 2 days after transfection, prior to the xanthine (guanine) phosphorysityltransferase selection. Thus, some fraction of the cells that took up, stably integrated, and continuously expressed pSV2-gpt did so also for the second, unlinked and unselected, vector, pSVK. This is reminiscent of the results of Wigler et al. (12), who demonstrated cointegration of unlinked selectable genes along with the herpes simplex virus thymidine kinase gene. Detection of pSVK-directed galK gene expression in the cotransfected NIH 3T3 cells was possible, despite the nearly identical electrophoretic mobilities of the E. coli and mouse galactokinase enzymes (Fig. 2C), because the bacterial enzyme produced in mammalian cells always moved somewhat faster in the gels than the one made in E. coli (compare lanes 4 and 6 in Fig. 2A). The reason for this electrophoretic difference is not known. However, it is likely to result from some posttranslational modification of the protein. We have observed the identical phenomenon when the bacterial zgtP gene product, expressed in these same cells from the pSV2-gpt vector, was assayed by starch gel electrophoresis (data not shown). This electrophoretic difference was not detected previously by assay on polyacrylamide gels (3).

Thus, both monkey and mouse cells and, as shown below, also hamster cells, can be induced by pSVK transfection to express the bacterial galK enzyme. This expression appears to depend on the gene being fused to a eukaryotic promoter signal. Moreover, expression can be obtained either transiently within a few days after transfection or stably by cotransformation with a dominant selectable marker.

**t Antigen Splice Is Not Required for galK Expression.** The SV40 t antigen splice region was originally built into the pSV2 vector because rabbit $\beta$-globin cDNA inserted into a similar transcription unit without splicing signals did not lead to the production of cytoplasmic $\beta$-globin mRNA (5). We investigated the importance of the SV40 t antigen splicing signals for E. coli galK gene expression. The pSVK plasmid was cut with the re-
striction enzyme Hpa I (Fig. 1) and religated to form a derivative that was missing the entire region containing the t antigen splicing signals (pSVK 102). CV-1 monkey cells (Fig. 3) or V6 and R1610 hamster cells (data not shown) transfected with pSVK 102 produced an E. coli galactokinase enzyme identical to that obtained from pSVK. Comparison of the levels of expression obtained from pSVK and pSVK 102 in the same experiments indicated that deletion of the t antigen splice region resulted in somewhat higher, rather than lower, production of the bacterial enzyme. Clearly, splicing at the SV40 t antigen-specific signals was not required for galK expression. However, we do not yet know whether the pSVK 102-directed galK gene expression occurs off an unspliced mRNA. It is possible that other sequences present on the galK transcription unit can substitute as splice signals.

Complementation of Genetic Galactokinase Deficiency in Cell Culture. Cells of most organisms from bacteria and protozoans to humans have the ability to metabolize galactose (for review see ref. 24). Galactokinase catalyzes the first step of this metabolic pathway, the phosphorylation of galactose to galactose 1-phosphate. In humans, genetic defects of galactokinase or uridylyltransferase, another enzyme in this pathway, are the causes of the two forms of galactosemia (25, 26). In addition, galactokinase defects have been well characterized in bacteria, certain yeasts, and a Chinese hamster cell line. We have used galactokinase-deficient Chinese hamster cell mutants (14) to study the expression of the bacterial galK gene.

Wild-type Chinese hamster cells (V6 cell line) contained a galactokinase activity that was completely absent from extracts of the galactokinase-deficient mutant R1610 cell line (Fig. 2D, lanes 12 and 14). When transfected with pSVK DNA, both cell lines produced a new galactokinase band corresponding to E. coli galK enzyme (Fig. 2D, lanes 13 and 15). In R1610 cells, this pSVK-dependent band was the only detectable galactokinase activity. Thus we have obtained transient expression of the bacterial galK gene in hamster cells and in particular in a cell line that is deficient in any endogenous galactokinase activity.

We then tested whether the bacterial galactokinase activity produced in response to pSVK transfection was sufficient to overcome the galactokinase defect of R1610 cells and thereby allow them to grow in the presence of galactose as the only carbon source. Cells were again transfected with pSVK DNA and after 2 days replated in glucose-free medium supplemented with galactose and dialyzed fetal bovine serum. Control cells that had not received pSVK DNA did not survive in this med-

![Fig. 3. Galactokinase activity from a pSVK derivative missing the SV40 t antigen splicing signals. The fluorograph of a starch gel electrophoretic galactokinase analysis of monkey CV-1 cell extracts is shown. Lane 1, cells transfected with pSVK DNA; lane 2, sheared calf thymus DNA; lane 3, DNA of pSVK 102, a pSVK derivative from which the Hpa I fragment carrying the SV40 t antigen intervening sequence was deleted (see text). The positions of monkey and E. coli galactokinases are indicated.](image)

In marked contrast, among pSVK-transfected cells surviving colonies were detected at a frequency of 5 × 10⁻⁴. Several of the resulting cell clones were isolated and found to produce E. coli galK enzyme. Thus, pSVK-directed galK gene expression was sufficient to complement the genetic galactokinase defect of R1610 cells and selectively enabled pSVK-transformed cells to grow in a galactose medium.

