BK Viral Enhancer Element and a Human Cellular Homolog

Nadia Rosenthal, Michel Kress, Peter Gruss, George Khoury

Summary. Comparison of two closely related primate papovaviruses, simian virus 40 (SV40) and human BK virus (BKV), reveals that the only region of extensive divergence, the tandem sequences adjacent to the origins of DNA replication, is responsible in SV40 for enhancing early gene expression. This study demonstrates a similar enhancer function for the analogous repeated region in BKV. The dissimilarity in sequence of the BKV and SV40 enhancer elements suggests that they may have been acquired since SV40 and BKV diverged. A locus cloned from the human genome homologous to the BKV tandem repeats has been shown to function as low level enhancer element in mammalian cells. These data support the hypothesis that viral enhancer sequences may be evolutionarily related to host cell sequences.

Identification of elements that control the initiation of transcription is a crucial step in understanding the regulation of eukaryotic gene expression. Approaches to the problem of transcriptional control with animal viruses as a model system have recently uncovered transcriptional regulatory elements, called enhancers. The prototype enhancer was originally identified in the DNA virus SV40 as a set of *cis*-essential sequences required for efficient expression of early viral genes (*1*). These elements increase the level of transcription of an adjacent gene from its promoter in a fashion that is relatively independent of position and orientation (*2*). Subsequent studies uncovered similar elements in a number of DNA viruses including polyoma virus (*3*), papilloma virus (*4*), and adenovirus (*5, 6*). Retroviruses also contain enhancer elements in their long terminal repeats (*7–10*).

Although there is no extended sequence homology among these elements, a 7- to 10-bp (base pair) consensus or "core" sequence has been identified in a number of enhancers. Deletions or mutations in the core sequence of SV40 obliterate enhancer activity (*2, 11*), suggesting that it is an important component of the enhancer. Apart from the core, viral enhancer elements show very little sequence similarity even in closely related viruses.

The limited coding capacity of certain viruses makes them dependent on host cell regulatory molecules that are normally directed toward the expression of cellular genes. In some cases these cellular factors, as yet undefined, appear to mediate the activity of viral enhancers in a host-specific way. For example, when compared to the 72-bp repeats of SV40, the enhancer associated with the long terminal repeat (LTR) of Moloney murine sarcoma virus is more active in murine cells than it is in primate cells (*8, 10*). The host cell preference of several viral enhancers (*8, 10, 12, 13*) suggests that similar transcriptional regulatory elements associated with host genes may interact with a limited set of cellular factors, contributing to the cell specificity of gene expression. The recent description of tissue-specific enhancer elements in mouse immunoglobulin gene introns (*14–16*) strongly supports this supposition, and suggests that enhancers may be a general feature of cellular gene regulation.

We have used the human papovavirus BKV as a viral model of eukaryotic transcriptional control. This virus was first isolated from the urine of an immunosuppressed patient (*17*) and is suspected of causing an inapparent childhood infection (*18*). The virus grows optimally in human tissue culture and undergoes an abortive cycle in nonprimate cells, transforming rodent cells in vitro and inducing tumors in hamsters (*19*). Papovaviruses BKV (BK virus) and SV40 are remarkably similar, sharing more than 80 percent of their nucleotide sequences (*20, 21*). They differ substantially in only a small stretch of tandem repeated nucleotides to the "late" side of the replication origin (*19–22*). In SV40, these sequences include the 72-bp enhancer element. The dissimilarity of the SV40 and BKV repeated sequences suggests that they may not have evolved from a common viral ancestor (*20*), but may have derived from the host cell genome. Our study was undertaken to assign an enhancer function to the BKV repeats, to demonstrate the presence of homologous human genomic sequences, and to assess their ability to function as an enhancer element.

Construction of BKV-CAT Expression Plasmids

Comparison of the BKV and SV40 genomes reveals a region of extensive sequence heterology situated to the late side of the replication origins (*19–22*). For both viruses, this region begins with an AT-rich (A, adenine; T, thymine) stretch that includes the Goldberg-Hogness or TATA box for the early (T antigen) transcriptional units. In SV40 the TATA box is preceded by three GC-rich 21-bp repeats (G, guanine; C, cytosine), each containing two copies of the sequence CCGCCC. The role of these repeats is not clear, but they may be involved in RNA polymerase binding and viral DNA replication (*2, 23*). Adjacent to the 21-bp repeats are two tandem 72-bp repeats, which have been characterized as enhancer elements for early SV40 gene expression (*1, 2*) (Fig. 1).

The Dunlop strain of BKV contains a 68-bp triplication (the central copy of which is missing 18 nucleotides) preceding the AT-rich region (*20*). This set of nucleotides bears some similarity to both the 21-bp and the 72-bp SV40 tandem repeats. In each of the BKV units, there is one copy of a GC-rich hexanucleotide (CCTCCC) analogous to those in the SV40 21-bp repeats. In addition, each of the BKV units contains the sequence GGTCATGGTTTG, similar to the proposed SV40 enhancer core sequence (GGTGTGGAAAG).

To test the triplicated BKV sequence for enhancer activity, we isolated a 216-bp fragment, containing all three repeats without a TATA box or replication origin, by Hae III restriction enzyme cleavage of the BKV genome (Fig. 1). We used the chloramphenicol acetyltransferase (CAT) assay to evaluate enhancer activity (*24*). In this assay, a bacterial gene encoding the enzyme CAT is linked

The authors are staff members of the Laboratory of Molecular Virology, National Cancer Institute, Building 41, Suite 200, Bethesda, Maryland 20205.

to several eukaryotic regulatory signals from the early transcription unit of SV40. The resulting plasmid ($pA_{10}cat_2$) includes a portion of the "early" SV40 promoter (the 21-bp repeats and TATA box) from which the enhancer sequence has been deleted, an intron, and a polyadenylation signal (*8, 24*). The addition of enhancer elements to $pA_{10}cat_2$ results in efficient CAT gene expression when introduced into eukaryotic cells. The amount of gene product (CAT) is measured by its in vitro conversion of chloramphenicol to an acetylated form. Since eukaryotic cells do not contain a gene for the CAT enzyme, there is no background. The assay therefore provides a sensitive, quantitative, and reproducible estimate of gene activity induced by added enhancer sequences.

Since the prototypic enhancer sequences appear to function in a position and orientation independent fashion (*2*), the 216-bp BKV fragment described above was placed in both sense (s) and antisense (a) orientations (relative to its position in the BKV genome) at the 5′ end of the gene (pBKs5′cat and pBKa5′cat, respectively), and in the s orientation at the 3′ end of the gene (pBKs3′cat). In the latter construction, the BKV sequences are 3 kb upstream or 2 kb downstream from the SV40 promoter elements in the circular molecule (Fig. 1).

