TECHNIQUES FOR ISOLATING HUMAN CHROMOSOMAL DNA SURROUNDING AN INTEGRATED HSV-1 THYMIDINE KINASE GENE

BY

Marina Vasiliki Pascali

Department of Microbiology and Immunology

McGill University

Montreal, Canada

July, 1990

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Science

ABSTRACT

The KT cell line was developed to study the nature of spontaneous mutations in eukaryotic cells. The plasmid pSV2*neo*KT, containing the herpes simplex virus type 1 (HSV-1) thymidine kinase (*tk*) gene, was used to transfect human AK143*tk*$^-$ cells. A *tk*$^+$, G418^R clone, containing a single integrated copy of the plasmid, was isolated and expanded to form the KT cell line.

Many *tk*$^-$ mutants were isolated and several were shown to revert at a high frequency. A potentially repetitive sequence organization was found at the termini of two deletion mutants, suggesting that neighbouring DNA sequence may play a role in the effects observed in the *tk* gene. To determine the influence of chromosomal location on the rates and types of mutations generated at the *tk* locus, several techniques were attempted to isolate the gene and its flanking cellular sequences.

The first method, plasmid rescue, was designed to recover a 15kb autonomously replicating plasmid containing the integrated pSV2*neo*KT sequences and the cellular DNA surrounding the integration site. Concurrently, a second method was pursued to clone each junction site separately as two smaller fragments using the cloning vectors pUC119 or pBR322. All of the rescued plasmids were smaller than expected, and a large number of recombinant plasmids con-

tained deletions of the cloned genomic DNA. It is possible that repetitive sequences in the flanking cellular DNA played a role in generating the observed deletions.

Three variations of the polymerase chain reaction (PCR) were also used to amplify the cellular DNA flanking the integration site : anchored PCR, single-specific-primer PCR and inverse PCR. All three methods were developed to amplify specific urknown double-stranded DNA in cases where sequence information is available at only one extremity. The validity of each technique is evaluated and further modifications and improvements are discussed.

RESUME

La lignée cellulaire KT a été développée afin d'étudier la nature des mutations spontanées chez les cellules eucaryotes. Le plasmide pSV2*neo*KT, portant le gène de la thymidine kinase (*tk*) du virus herpes simplex type 1 (HSV-1) a éte utilisé pour transfecter la lignée cellulaire AK143*tk*$^-$. Un clone *tk*$^+$, G418^R, et contenant une seule copie intégrée du plasmide a été isolé et cultive formant ainsi la lignée cellulaire KT.

Plusieurs mutants *tk*$^-$ ont eté isolés et nombre d'entre eux démontraient une fréquence élevée de réversion au type sauvage. L'organisation d'une séquence potentiellement repetitive a ete démontrée aux extrémités de deux mutants de déletion, suggerant que la séquence d'ADN environnante est reliee aux effets observés dans le gene *tk*. Afin de determiner l'influence de la situation chromosomique sur les taux et les types de mutations générés au locus *tk*, plusieurs techniques ont éte utilisees pour isoler ce gene et les séquences cellulaires le flanquant. La première methode, la recuperation de plasmide, a éte développée pour isoler un plasmide de 15 kb se répliquant de façon autonome et contenant les séquences de pSV2*neo*KT intégré ainsi que l'ADN cellulaire entourant le site d'intégration. Simultanement, une seconde méthode a eté utilisée pour cloner séparément chacune des deux jonctions sur deux frag-

ments plus petits en utilisant les vecteurs de clonage pUC119 ou pBR322. Tous les plasmides récuperes etaient plus petits que prévu et plusieurs de ces plasmides recombinants présentaient des déletions de l'ADN génomique clone. Il est possible que des séquences répétitives presentes dans l'ADN cellulaire flanquant le plasmide integre aient joue un rôle dans la genération des délétions observees.

Trois variantes de la réaction en chaîne de la polymérase (RCP) ont aussi éte utilisees pour amplifier l'ADN cellulaire flanquant le site d'integration: RCP ancrée, RCP avec séquence-amorce specifique unique et RCP inversée. Ces trois méthodes ont ete developpees pour amplifier spécifiquement l'ADN double-brin inconnu dans les cas où la séquence d'ADN n'est disponible que pour une seule extrémité. La validité de chacune de ces techniques est évaluées et des modifications et ameliorations appropriees sont suggérées et discutées.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Michael DuBow for sparking my interest in research and encouraging me to begin my M.Sc. degree. A very special thanks to students and staff in the Department of Microbiology and Immunology for their valuable discussions on a variety of matters including science in general, Mother Pucker techniques, and Crossword analysis and T. H. (Tall Hunks) reviews in particular. Some of these people include Amelia Anteka, Chantal Autexier, Ginette Berubé, Josee Brisebois, Robin Cameron, Caroline Diorio, Robert Gendron, Lina Guzzo, George Kukolj, David Levin, Joy Marmor, Anne Nudo, and Peter Ulycznyj.