**DISCUSSION**

We have constructed the hybrid fusion vector pSVK and demonstrated that it will introduce the E. coli galK gene into a variety of eukaryotic cells and allow its efficient expression. Along with other recent demonstrations of bacterial gene expression in mammalian cells (3, 4), these results clearly demonstrate that certain bacterial genes contain all the necessary information for efficient translation in a eukaryotic cell.

**Regulatory Elements Controlling galK Expression.** An important feature of the pSVK vector is that each functional element of the galK transcription unit (e.g., promoter, gene, splice signals, polyadenylation site) can be separately removed or substituted with other DNA fragments. Thus, the importance of the regulatory elements controlling galK expression can be readily investigated. For instance, we demonstrated that galK expression from the pSVK vector is dependent on fusion of the gene to a fragment of SV40 DNA carrying the early SV40 promoter. In addition, we have shown that galK expression does not require the t antigen splice region. However, the question whether pSVK 102-directed galK expression is mediated by an unspliced mRNA needs to be further investigated. We find the possibility of a substitute splice unlikely for two reasons: (i) the observed high level of expression would be surprising if expression depended on splicing and that event were occurring at sites that had not naturally evolved to carry out this function; (ii) on pSVK 102, the SV40 early polyadenylation signal is located only 130 base pairs downstream from the galK termination codon (unpublished data). Thus, the putative 5′ and 3′ untranslated sequences are both very short and moreover do not contain any sequences with significant homology to other splice boundaries (27). Moreover, any splicing that involved galK coding sequences would presumably have drastic effects on the size or function of the polypeptide. The pSVK vector also contains the early SV40 polyadenylation signal, but we do not as yet know whether this signal is required for expression of the bacterial gene.

The modular structure of the pSVK vector should also prove useful in studying sequences from other eukaryotic genes that can fulfill those functions that are required for galK gene expression. For instance, it is possible to remove the SV40 promoter region and substitute other eukaryotic DNA fragments that contain analogous regulatory information.

In addition, the vector can be used to study the effects on galK expression that result from varying the size and structure of the 5′ untranslated leader region in front of galK. For instance, in the BAL31 deletion experiment performed to remove additional ATG triplets 5′ to the galK coding sequence (see above and Fig. 1), we obtained a large number of deletion plasmids. These plasmids all contain a novel HindIII site that is separated from the galK coding sequence by various stretches of E. coli "leader" sequence. Derived from these, we have constructed a related set of pSVK derivatives that should help elucidate the role and importance of positional effects on the translation efficiency of a particular gene (i.e., galK) in mammalian cells.

**Gel Assay and Quantitation.** Another important element of the pSVK vector system is the ability to readily monitor expression of the bacterial galK gene in various cell types despite the
presence of an endogenous galactokinase activity. Using this simple and discriminative assay, it is possible to analyze in a variety of cell types the precise in vitro fusions that recombine various eukaryotic regulatory signals with the galK gene present on pSVK. The fact that transient galK expression in CV-1 cells was roughly proportional to the amount of pSVK DNA used for transfection (Fig. 2A, lanes 4 and 5) suggests that the system might be used to quantitatively assess the function of various gene control signals by monitoring different levels of galK expression. However, precise quantitation of transient gene expression after DNA-mediated transfection is potentially affected by a large number of experimental variables. Particularly problematic is the differential gene expression that will result from variations in gene copy number. Recent experiments (unpublished) indicate that these problems can be overcome by introducing into the same vector a second assayable gene (i.e., E. coli xgpt) in a similarly constructed transcription unit and using it as an internal standard. Thus, variations in galK expression resulting from alterations in its gene control sequences can be accurately measured.

Genetic Complementation and galK Selections. We have demonstrated pSVK-dependent expression of the bacterial galK gene in a hamster cell line that is genetically deficient in its own galactokinase activity. Moreover, expression of the bacterial enzyme complements the galactosemic defect of these cells, enabling them to survive and grow in a medium containing galactose as the only carbon source. Thus, the galactokinase-deficient hamster cell line allows positive selection for those cells that have been stably transformed by pSVK (or potentially by other pSVK derivatives) and continually express the bacterial galK enzyme. Stably transformed cell clones can be studied with respect to the state of the introduced DNA and the precise nature of galK-specific transcripts.

In addition to this positive selection for galK expression, the system is also amenable to a negative selection (14). 2-Deoxy-D-galactose, an analogue that, when metabolized, blocks sugar chain elongation of glycoproteins (25), could be used to select among the pSVK-transformed hamster cells for those cells that no longer express the galK gene. Such positive and negative selections may prove extremely useful in obtaining eukaryotic regulatory site mutations.

We express our gratitude to our colleagues Keith McKennedy, Raji Padmanabhan, Gail Taff, Charles Mock, O. Wesley McBride, and Jane Peterson, to Jean-Paul Thirion for sending cell lines, and to Richard Mulligan and Paul Berg for making available pSV2-gpt and for communicating results prior to publication. D.S. was supported by Fellowship no. 83.746.0.79 from the Swiss National Science Foundation.