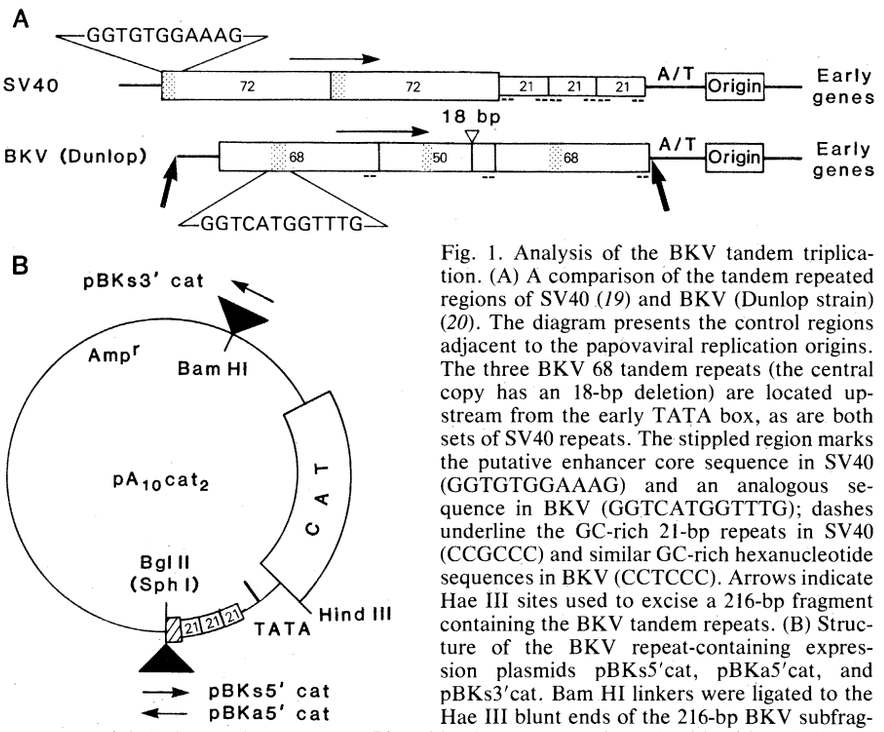


Fig. 1. Analysis of the BKV tandem triplication. (A) A comparison of the tandem repeated regions of SV40 (*19*) and BKV (Dunlop strain) (*20*). The diagram presents the control regions adjacent to the papovaviral replication origins. The three BKV 68 tandem repeats (the central copy has an 18-bp deletion) are located upstream from the early TATA box, as are both sets of SV40 repeats. The stippled region marks the putative enhancer core sequence in SV40 (GGTGTGGAAAG) and an analogous sequence in BKV (GGTCATGGTTTG); dashes underline the GC-rich 21-bp repeats in SV40 (CCGCCC) and similar GC-rich hexanucleotide sequences in BKV (CCTCCC). Arrows indicate Hae III sites used to excise a 216-bp fragment containing the BKV tandem repeats. (B) Structure of the BKV repeat-containing expression plasmids pBKs5′cat, pBKa5′cat, and pBKs3′cat. Bam HI linkers were ligated to the Hae III blunt ends of the 216-bp BKV subfragment containing the tandem repeats. Plasmid $pA_{10}cat_2$ was cleaved with either Bgl II or Bam HI generating cohesive termini. Linear vectors were ligated with the BKV 216-bp fragment (*46*) and recombinant plasmids were transfected into *Escherichia coli* HB101 by the calcium phosphate method (*25*). Alternative orientations of the repeats (with respect to their position in the BKV early transcription unit) were selected in 5′ position, and the sense orientation was in the 3′ position. pBKs5′cat, BKV repeats in the sense orientation relative to the CAT gene, inserted at the Bgl II site (5′ position). pBKa5′cat, BKV repeats in the antisense orientation at the Bgl II site (5′ position). pBKs3′cat, BKV repeats in the sense orientation at the Bam HI site (3′ position).

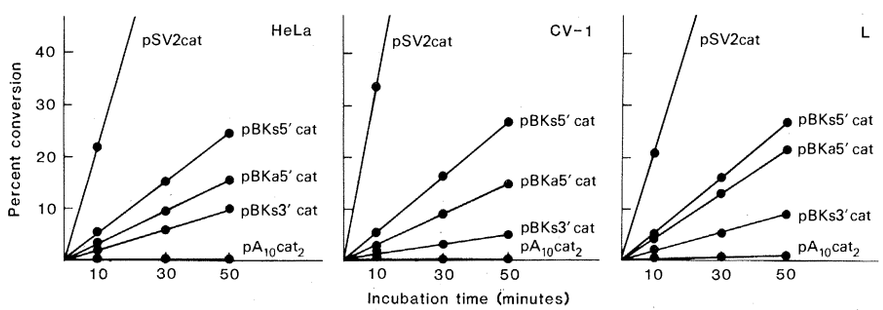


Fig. 2. Activation of the CAT gene by the repeat sequences from BKV and SV40. Equivalent amounts (25 μg) of the CAT plasmids containing the BKV repeats (pBKs5′cat, pBKa5′cat, and pBKs3′cat), the SV40 repeats (pSV2cat), or no enhancer element ($pA_{10}cat_2$) were introduced into human HeLa cells, monkey CV-1 cells, or mouse L cells (10^6 cells in 100-mm dishes) by calcium precipitation (*8, 24*). Cultures were harvested 48 hours after transfection, and protein extracts were prepared and analyzed for CAT activity (*24*). Conversion of chloramphenicol (percent) to an acetylated form was calculated by thin-layer chromatography and radioactive scintillation counting of the spots cut from the plate. Values are an average of three separate experiments, each performed on duplicate plates.

Enhancer Activity of the BKV Repeats

Each plasmid containing the BKV triplication was separately transfected into semiconfluent HeLa (human), CV-1 (monkey), or L (mouse) cells by the calcium phosphate method (*25*). After 48 hours, the cells were harvested, and total protein was extracted and assayed for CAT activity (*8, 24*). Enzyme assays for each set of transfected cells are shown as time course in Fig. 2. The effect of the BKV repeats on CAT expression is compared to the SV40 72-bp repeats (pSV2cat) and to the plasmid with no enhancer ($pA_{10}cat_2$). Kinetic analyses of the CAT activity of the BKV constructs normalized against pSV2cat levels in each cell type are presented in Table 1.