A most gracious THANK-YOU! to Josee Brisebois and Peter Ulycznyj for painstakingly and meticulously reading and correcting the early drafts of this thesis. Your comments and suggestions were a breath of fresh air to a clouded mind. I would also like to thank Chantal Autexier, Robin Cameron, and Bob Gendron for reading the last draft and checking all the details.

Finally, I would like to thank my parents for their emotional and financial support throughout my M.Sc. degree, Josée Brisebois for a friendship that is rare and precious and Nicolai van Oers for his un nding support and encouragement which extended over the years and across the continent.

LIST OF ABBREVIATIONS

ACV:	Acyclovir
Aprt:	Adenine phosphoribosyl transferase
bp:	base pairs
DHPG:	9-(1,3-dihydroxy-2-propoxy-methyl)guanine
DNA:	Deoxyribonucleic acid
DTT:	Dithiothreitol
EDTA:	Ethylenediaminetetraacetic acid
HAT:	Hypoxanthine, amenopterin, thymidine
Hprt:	Hypoxanthine phosphoribosyl transferase
HSV-1:	Herpes simplex virus type 1
kb:	kilobase pairs
Neo:	neomycin
nt:	nucleotides
SDS:	Sodium dodecyl sulfate
SSC:	Saturated sodium citrate
ssDNA:	Single-stranded DNA
Tk:	Thymidine kinase
TFT:	Trifluorothymidine

LIST OF FIGURES

PAGE

LIST OF TABLES

PAGE

I. LITERATURE REVIEW

The generation of genetic diversity is critical for the adaptation and survival of a species in a changing environment. Spontaneous mutations provide the raw material for evolution through changes in the hereditary material of a cell. If the mutation is neutral,having no immediate effect, it may be retained or lost at random. However, if the mutation confers a survival advantage to the population, it can be maintained in the gene pool by natural selection. In contrast, deleterious mutations are generally lost from the gene pool because they hinder the individual's ability to survive and reproduce. These types of mutations are frequently associated with the development of genetic disorders including Duchenne muscular dystrophy (Worton, 1988) and embryonic tumours such as retinoblastoma (Godbout *et al.*, 1983). To elucidate the mutational mechanisms underlying evolutionary change and the development of disease, researchers have focused their studies on deleterious mutations, since these can often result in detectable altered phenotypes.

Our understanding of the molecular basis of spontaneous genetic mutations is based primarily on the studies of bacterial systems, particularly those involving the *lacI* gene of *Escherichia coli*. Mutants identified on the basis of their constitutive expression of the *lac* operon (Miller *et al.*, 1977) were efficiently cloned into plasmids for

nucleotide sequence analysis. In this manner, a variety of spontaneous mutational events were identified, including base substitutions, frameshifts, deletions, insertions, duplications and complex rearrangements (Schaaper et al., 1986). Many mutations did not occur randomly but appeared in clusters as postulated by Seymour Benzer in 1961 (Benzer, 1961). The nature of these mutation "hotspots" was shown to be influenced by the characteristics of the surrounding DNA, and in particular by the presence of sequence motifs such as repetitive elements and regions of dyad symmetry (Farabaugh et al., 1978).

Comparable analyses of mutational mechanisms in the eukaryotic genome are hindered by its greater structural complexity, the diploid nature of the genome, and lack of facile genetic manipulations. To reach a molecular understanding of eukaryotic mutational mechanisms, it is therefore necessary to develop a system that permits the recovery and analysis of both a mutant allele and its adjoining sequences. The following chapter will review the systems developed in cultured mammalian cells that address the difficulties inherent in working with diploid organisms and that offer a reliable method of allele recovery, which is the underlying interest of this thesis.

A. MAMMALIAN CELL SYSTEMS

In diploid organisms, each genetic locus is represented by two alleles that are located on homologous chromosomes. Consequently, mutations that alter or inactivate one allele may remain undetected because their effect is frequently masked by the functioning homologous allele. Furthermore, unless a mutation confers a dominant selectable phenotype to the affected allele, it is difficult to distinguish a functional allele from a non-functional one at the molecular level.

Systems developed to study spontaneous mutations in eukaryotic cells may be divided into two major groups. The first group analyses endogenous genes, and is generally restricted to the study of genes that occur in a single copy, either on the X chromosome or in specialized cell lines. The second group analyses exogenously introduced genes and is therefore more flexible because any gene of interest may be introduced. Well characterised bacterial and viral genes offer the advantage of a comparison between the types of mutations produced in eukaryotic and either bacterial or viral systems.