These data show that the BKV fragment enhances CAT expression from the heterologous SV40 promoter in both orientations and at a distal 3′ location in all three cell types. The construct containing the SV40 enhancer, pSV2cat, induces CAT expression in HeLa and L cells to levels four times as high and in CV-1 cells six times as high as pBKs5′cat, the analogous construct containing the BKV repeats. This difference may reflect properties of the individual constructions; for example, in the BKV plasmids the 68-bp triplication, each of which contains its own GC-rich sequence, is separated from the TATA box by the SV40 GC-rich 21-bp repeats. Alternatively, the relative strengths of the BKV and SV40 enhancers measured in this assay may reflect a true difference of enhancer activity in vivo. This difference is consistent with the biological properties of the two viruses, since it has been shown that BKV grows more slowly in primate cells than does SV40 (*26*). The induction of CAT activity by plasmids with the BKV sequences in the 5′ location were reproducibly higher than induction by the plasmid in which they were positioned 3′ to the CAT gene, although the 3′ antisense orientation was not tested. We have obtained similar

results in the analysis of enhancer sequences from murine sarcoma virus (*9*). Whether the differences in activity reflect the increased distance between enhancer and promoter elements, cryptic plasmid promoters which compete for activity (*27*), or a contribution from promoter sequences within the BK repeats, remains unclear.

In summary, the 68-bp BKV triplication functions as an enhancer element for gene expression in a number of cell types without a pronounced host cell preference. Like other enhancer elements, it has the properties of position and orientation independence although the level of gene enhancement appears to depend on the location of the BKV sequence in any particular construct.

Homologous Sequences in the Human Genome

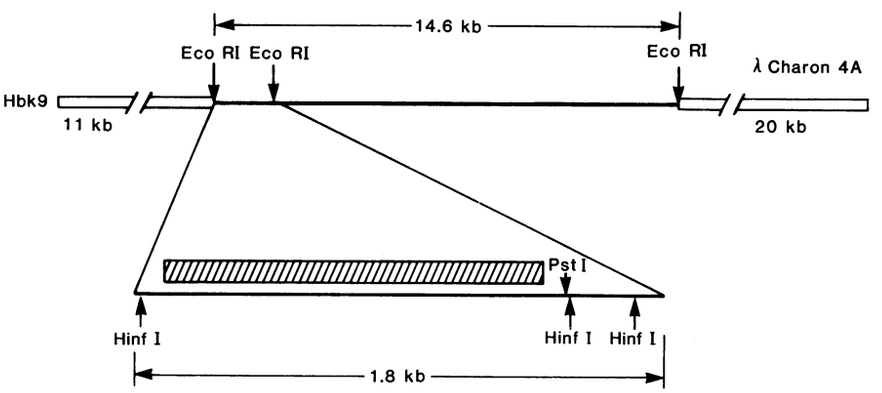
To investigate the possibility that viral enhancers may have cellular homologs, we screened a human genomic λ library (*28*) for sequences related to the BKV tandem triplication with the 216-bp BKV Hae III fragment as a probe. One genomic clone (Hbk9) that hybridized strongly with the probe after repeated plaque purifications was selected for further analysis. Restriction enzyme mapping and Southern blot analysis (*29*) generated a map of the clone (Fig. 3). All of the BKV-homologous sequences are located within a 1.8-kb Eco RI fragment at one end of the 14.6-kb human sequence inserted in Hbk9. The possibility that Hbk9 contained sequences homologous to regions of BKV other than the enhancer segment was tested by hybridizing a ^{32}P-labeled nick-translated probe of the BKV genome minus the enhancer to a Southern blot of Eco RI–cleaved Hbk9 DNA. The absence of a signal indicates that the human BKV-homologous segments are not associated with an integrated virus (data not presented).

To determine whether the BKV-homologous sequences in clone Hbk9 were represented elsewhere in the human genome, we used the 1.45-kb Eco RI–Pst I fragment from this clone to probe a Southern blot of human genomic DNA cleaved with Eco RI or Hinf I. The probe hybridizes to only one band in each digest (Fig. 4), suggesting that the region homologous to the BKV enhancer represented in clone Hbk9 is present at a single chromosomal site. The size of the fragment (2.7 kb) in the Eco RI digests (Fig. 4, lanes 1 and 3) is larger than 1.8 kb, reflecting the fact that construction of the library generated artificial Eco RI sites at the ends of the clones. Therefore, the 2.7-kb fragment observed in the genomic DNA's is only partially represented in Hbk9. The single 1.45-kb fragment detected by hybridization of the enhancer probe to the Hinf I digest of human genomic DNA (lanes 2 and 4) is entirely included in clone Hbk9. No additional bands were detected in identical blots hybridized under less stringent conditions (data not presented). These results suggest that there are no other significant stretches of sequences homologous to the BKV repeats in the human genome either adjacent to those included in Hbk9, or at other locations.

A similar analysis was performed on DNA from two different monkey cell sources (AGMK and CV-1) and a mouse cell line (NIH 3T3) cleaved with Eco RI or Hinf I and hybridized under the same conditions. The BKV-related sequence is present in the monkey genome (Fig. 4, lanes 5 to 8), but absent from the mouse genome (lanes 9 and 10). The size of the internal Hinf I fragment is slightly smaller in the monkey genome (1.3 kb; Fig. 4, lanes 6 and 8) than in the human genome (1.45 kb), perhaps reflecting a deletion or polymorphism in the number of repeat units. The 2.7-kb Eco RI fragment present in human DNA appears as a 7.0-kb band in monkey DNA (lanes 5 and 7), presumably due to the loss of an Eco RI site.

Tandem Repeats in the Human Sequence

The nucleotide sequence of the 1.8-kb Eco RI fragment was determined after subcloning into M13 (*30*) by the dideoxy primer extension method (*31*). A portion of the region homologous to the BKV



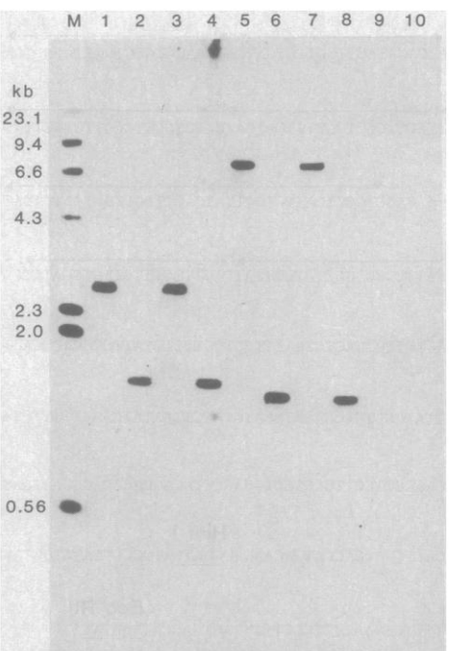


Fig. 3 (above). Map of clone Hbk9. A human genomic library, constructed by cloning a Hae III–Alu I partial digest of human DNA into λ vector Charon 4A by way of Eco RI linkers (*28*), was screened with the 216-bp BKV Hae III fragment (see Fig. 1) as a ^{32}P-labeled probe. Under stringent conditions [6× SSC (0.9*M* NaCl, 0.09*M* sodium citrate), 50 percent formamide at 37°C], 12 clones from a genomic library equivalent (800,000 clones) hybridized to the probe. After repeated plaque purifications only one clone, Hbk9, annealed strongly and was selected for further study. A detailed map was generated from restriction enzyme analysis and Southern blotting with the BKV repeat probe (data not shown). The entire sequence homologous to BKV (hatched bar) is contained within the 1.8-kb Eco RI fragment at one end of clone Hbk9. Fig. 4 (left). Hybridization of human, monkey, and mouse genomic DNA with the human BKV-homologous repeats. Genomic DNA's (10 μg) were digested with either Eco RI or Hinf I and subjected to electrophoresis on a 1 percent agarose gel. After transfer to nitrocellulose paper (*29*), the digested DNA was hybridized with a ^{32}P-labeled nick-translated probe representing the 1.45-kb Eco RI–Pst I fragment from Hbk9. This fragment contains all the BKV-homologous sequences in the clone. Hybridization conditions were 3× SSC at 60°C. Lane M, ^{32}P end-labeled λ-Hind III marker; lanes 1 and 2, adult human leukocyte DNA, sample A; lanes 3 and 4, adult human leukocyte DNA, sample B; lanes 5 and 6, CV-1 cell monkey DNA; lanes 7 and 8, AGMK cell monkey DNA; lanes 9 and 10, NIH 3T3 mouse DNA. For each pair of DNA's, the left lane is an Eco RI digest, and the right lane is a Hinf I digest.

enhancer is shown in Fig. 5. A series of 20- to 21-bp repeats begins about 100 bp inside the terminal Eco RI site and extends for 1.4 kb, ending about 60 bp before the single Pst I site (Fig. 3). A representative group of 12 repeats is shown in Fig. 6, with the sequence of the BKV tandem repeat for comparison. Adjacent human genomic repeats are nonidentical but are > 70 percent homologous to one other. Like the BKV enhancer, the human tandem repeats are not especially GC-rich (this subclone is 53 percent GC). A 10-bp oligonucleotide represented in several repeats (GGTCATGGTT) matches the putative core sequence found in the BKV enhancer. Each repeat has a sequence that is homologous, but not always identical, to this core sequence. Further, the hexanucleotide CCTCCC present in the BKV triplication and similar to the hexanucleotide present in SV40 21-bp repeats (CCGCCC) appears frequently in the human subclone (double underline). We cannot exclude the possibility that a few copies of these repeats exist in an isolated form elsewhere in the human genome. On the basis of the Southern blot analysis (Fig. 4), however, it is unlikely that extensive tandem repeats such as those found in clone Hbk9 are present at another locus.

Enhancer Activity of the Human Repeats

Although the sequence length and arrangement of the human genomic tandem repeats is different from that of the BKV enhancer, certain similarities such as the core region and the GC-rich hexanucleotides are impressive. To examine the possibility that these structural similarities underlie an analogous function, we tested the human tandem repeats for cellular enhancer activity. The entire 1.8-kb Eco RI fragment containing all of the repeat units was cloned in both orientations at the 5′ end of the CAT gene (the Bgl II site) in $pA_{10}cat_2$, the vector used to test the BKV repeats (see Fig. 1) or at the 3′ end of the CAT gene (the Bam HI site) in the same vector. In the latter constructions, the human sequence is separated from the CAT gene promoter elements by 2 to 3 kb in the circular plasmid. Alternate orientations were designated s (sense), and a (antisense) with respect to their homology to the BKV repeats.

The recombinant plasmids p1.8s5′cat, p1.8a5′cat, p1.8s3′cat and p1.8a3′cat were introduced into HeLa, CV-1, or L cells under the same conditions that were used previously for plasmids containing the BKV enhancer. The CAT assays were performed on the cellular extracts prepared 48 hours after transfection.

Figure 7 shows representative time courses of CAT assays for the four plasmids as well as pBKs5′cat in each cell type. These data are shown in Table 1 as kinetic analyses normalized to the CAT activity of pBKs5′cat. In both HeLa and L cells the human tandem repeats enhance the expression of CAT in all positions and orientations. In these cells lines, they are generally five times less active than the BKV enhancer in analogous constructs but are 5 to 20 times more active than the control plasmid $pA_{10}cat_2$, which contains no enhancer element. Unlike the BKV enhancer, none of these constructs enhanced CAT gene expression significantly over background levels in CV-1 cells (Table 1).

Several subfragments of the 1.8-kb segment containing 4 or 12 human tandem repeat units were also subcloned into the $pA_{10}cat_2$ vector and tested for enhancer activity in all three cell types. Each of the smaller inserts induced approximately the same levels of CAT gene expression as the complete 1.8-kb fragment in HeLa and L cells, and were also inactive in CV-1 cells (data not presented).