A.1. Endogenous Gene Systems

To simplify mutation studies in mammalian genomes, a number of systems exploit the fact that males contain only one copy of both the X and Y chromosomes. Therefore, genet-

ic loci located on either of these two chromosomes are represented by only one allele - the locus is hemizygous. The gene coding for hypoxanthine-guanine phosphoribosyl transferase (hprt) is located on chromosome X and has been the focus of mutational studies in mammalian genomes because there is no homologous allele to interfere with the interpretation of molecular changes(Yang et al., 1984; Thacker, 1985). Since the hprt gene product is an enzyme that functions in the nucleotide salvage pathway, it is not essential for normal cell growth in ordinary medium. Therefore, all classes of mutational events are expected to occur at this locus. Hprt$^-$ mutants are easily selected by treating cells with nucleotide analogs such as 6-thioguanine. Cells maintaining a functional enzyme incorporate the analogs as nucleotides in their DNA which leads to cell death. However, cells containing a mutation that inactivates hprt are unaffected by the treatment.

Results from tissue culture studies involving the hprt gene locus indicate that the majority of spontaneous mutants are the result of point mutations or small alterations (less than 50 base pairs (bp)), which fall below the resolution of Southern blot analysis (Fuscoe et al., 1983; Albertini et al., 1985; Gennett and Tilly, 1988). These results correlate with the spectrum of mutations identified in patients suffering from Lesch-Nyhan syndrome (a complete HPRT enzyme deficiency), in which 25 out of 28 individuals showed no detectable alterations at the hprt locus (Yang et al.,

1984). Although the tissue culture and population studies are in agreement, there is a question regarding the validity of extrapolating data from hemizygous loci such as hprt, to account for mutational events occurring throughout the genome (see below). Despite the fact that the hprt gene is present in single copy, molecular analysis has been frustrated by the presence of cross-hybridizing pseudogenes (Gennett and Tilly, 1988; Yang et al., 1984), and by the large size of the gene itself - greater than 33 kilobase pairs (kb) in the mouse (Melton et al., 1984 ; 44kb in Patel et al., 1986).

Adenine phosphoribosyl transferase (aprt), like hprt, encodes a non-essential enzyme in the nucleotide salvage pathway and has also been used extensively in mutation studies. The apparent absence of pseudogenes (Grosovsky et al., 1986) and its relatively small size: 2.1-4.3kb (Thacker, 1985), facilitate Southern blot analysis and molecular characterization. Unlike hprt, aprt is an autosomal gene and is normally present in diploid copy number. However, chinese hamster ovary cell lines hemizygous for the aprt locus have been isolated (Bradley and Letonavec, 1982; Nalbantoglu et al., 1983), thereby simplifying analyses and enabling convenient isolation of $aprt^-$ mutants. The spectrum of aprt mutations is similar to that observed at the hprt locus : most of the mutants isolated revealed no detectable alterations by Southern blot analysis (Nalbantoglu

et al., 1983; Grosovsky et al., 1986). Sequence analysis showed that these mutations included single-base substitutions and small intragenic rearrangements or deletions of less than 50bp (Phear et al., 1989; Nalbantoglu et al., 1986; reviewed in Thacker, 1985).

Although the spectra of both aprt and hprt mutants are similar, it has been suggested that hemizygous loci may be inadequate indicators of certain mutagenic events. For example, large multi-locus deletions may inactivate one or more genes within the region of hemizygosity that are essential for cell viability (Evans et al., 1986; Stankowski et al., 1986). However, Cox and Masson (1978) recovered hprt mutations with cytogenetically visible alterations of the X chromosome, indicating that at least some large chromosomal alterations can lead to viable loss of a hemizygous locus.

Thymidine kinase (tk) is another autosomal gene which codes for a non-essential salvage pathway enzyme involved in nucleotide metabolism. In cell lines that are heterozygous at this locus, one tk allele is functional (tk^+) and the homologous allele, though structurally unrearranged, is non-functional (tk^-) (Yandell et al., 1986; Moore et al., 1985). In both rodent and human cell lines, two classes of tk^- mutants were recovered which were distinguished phenotypically by their growth rates. The majority (95%) of slow growing tk^- mutants resulted from chromosomal deletions and rearrangements that lost the entire functional tk allele. In contrast, analysis of tk^- mutants growing at a normal

rate revealed that smaller intragenic alterations predominated as the major mutational event, as previously found at the aprt and hprt hemizygous loci (reviewed in Thacker, 1986; Yandell et al., 1986; Moore et al., 1989; Applegate et al., 1990).

The higher frequency of chromosomal rearrangements observed at the heterozygous tk locus may be due to the action of additional mutational mechanisms which require homologous DNA sequences to operate (Liber et al., 1989; Moore et al., 1989). Although gene conversion and homologous recombination may occur between a hemizygous locus and a pseudogene or other partially homologous sequence, they are expected to occur less frequently than at a heterozygous locus which contains two homologous alleles (Liber et al., 1989).