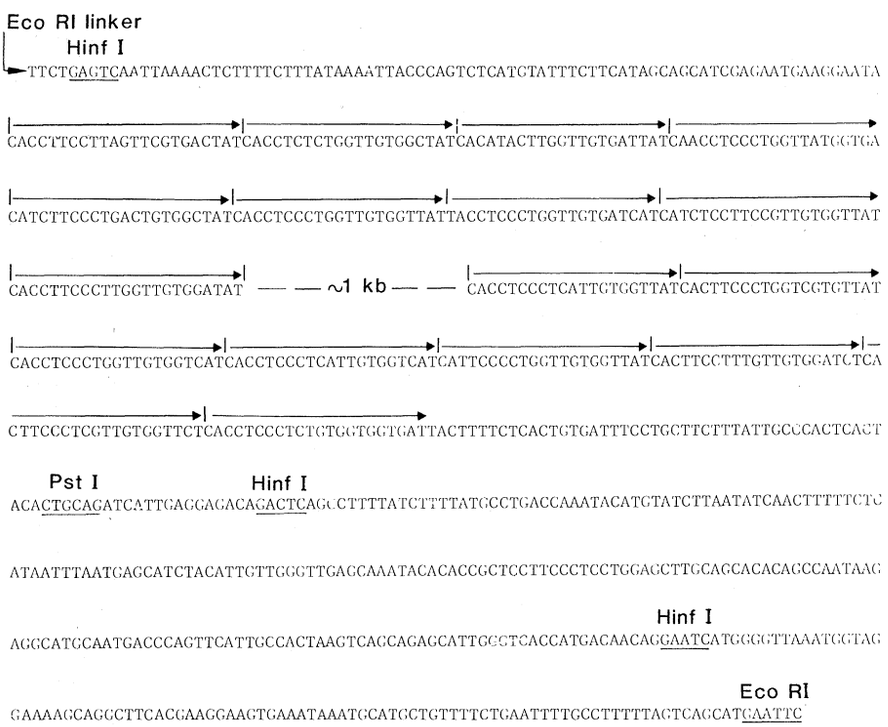


Fig. 5. Sequence of the BKV enhancer-homologous human locus and flanking DNA. Sequence analysis was performed on subclones of the 1.8-kb Eco RI fragment in an M13 vector (*30*), by the primer extension method in the presence of 2′,3′-dideoxy nucleoside triphosphates, and buffer gradient gels (*31*). Arrows above the sequences show the repeat periodicity beginning and ending within the sequenced fragment. The central 1-kb segment containing analogous repeats has been omitted for simplicity.

Evidence for Cellular Enhancers

The experiments presented above describe the characterization of the BKV enhancer element and the isolation of a human tandem repeated sequence homologous to it. When linked to a test gene and introduced into mammalian tissue culture cells, this human locus exhibits several characteristics ascribed to viral enhancers (*2*). First, the human sequence elevates the expression of the linked CAT gene from a heterologous SV40 promoter. Second, the expression of the CAT gene is increased by the human tandem repeats in both proximal and distal positions from the promoter. Finally, the orientation of the repeats relative to the CAT transcription unit, and the number of repeat units included in the constructs, do not qualitatively affect their enhancing capacity.

These host cellular sequences enhance CAT expression to levels between 5 and 20 times the baseline expression seen in the absence of enhancers. Compared to the levels of CAT expression in cells transfected with the BKV enhancer constructs, however, the homologous human sequences are only about 20 percent as active in HeLa and L cells and are essentially inactive in CV-1 cells. Several possible explanations can be advanced

for the lower level of enhancer activity of the human tandem repeats compared to the viral enhancer element. It seems reasonable to suggest that many host genes, expressed at low levels in a particular cell, are not associated with a strong enhancing element. Viruses have evolved to compete with cellular transcriptional units in the cells they infect. As a result, the affinity of viral enhancers for cellular factors may exceed the affinity of homologous cellular regulatory sequences for these factors. The comparatively low enhancing ability of the human tandem repeated element may be typical of certain cellular enhancers. In support of this model, similar enhancing ability has recently been reported for a random fragment of mouse DNA, isolated by its ability to restore the transformation potential of a polyoma virus, whose enhancer had been deleted. The sequence responsible for activating polyoma virus early gene expression has been mapped to a 58-bp segment of mouse DNA with no obvious homology to known viral enhancer sequences. Transformation of rat cells by the recombinant virus was 20 to 40 percent as efficient as that of wild-type virus (*32*).

Alternatively, the locus we have isolated may function at considerably higher levels in a tissue that we have not yet tested. Our observation that the human tandem repeated sequences act as enhancers in HeLa and L cells but not in CV-1 cells (even though the monkey genome contains homologous counterparts) may indicate that their activity is not host-specific but rather cell-specific. The possibility that enhancers may affect the expression of associated genes in a tissue-specific way is supported by studies with several viral enhancers (*8–12*) as well as by experiments involving polyoma viruses with mutations in their enhancer elements. These mutant viruses exhibit an altered phenotype that allows them to grow on undifferentiated teratocarcinoma stem cells, whereas wild-type polyoma virus cannot (*33*). Also, a naturally occurring variant of BKV with a single small deletion in the tandem repeated region can transform hamster and rat cells more efficiently than wild-type BKV but grows more slowly in human cells (*34*). This same variant induces an unusually high number of insulinomas when injected into hamsters (*35*).

The tissue-specific expression of some cellular genes may be controlled in the same way. The sea urchin histone H2a gene (*36*), and the human α-globin gene (*37*), are examples of cellular coding sequences that may be regulated by enhancer-like elements. Several groups of investigators have identified a region in introns of the mouse immunoglobulin heavy chain (*14, 15*) and light chain (*16*) genes which acts as an enhancer for the immunoglobulin gene expression. A short, repeated sequence, present in the heavy chain gene intron (*14, 15*), closely resembles the enhancer core sequence defined in SV40 (*2, 8, 11*). An interesting aspect of this discovery is that the immunoglobulin enhancers activate adjacent genes at high levels only in lymphoid cells (*14–16*). These studies suggest that differences in the sequences of either viral or cellular enhancer elements may dictate the tissue-specific transcription of adjacent genes through interaction with factors which are specifically present in those cells.

Other examples of short tandem repeated sequences have been found to be associated with several human structural genes. These include the polymorphic 14- to 15-bp repeats upstream of the insulin gene which appear to be unique in the genome (*38*), a block of 14-bp repeats in the intron of the zeta-globin gene (*39*), and an 800- to 900-bp stretch of 28-bp repeats in the 3′ flanking region of the proto-oncogene c-Ha-*ras*-1 (*40*). These sequences do not share extended homology with the BKV-like repeats, and their role in regulating the expression of linked genes remains to be determined.

The potential association of the BKV-homologous human tandem repeats with a functional gene will first require the identification of an adjacent sequence coding for a specific transcript. Although there are no open reading frames in the repeats or flanking DNA sequence so far examined, a potential TATA box and an appropriately spaced initiation codon near the internal Eco RI site at the end of the 1.8-kb fragment may signal the beginning of a coding region just beyond the end of the sequenced subclone.

Structure of Enhancer Elements

It seems clear that two different tandem repeated sequences fulfill an analogous role in the two closely related papovaviruses SV40 and BKV. Enhancer equivalents in cellular DNA might be expected to retain a structural similarity to the tandem repeated sequences of viral enhancers, reflecting a related function. Several previous studies were designed to test this possibility by isolating cellular sequences homologous to viral control elements. Sequences from both monkey (*41*) and human (*42*) genomes have been obtained by hybridization to the SV40 Hind III C fragment, which contains the origin of DNA replication as well as the transcriptional regulatory se-

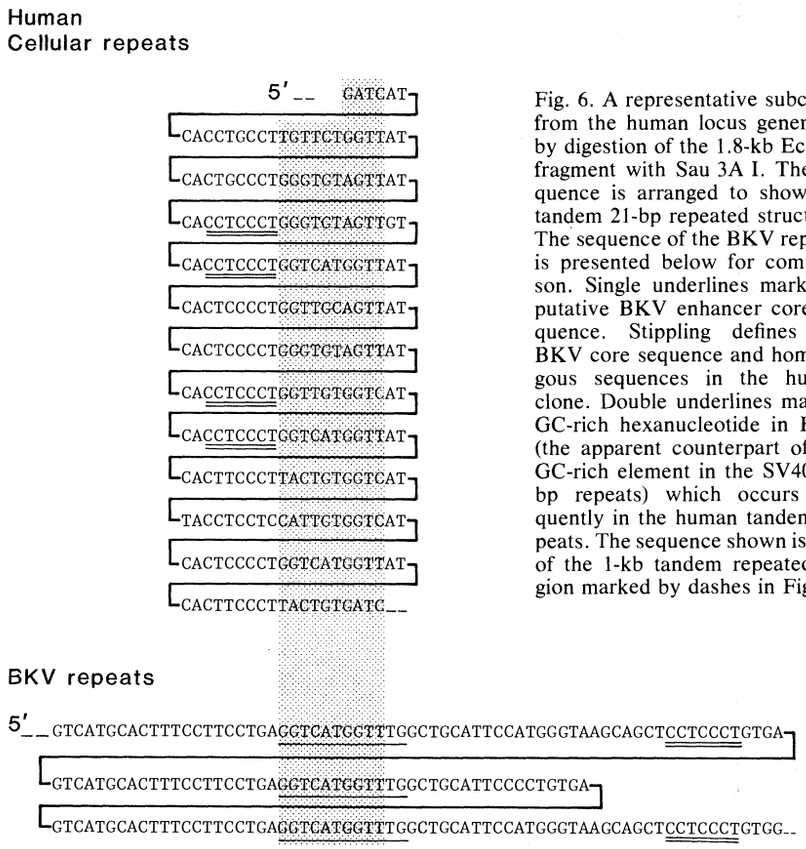


Fig. 6. A representative subclone from the human locus generated by digestion of the 1.8-kb Eco RI fragment with Sau 3A I. The sequence is arranged to show the tandem 21-bp repeated structure. The sequence of the BKV repeats is presented below for comparison. Single underlines mark the putative BKV enhancer core sequence. Stippling defines the BKV core sequence and homologous sequences in the human clone. Double underlines mark a GC-rich hexanucleotide in BKV (the apparent counterpart of the GC-rich element in the SV40 21-bp repeats) which occurs frequently in the human tandem repeats. The sequence shown is part of the 1-kb tandem repeated region marked by dashes in Fig. 5.

quences. Both the monkey and human sequences isolated with this probe were homologous to the region surrounding the viral origin of replication, containing multiple copies of the GC-rich hexanucleotide (CCGCCC) found in the SV40 21-bp repeats. The monkey sequences are present in the genome in about 100 copies and can promote bidirectional transcription without a TATA box in CV-1 cells (*43*), although no enhancer activity is observed (*44*). The human sequences related to SV40 have been shown to enhance the frequency of thymidine kinase positive (tk^+) colony selection when a linked herpes simplex virus tk gene is introduced into LTK^- cells (*42*). These sequences are present in multiple copies in the human genome.

In contrast, the human locus isolated with the BKV repeat probe appears to be unique in the genome. It does not hybridize with any other portion of the BK viral DNA besides the tandem triplication. Although no clear structural relationship exists between the BKV 68-bp triplication and the human 20- to 21-bp repeats, short stretches of homology, such as the enhancer core element (GGTCATGGTTAT) and the GC-rich hexanucleotide (CCTCCC), appear in many of the repeat units, and are probably responsible for the strong hydridization of the BKV probe to the human locus. Further, limited areas of homology to the human tandem repeats can be identified in other viral enhancer elements. For example, sequences in both the BKV tandem triplication (ATGCACTTTCCT) and several human cellular repeat elements (TATCACTTCCCT), are similar to a repeated element located in the adenovirus type 5 E1A enhancer region [TTTCACTTCCT, antisense strand (*6*)].

Table 1. Enhancer activity (relative) of the BKV repeats and a homologous human genomic segment. In the upper portion of the table is a kinetic analysis of CAT activity of the BKV constructs relative to pSV2cat in each cell type (Fig. 2). The lower portion of the table shows the kinetic analysis of the CAT activity of the human tandem repeat constructs relative to pBKs5′cat in each cell type (Fig. 7).

Plasmid	HeLa	CV-1	L
	BKV constructs		
pSV2cat	100	100	100
pBKs5′cat	25	18.5	25
pBKa5′cat	15.5	9.6	23.3
pBKs3′cat	9.8	3.5	8.8
$pA_{10}cat_2$	0.2	0.1	1.2
	Human constructs		
pBKs5′cat	100	100	100
p1.8s5′cat	25	5	33
p1.8a5′cat	23	5	33
p1.8s3′cat	19	4	30.8
p1.8a3′cat	19	4	27.2
$pA_{10}cat_2$	1	1	5.0

The varied arrangement of these homologous oligonucleotides in viral and cellular enhancers suggests that if certain essential elements are retained, enhancer structure may be flexible without loss of function. For example, the two major variants of BKV (Dunlop and MM) differ significantly in the arrangement of their repeated segments (*20, 21*). Other viable strains of BKV have been isolated which display a wider variation of sequence arrangements in the tandem repeat region (*34, 45*). A common ancestral sequence may have been acquired by the BKV prototype and evolved in different ways to produce a series of effective viral transcriptional control elements.

Clearly, the BKV enhancer did not derive directly from a transduced portion of the cellular repeats described above; the viral repeats contain stretches of sequence, absent from the human locus, which presumably contribute to the ability of viral enhancer to function both efficiently and independently of cell type. Yet, the persistence in the human genome of tandem repeats with both structural and functional analogies to the BKV enhancer implies a degree of evolutionary relatedness between viral control elements and host cell sequences. Whether the unique locus of cellular repeats represents a functional element for adjacent genes in the human genome remains to be determined.