At least three other factors may affect the relative rates and types of mutations observed at different gene loci in different cell system. First, the higher frequency of mutation observed at autosomal loci in rodent versus human cell lines may be due to the relatively frequent occurrence of heterozygosity in the human genome (Vogel and Motulsky, 1986). If many human chromosomes contain non-functional alleles of essential genes, then large deletions, chromosomal loss or large-scale recombinational events that inactivate the homologous functional allele may result in predominantly lethal mutations. However, rodent cell lines are

derived from inbred strains and are predominantly homozygous at all loci. Therefore, loss of an entire chromosome or large-scale recombination would not necessarily result in a lethal mutation because of the presence of a functioning homologous allele. Second, the mutational differences observed at distinct loci may be due to the size and structural features of the particular gene being studied. Third, locus-specific effects may be influenced by the presence or absence of mutational hotspots (Wahls et al., 1990), closely linked essential genes (Evans et al., 1986), or other sequence features such as neighbouring repetitive DNA elements (Myerowitz and Hogikyan, 1987).

A.2. Exogenously Introduced Genes

Difficulties encountered in the analysis of endogenous mammalian genes may be overcome by introducing foreign genes as targets of mutation. Advances in recombinant DNA technology led to the construction of plasmids called "shuttle" vectors, that are capable of replicating and expressing genes in both bacterial and eukaryotic cells (Mulligan and Berg, 1980). In these systems, a cloned gene that is part of a shuttle vector can be exposed to the mammalian cell environment and then rescued into bacteria for the rapid detection and analysis of mutations that were generated in the eukaryotic cells. The major advantage of shuttle vector systems is that large amounts of the target gene can be rapidly isolated and analysed using established bacterial

techniques, thus bypassing the complexity and limitations of somatic cell genetics. In addition, direct comparisons of mutagenesis in eukaryotic and prokaryotic cells can be made if the target is a well characterized bacterial gene.

Shuttle vector systems are grossly divided into two groups : those that replicate autonomously and those that integrate into the host genome. The simplest shuttle vectors differ from bacterial plasmids by the inclusion of a mammalian origin of replication (i.e. simian virus 40 (SV40)). When the vector is transfected into tissue culture cells that are permissive for the virus (COS cells), the SV40 origin is activated by the large T antigen that the cells constitutively express. This activation permits the autonomous, extrachromosomal, replication of the vector which can then be retrieved into bacterial cells through a Hirt extraction (Hirt, 1967) and transformation.

Early studies with shuttle vectors carrying the *lacI* and *galK* bacterial genes uncovered an unexpected high frequency of spontaneous mutations in the transfected DNA (Calos *et al.*, 1983; Razzaque *et al.*, 1983). Approximately 1 to 10% of the vector molecules contained detectable alterations, compared with the endogenous spontaneous mutation rates in *E. coli* of 2.5×10^{-6} per cell per generation (Miller *et al.*, 1977), and in mammalian cells of 10^{-6} to 10^{-8} per cell per generation (Lewin, 1985). Further studies determined that the high frequency alterations were confined to

the transfected sequences, suggesting that critical differences exist between the cellular and vector DNA that may be responsible for the latter's greater mutational susceptibility. (Lebkowski et al., 1984; Razzaque et al., 1984). It was proposed that most of the mutations occur prior to replication and involve the nucleolytic degradation and ligation of vector DNA molecules (Lebkowski et al., 1984; Razzaque et al., 1984; Ashman and Davidson, 1985).

Several modifications to the SV40 shuttle vectors (Sarkar et al., 1984; Ashman and Davidson, 1985; Lebkowski et al., 1985; Seidman et al., 1985), and the use of other viral-based vectors (bovine papilloma virus) (Ashman and Davidson, 1985); Epstein-Barr virus (Menck et al., 1987; Drinkwater and Klinedinst, 1986)), have succeeded in increasing vector stability. But the background mutation frequency remains high (10^{-2} to 10^{-5} per cell per generation), precluding their use in the study of spontaneous events. Furthermore, since these vectors replicate extrachromosomally, they may not be exposed to the same mutation pressures that affect chromosomal loci and so may not accurately represent all of the mutational events that occur at an endogenous mammalian gene.

The second major group of shuttle vectors also contain the SV40 origin of replication, but these plasmids are transformed into mammalian cells that do not express large T antigen. Consequently, the vectors do not replicate autonomously and cells can be isolated which contain a

single integrated copy of the foreign DNA. The integrated plasmid can then be rescued by fusing the transformed mammalian cells with COS cells and isolating the plasmids in E. coli as previously described (Breitman et al., 1982). Initial experiments using these SV40-based shuttle vectors also showed a high frequency of spontaneous mutations (1.7% of the vectors contained detectable alterations), which resembled the types of mutations generated by COS cell transfections (Ashman and Davidson, 1984). It is possible that nucleolytic degradation, previously implicated in the vector instability of COS cell transfections, is responsible for the high mutation rate observed in COS cell fusions, but the mechanism(s) remains unknown.