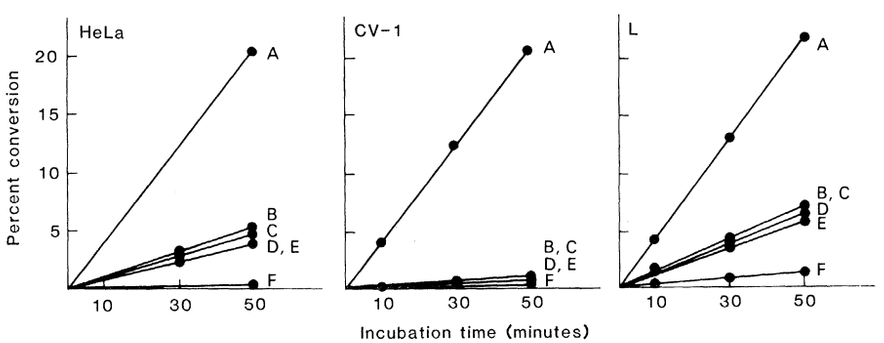


Fig. 7. Activation of the CAT gene activity by the human genomic repeat locus. The human genomic 1.8-kb Eco RI fragment (see Fig. 3) was subcloned via Bam HI linkers into either the 5′ or 3′ positions of $pA_{10}cat_2$ (Fig. 1). Plasmids containing this fragment in both orientations (sense and antisense) at each position (5′ and 3′) were selected. CAT assays were performed as described in Fig. 2 on cells transfected by the four recombinant plasmids p1.8s5′cat (B), p1.8a5′cat (C), p1.8s3′cat (D), p1.8a3′cat (E); and by pBKs5′cat (A) and $pA_{10}cat_2$ (F), the positive and negative controls, respectively. Values represent an average of three separate experiments, each performed on duplicate plates.

References and Notes

1. C. Benoist and P. Chambon, *Nature (London)* **290**, 304 (1981); P. Gruss, R. Dhar, G. Khoury, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 943 (1981).
2. P. Moreau, R. Hen, B. Wasylyk, R. Everett, M. P. Gaub, P. Chambon, *Nucleic Acids Res.* **9**, 6047 (1981); J. Banerji, S. Rusconi, W. Schaffner, *Cell* **27**, 299 (1981); M. Fromm and P. Berg, *J. Mol. Appl. Gen.* **1**, 457 (1982); *Mol. Cell. Biol.* **3**, 991 (1983).
3. J. de Villiers and W. Schaffner, *Nucleic Acids Res.* **9**, 6251 (1981); C. Tyndall, G. La Mantia, C. M. Thacker, J. Favaloro, R. Kamen, *ibid.*, p. 6231.
4. M. Lusky, L. Berg, H. Weiher, M. Botchan, *Mol. Cell. Biol.* **3**, 1108 (1983); Y.-C. Yang *et al.*, personal communication.
5. D. L. Weeks and N. C. Jones, *Mol. Cell. Biol.* **3**, 1222 (1983); J. Nevins *et al.*, personal communication; P. Chambon *et al.*, personal communication.
6. P. Hearing and T. Shenk, *Cell* **33**, 695 (1983).
7. B. Levinson, G. Khoury, G. Vande Woude, P. Gruss, *Nature (London)* **295**, 568 (1982); C. M. Gorman, G. T. Merlino, M. C. Willingham, I. Pastan, B. H. Howard, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 6777 (1982); D. J. Jolly, A. C. Esty, S. Subramani, T. Friedmann, I. M. Verma, *Nucleic Acids Res.* **11**, 1855 (1983); P. A. Luciw, J. M. Bishop, H. E. Varmus, M. A. Capecchi, *Cell* **33**, 705 (1983).
8. L. A. Laimins, G. Khoury, C. Gorman, B. Howard, P. Gruss, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 6453 (1982).
9. L. A. Laimins, P. Gruss, R. Pozzati, G. Khoury, *J. Virol.*, in press.
10. M. Kriegler and M. Botchan, *Mol. Cell. Biol.* **3**, 325 (1983).
11. H. Weiher, M. König, P. Gruss, *Science* **219**, 626 (1983); M. Kessel, personal communication.
12. B. J. Byrne, M. S. Davis, J. Yamaguchi, D. Bergsma, K. N. Subramanian, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 721 (1983).
13. J. de Villiers, L. Olson, C. Tyndall, W. Schaffner, *Nucleic Acids Res.* **10**, 7965 (1982); P. E. Berg *et al.*, *Mol. Cell. Biol.* **3**, 1246 (1983).
14. S. D. Gillies, S. L. Morrison, V. T. Oi, S. Tonegawa, *Cell* **33**, 717 (1983).
15. J. Banerji, L. Olson, W. Schaffner, *ibid.*, p. 729.
16. C. Queen and D. Baltimore, *ibid.*, p. 741.
17. S. D. Gardner, A. M. Field, D. V. Coleman, B. Holme, *Lancet* **1971-I**, 1253 (1971).
18. S. D. Gardner, *Br. Med. J.* **1**, 77 (1973); K. V. Shah, R. N. Daniel, R. Warszawski, *J. Infect. Dis.* **128**, 784 (1973); P. Brown, T. Tsai, D. C. Gajdusek, *Am. J. Epidemiol.* **102**, 331 (1975).
19. For a review see J. Tooze, Ed., *Molecular Biology of Tumor Viruses. DNA Tumor Viruses*, (Cold Spring Harbor Press, Cold Spring Harbor, N.Y., rev. ed. 2, 1981).
20. I. Seif, G. Khoury, R. Dhar, *Cell* **18**, 963 (1979).
21. R. C. A. Yang and R. Wu, *Science* **206**, 456 (1979).
22. P. M. Howley, M. F. Mullarkey, K. K. Takemoto, M. A. Martin, *J. Virol.* **15**, 173 (1975); G. Khoury, P. M. Howley, J. Garon, M. F. Mullarkey, K. K. Takemoto, M. A. Martin, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 2563 (1975); J. E. Osborn, S. M. Robertson, B. L. Padgett, D. L. Walker, B. Weisblum, *J. Virol.* **19**, 675 (1976).
23. R. D. Everett, D. Baty, P. Chambon, *Nucleic Acids Res.* **11**, 2447 (1983).
24. C. M. Gorman, L. F. Moffat, B. H. Howard, *Mol. Cell. Biol.* **2**, 1044 (1982).
25. M. Mandel and A. Higa, *J. Mol. Biol.* **53**, 154 (1970).