Ashman and Davidson modified the SV40-based shuttle vectors by adding retroviral elements and developed a system which distinguishes the truly spontaneous mutations occuring on the chromosome from those produced during the rescue procedure (Ashman et al., 1986). The retroviral shuttle vectors contain two long terminal repeats (LTRs) which are required for the initiation and polyadenylation of viral transcripts as well as for the integration of viral sequences into chromosomal DNA. Sequences for reverse transcription of the viral genome and encapsidation of viral RNA are also present. The retroviral vector is packaged into a transmissible virus following transfection into psi-2 cells which provide all of the functions in trans that are neces-

sary for packaging RNA into virus. Medium from psi-2 infected cells is then used as a source of virus to transfect eukaryotic cells. A cell line containing a single copy of the target gene was readily obtained since retroviral vectors usually integrate into chromosomal DNA in single copy. Unlike autonomously replicating vectors and the integrated vectors discussed above, mutations are selected while the vector is still integrated in the mammalian chromosome. The proviral structure of the integrated vector ensures that the SV40 origin is closely flanked by the viral LTRs. Therefore, following COS cell fusion and activation of the SV40 origin, the vector sequences replicate and are excised by homologous recombination between the viral LTRs. The resulting plasmids are recovered by transformation into E. coli.

Ashman and co-workers found that approximately 50% of the plasmids recovered from a single COS cell fusion contained the same DNA sequence alteration, while the remainder exhibited a variety of large deletions and rearrangements. The ability of this system to distinguish spontaneous mutations which occurred in the integrated vector from those that occur during the COS cell fusion is based on the assumption that the plasmids recovered which exhibit the same DNA sequence alteration are identical to the mutation selected for in mammalian cells. The spectrum of spontaneous versus induced mutations isolated with this system are different, suggesting the genetic alterations characterized

were indeed generated in the mammalian cells and are not due to the recovery method (Ashman and Davidson, 1987a; Ashman and Davidson, 1987b; Davidson et al., 1988).

B. ELEMENTS THAT INFLUENCE GENE EXPRESSION IN MAMMALIAN CELLS

Sequence analysis of both endogenous and exogenously introduced genes in mammalian cells revealed that frameshift and deletion events in eukaryotic genes occurred predominantly at the sites of direct and inverted repeats (deJong et al., 1988). Direct and inverted sequence repeats have been identified at deletion hotspots in the lacI gene of E. coli and are believed to play an important role in the mechanism of mutation through misalignment slippage and the formation of stable stem-loop structures (Albertini et al., 1982; Farabaugh et al., 1978). In addition to the mutation mechanisms shared by both bacterial and mammalian systems, eukaryotic gene expression is influenced by regulatory elements not characteristic of prokaryotes, including epigenetic events (Dobrovid et al., 1988) and interspersed repetitive sequences (Jeffreys et al., 1985).

B.1. Epigenetic events

Epigenetic events are distinguished from structural mutations by a higher mutation frequency and a high rate of reversion to the wild type phenotype. Site-specific modifi-

cations, such as DNA methylation, are considered epigenetic events. Methylation, particularly of cytosine residues at CpG dinucleotides, has been implicated in the regulation of gene expression in mammalian cells. For example, the expression of most housekeeping genes appears to depend on the unmethylated state of a CpG-rich promoter region (Bird, 1986), whereas genetic inactivation of a number of other loci correlates with the methylation of particular DNA sequences (Dobrovic et al., 1988; Tasseron- deJong et al., 1989).

However, a report by Woodcock et al. (1987) indicates that although 40 to 80% of all CpG dinucleotides are methylated in mammalian genomes, the majority of methylated cytosines occur in the CpA, CpC and CpT dinucleotides (54.5%). These calculations suggest that methylated cytosines at sites other than the CpG dinucleotide may play determining roles in the regulation of gene activity and warrant further investigation.

In contrast to these site-specific modifications a second, more general type of epigenetic event is involved in gene inactivation. The phenomenon is called position-effect variegation and is characterized by the inactivation of a gene located near a region of heterochromatin, an area of the chromosome that is condensed and generally associated with transcriptional inactivity. How heterochromatization regulates gene activity is not known, but studies in Drosophila development have identified a gene whose product may

modulate transcriptional activity by tightly packing the chromatin fibre into heterochromatin (Reuter et al., 1990). A correlation between alterations in chromatin structure and changes in gene expression has also been observed in mammalian genomes by monitoring site-specific changes in DNaseI sensitivity (Davies et al., 1982). Furthermore, particular stretches of alternating purine and pyrimidine residues may be involved in shaping chromatin structure by forming left-handed DNA called Z-DNA (Pardue et al., 1987).

B.2. Repetitive DNA

Eukaryotic genomes, unlike prokaryotes, are composed of a variety of repetitive DNA elements which have been associated with genome instability, rearrangements, and deletions. Up to 5% of the total eukaryotic genome is represented by tandem blocks of highly repetitive DNA called satellite DNA (Lewin, 1985). There is evidence that shorter, tandem repetitive units, called minisatellites, promote homologous recombination in mammalian cells (Jeffreys et al., 1985; Wahls et al., 1990). Genetic recombination is known to produce a variety of mutations including deletions, inversions, duplications and translocations (Allgood and Silhavy, 1988). The common feature shared by hypervariable minisatellite DNAs is a short consensus core sequence that resembles the known recombination hotspot "chi" in E. coli (Smith et al., 1981). The similarity led Jeffreys and co-

workers (1985) to hypothesize that the minisatellite core sequence is involved in eukaryotic recombination events. Homologous core sequences have been identified in the recombination hotspots of the major histocompatibility complex (MHC)(Kobori et al., 1986; Steinmetz et al., 1936), and other non-homologous repeat sequences have been associated with recombination hotspots in the human beta-globin and insulin loci (Bell et al., 1982; Lebo et al., 1983; Treco et al., 1985).

In addition to the clustered tandem repeats of satellite DNA, highly repetitive elements are found dispersed throughout the mammalian genome. The predominant interspersed repetitive element in man is the Alu sequence, which consists of a basic unit of ~300 base pairs and exists in ~300,000 copies per haploid genome (Jelinek and Schmid, 1982). Recombination events between Alu sequences have been implicated in the deletion mutation responsible fc Tay-Sachs disease in French Canadians (Myerowitz and Hogikyan, 1987), the deletion and duplication mutations in the low density lipoprotein (LDL) receptor gene (Lehrman et al., 1985, 1987), and certain deletions at the globin locus (Vanin et al., 1983).

C. ISOLATION OF SPECIFIC EUKARYOTIC DNA SEQUENCES

The observation that epigenetic events and repetitive sequence motifs can influence gene expression underscores the importance of characterizing both the mutation target

and flanking DNA sequences in order to elucidate the mechanisms of mutational events. Integrated shuttle vectors and their adjoining cellular sequences can be rescued via COS cell fusion. However, it is difficult to determine with this method whether the recovered DNA was rescued intact or whether it was altered by the procedure itself. Several methods developed to isolate chromosomally located genes flanked by the adjoining cellular sequences will be reviewed in the following sections.

C.1. Construction of Lambda Libraries

Bacteriophage lambda is a temperate bacteriophage of E. coli and has a double-stranded DNA genome of 48,502bp (Sanger et al., 1982). Generally, bacteriophage lambda vectors are utilized for the cloning of DNA sequences from complex genomes because of the ability to insert large fragments of DNA and the high efficiency of in vitro packaging of recombinant molecules into infectious phage particles. A stringent requirement for DNA size imposed by the packaging reaction (78% to 105% of the wild type genome length) demands a different strategy for the cloning of small versus large DNA fragments (Feiss and Siegele, 1979).

Small fragments of DNA are readily cloned into "insertion vectors" at a single restriction site in a non-essential part of the lambda genome. Ideally, insertion of a DNA fragment will inactivate a non-essential gene and

recombinants can be recognised by a changed phenotype.

Although the lambda genome contains about 50 genes, only 50% are essential for phage growth and plaque formation. In the construction of lambda vectors capable of accepting 10 to 20kb of foreign DNA, the entire region containing the non-essential genes is replaced. Digestion of the vector with two restriction endonucleases can reduce religation of the non-essential, or "stuffer", fragment, and recombinants can be directly selected by infecting host strains which prevent growth of wild type phage.

In a typical cloning procedure, lambda phage heads and tails are supplied *in vitro*, the recombinant DNA is packaged and the infectious lambda particles are plated on a lawn of bacteria. The specific genomic clone is then identified with nucleic acid probes by one of two methods. The first relies on hybridization of a DNA or RNA probe to phage DNA immobilized on nitrocellulose filters (Benton and Davis, 1977). For example, investigators used this procedure to analyse spontaneous mutations at the *aprt* locus (Nalbantoglu *et al.*, 1987), and the *Caenorhabditis elegans unc-54* locus (Pulak and Anderson, 1988), as well as to isolate unknown cellular DNA flanking the integrated hepatitis B virus (Takada *et al.*, 1990).

The second method, *in vivo* recombination technique, relies on homologous recombination between a defective lambda vector and a plasmid DNA harbored by the *E. coli* host strain (Seed, 1983). Incorporation of a small plasmid

carrying the supF gene into a phage vector containing an amber mutation results in suppression of the mutation, and recombinant phages can be selected on a nonsuppressor host. In vivo recombination has been used in mammalian cells to repeatedly isolate spontaneous mutations at the CHO aprt locus (deJong et al., 1988). Although cloning in lambda vectors has proven successful, there is some evidence that certain regions of genomic DNA are underrepresented in phage libraries and must be cloned by other methods (Nicholls et al., 1987).