26. N. M. Maraldi, G. Barbanti-Brodano, M. Portolani, M. LaPlaca, *J. Gen. Virol.* **27**, 71 (1975).
27. B. Wasylyk, C. Wasylyk, P. Augereau, P. Chambon, *Cell* **32**, 503 (1983).
28. R. M. Lawn, E. F. Fritsch, R. C. Parker, G. Blake, T. Maniatis, *ibid.* **15**, 1157 (1978).
29. E. M. Southern, *J. Mol. Biol.* **98**, 503 (1975).
30. J. Messing and J. Vieira, *Gene* **19**, 269 (1982).
31. F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977); M. D. Biggin, T. J. Gibson, G. F. Hong, *ibid.* **80**, 3963 (1983).
32. M. Fried, M. Griffiths, B. Davies, G. Bjursell, G. La Mantia, L. Lania, *ibid.*, p. 2117.
33. M. Katinka, M. Yaniv, M. Vasseur, D. Blangy, *Cell* **20**, 393 (1980); F. K. Fujimura, P. L. Deininger, T. Friedmann, E. Linney, *ibid.* **23**, 809 (1981); M. Katinka, M. Vasseur, N. Montreau, M. Yaniv, D. Blangy, *Nature (London)* **290**, 720 (1981); K. Sekikawa and A. J. Levine, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 1100 (1981); F. K. Fujimura and E. Linney, *ibid.* **79**, 1479 (1982).
34. S. Watanabe and K. Yoshiike, *J. Virol.* **42**, 978 (1982).
35. _____, A. Nozawa, Y. Yuasa, S. Uchida, *ibid.* **32**, 934 (1979).
36. R. Grosschedl and M. L. Birnstiel, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1432 (1980).
37. P. Mellon, V. Parker, Y. Gluzman, T. Maniatis, *Cell* **27**, 279 (1981); R. K. Humphries *et al.*, *ibid.* **30**, 173 (1982).
38. G. I. Bell, M. J. Selby, W. J. Rutter, *Nature (London)* **295**, 31 (1982); P. S. Rotwein *et al.*, *N. Engl. J. Med.* **308**, 65 (1983).
39. N. J. Proudfoot, A. Gil, T. Maniatis, *Cell* **31**, 553 (1982).
40. D. J. Capon, E. Y. Chen, A. D. Levinson, P. H. Seeburg, D. V. Goeddel, *Nature (London)* **302**, 33 (1983).
41. C. Queen, S. T. Lord, T. F. McCutchan, M. F. Singer, *Mol. Cell. Biol.* **1**, 1061 (1981).
42. S. E. Conrad and M. R. Botchan, *ibid.* **2**, 949 (1982).
43. J. Saffer and M. Singer, in press.
44. S. Subramani and J. Saffer, personal communication.
45. J. van der Noordaa, personal communication.
46. P. Gruss and G. Khoury, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 133 (1981).
47. We thank Y. Barra, J. Brady, R. Brent, D. Cosman, P. Howley, G. Jay, M. Kessel, L. Laimins, M. Martin, R. Muschel, R. Pozzatti, J. Savarese, and M. Singer for helpful discussions and editing; A. B. Rabson and R. Koller for gifts of AGMK and human DNA's; J. Duvall, M. Chang, M. Priest, and D. Hawkins for technical assistance. N.R. was supported by a Damon Runyon–Walter Winchell Cancer Fund fellowship.

Insertion Sequence Duplication in Transpositional Recombination

Ted A. Weinert, Nancy A. Schaus, Nigel D. F. Grindley

Summary. Insertion sequences (IS) are discrete segments of DNA that can transpose from one genomic site to another and promote genetic rearrangements. A question that is central to understanding the mechanism of transpositional recombination is whether genetic rearrangements are accompanied by duplication of the IS that promotes them. Analysis of adjacent deletions mediated by IS*903* provides the strongest evidence to date than any IS-mediated transpositional recombination can occur by an efficient replicative mechanism.

The genomes of a wide variety of eukaryotic and prokaryotic organisms contain segments of DNA that can move from one location to another and can mediate other genetic rearrangements. Indeed, these transposable elements (transposons) provide the molecular basis for the genetic instability that has puzzled and intrigued geneticists for several decades (*1*).

Many different transposons have been identified in Gram-negative bacteria and extensively characterized. Most of them fall into one of two groups (*2*): the rather homogeneous Tn*3* family of transposons, and the more heterogeneous collection of insertion sequences (IS) (together with the composite transposons that contain a segment of DNA flanked by two copies of an IS). Attempts to understand the mechanisms of transpositional recombination have focused on genetic and structural characterization of the transposons themselves and of the products formed during recombination. One of the primary questions has been whether a transposon is duplicated during transpositional recombination.

In the case of Tn*3* and related transposons there is a strong evidence that transposition is a replicative process (*3*). During the first stage of transposition, donor and target replicons are fused to form a cointegrate. This intermediate contains two copies of Tn*3*, one at each junction between donor and target DNA sequences (*4*). From this cointegrate, simple insertions can then be generated by a site-specific recombination between the two transposon copies. For the second group of transposons, the IS elements, the evidence for replicative transposition is much weaker. Although IS elements do promote the fusion of replicons, this occurs only infrequently (at about 1 percent of the frequency of simple insertions) (*5–7*), and once formed the resultant cointegrates are generally stable in a RecA$^-$ host strain. Simple insertions of an IS are therefore thought to be generated not by breakdown of a cointegrate intermediate but rather by a one-step process that results directly in the integration of a single IS copy at a target site. From the strong dependence of cointegrate formation on *recA* activity that is found with the IS*50* composite transposon, Tn*5* (*7*), it has been argued that the cointegrates observed may result from simple insertions from dimers of the donor replicon rather than from true replicon fusion between the target and a monomer of the donor (*8*). Apart from cointegrate formation, the only other overtly replicative transpositional process is "inversion-insertion" (also called "duplicative inversion") (*2*): the inversion of a DNA segment adjacent to a transposon, coupled to an insertion of the element in the opposing orientation at the other end of the inverted DNA segment (see Fig. 5B). However, only three examples of such IS-mediated inversion-insertions have been documented (two with IS*1* and one with IS*10*) and all were detected in RecA$^+$ hosts (*9, 10*). Thus the possibility remains that these rearrangements took place in two unlinked steps: an intermolecular insertion of the IS in inverted orientation, followed by *recA*-mediated recombination between the two IS copies to invert the intervening DNA segment.

With evidence for the apparently replicative IS transpositional processes of replicon fusion and inversion-insertion in some doubt, it is not surprising that evidence for a replicative process in the formation of simple insertions or a fourth IS-promoted genetic rearrangement, adjacent deletions, is essentially nonexis-

T. A. Weinert is a graduate student, N. A. Schaus was a postdoctoral fellow, and N. D. F. Grindley is an associate professor in the Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06510. Nancy Schaus is currently a senior biologist in the Molecular Genetics Department, Lilly Research Laboratories, Indianapolis, Indiana 46285.