C.2. Plasmid Rescue

The plasmid rescue technique was developed to recover both integrated vector DNA and its flanking cellular sequences from transformed cell lines (Perucho et al., 1980). This method requires that a vector contain a functional plasmid origin of replication as well as an antibiotic resistance gene for selection in E. coli. In a typical procedure, DNA is digested with a restriction endonuclease that does not cleave within the vector sequences. The digested DNA is then ligated under conditions that promote the self-ligation (and cyclization) of each DNA fragment. Only the circular DNA fragment containing the vector sequences will replicate after transformation into E. coli and can be selected on media containing the appropriate antibiotic. By comparing the restriction pattern of the plasmids

rescued in E. coli with that of the original transformed cell line, Kato et al.(1986) showed that the rescued cellular DNA does indeed flank the integrated vector in vivo and is not generated by rearrangements during plasmid rescue. This procedure has been used to isolate the chicken thymidine kinase gene (Perucho et al., 1980), and to identify the repetitive elements flanking the integration site of exogenously introduced DNA (Anderson et al., 1982).

C.3. Cloning into Plasmid Vectors

In contrast to bacteriophage vectors, plasmids are rarely capable of accepting inserts as large as 20 kb and the overall efficiency of transformation into host bacteria is insufficient for the production of libraries that are representative of the entire mammalian genome (Hanahan, 1983). However, the laborious screening of phage libraries and the ultimate necessity to subclone into plasmids prompted the development of a new strategy whereby a small, but representative, plasmid library is generated which allows for direct cloning and rapid repetitive isolation of mutants from the same genetic locus (Nicholls et al., 1985).

In principle, the plasmid vector and genomic insert DNA are cleaved with the same restriction endonuclease, ligated together in vitro to form a recombinant plasmid, and transformed into bacteria. A recombinant may be distinguished from a recircularized plasmid by inserting the foreign DNA fragments into a vector gene that is easily

detected when inactivated (for example, insertion into an antibiotic resistance gene). Nicholls et al. (1985) demonstrated that by hydrolyzing the mammalian DNA with multiple restriction endonucleases that cleave outside the sequence of interest and isolating the specific fragment from an agarose gel, the specific fragment can be isolated from a small plasmid library containing fewer than 1000 recombinants.

To locate the DNA insert of interest, it is generally necessary to screen all transformants containing recombinant plasmids. This may be accomplished by first isolating the plasmids and then analyzing their restriction enzyme patterns or by identifying the transformant by in situ hybridizing of each colony with a radioactively labelled probe. A recombinant containing the junction site between unknown cellular sequences and a known genetic locus can be identified with a DNA probe specific for the genetic locus.

Restriction-modification systems in wild type bacterial strains can lead to the inactivation and eventual degradation of foreign DNA. Recently, studies found two site-specific systems, McrA and McrB, in E. coli K12 that restrict methylated cytosine at particular sequences (Raleigh and Wilson, 1986). Genomic DNA from mammals and plants contains substantial amounts of methylcytosine (Ehrlich and Wang, 1981) and therefore are targets for the Mcr system (Raleigh et al., 1988). To avoid the potential restriction

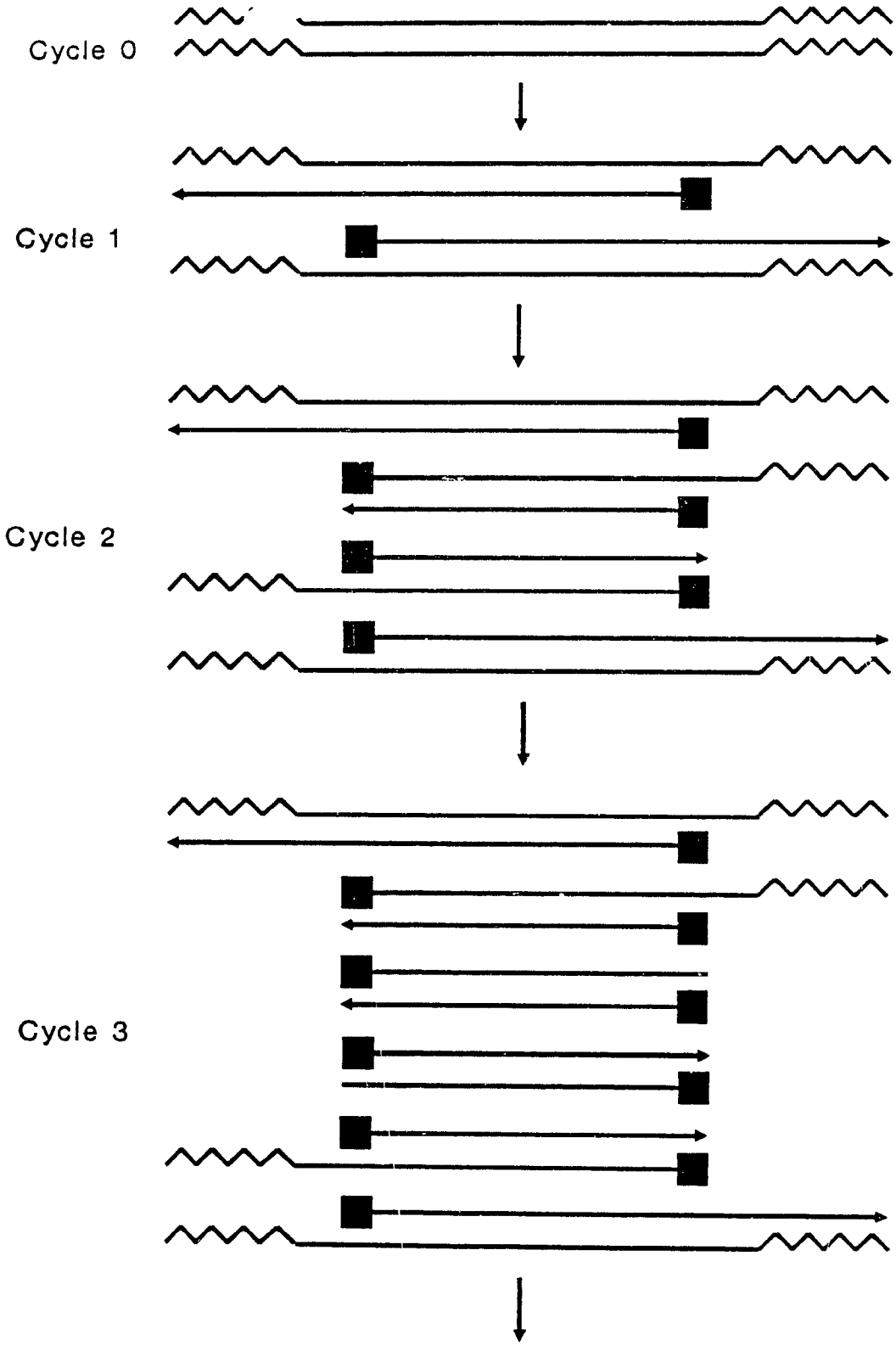
and degradation of mammalian chromosomal DNA in E. coli, it is suggested that bacterial strains deficient in McrA and McrB be utilized in transformation and phage screening procedures (Woodcock et al., 1988).

C.4. Polymerase Chain Reaction

The polymerase chain reaction (PCR) was first described as a method of amplifying a DNA segment between two regions of known sequence (Saiki et al., 1985)(Figure 1). Since 1985, the PCR system has been refined, simplified, and automated, rendering it a powerful tool in basic research, genetic diagnosis and forensic science.

Briefly, PCR employs two oligonucleotide primers that bracket the DNA sequence of interest : one is complementary to a sequence on one strand of DNA and the other to a downstream sequence on the other strand. Amplification of the intervening region is achieved by the repetition of three steps : 1. the double-stranded DNA target is denatured by incubation at 92-96^{o}C; 2. the temperature is lowered to 50-60^{o}C where an excess of oligonucleotide primers favours primer-template annealing over the reassociation of template DNA; 3. the temperature is raised to 72-75^{o}C and a thermostable DNA polymerase from Thermus aquaticus elongates from the 3'hydroxyl of the annealed primers, yielding single-stranded DNA copies of the intervening sequence. Initially, the polymerase products are of indefinite length, but the single-stranded DNA products (bounded by a primer at

FIGURE 1. **Schematic representation of the polymerase chain reaction.** The region to be amplified is represented by a straight line and the adjoining sequences are represented by a jagged line. Oligonucleotide primers used in the amplification are represented by black boxes and the direction of elongation is represented by arrows.


Cycle 0
Cycle 1
Cycle 2
Cycle 3
Cycles 4-30

amplification. Repeated cycles result in the preferential and exponential accumulation of double-stranded DNA fragments bounded at both ends by the oligonucleotide primers. Purification of a thermostable DNA polymerase from Thermus aquaticus (Taq polymerase) (Chien et al., 1976), simplified the process because the enzyme does not need to be replenished after each denaturation, as was the case with the thermolabile Klenow fragment of E. coli DNA polymerase I. Taq polymerase also substantially improved the overall performance of the reaction by increasing the specificity, yield, sensitivity, and length of targets that can be amplified (Saiki et al., 1988).

The misincorporation rate for Taq polymerase is estimated at $2x10^{-4}$ per nucleotide per cycle (Saiki et al., 1988). Therefore, if the sequence of the PCR products is required, direct sequencing via bulk product analysis is preferred to the sequencing of cloned amplification products, which examines only one sequence from a potentially heterogeneous population. Single-stranded DNA is generated for sequence analysis by amplifying with an excess of one oligonucleotide primer. Misincorporation errors would have to occur at the same position in the majority of target sequences, or during early cycles of PCR, in order to be detected by direct sequence analysis (Saiki et al., 1988). Furthermore, background due to amplification of non-specific DNA may be reduced by sequencing with a nested oligonucleo-

tide primer that hybridizes internally on the amplified DNA (Engelke et al., 1988).

a) Amplification of Unknown Sequences

PCR has been used for a variety of purposes (White et al., 1989) however, amplification is generally limited to regions of known DNA sequence because the technique depends on the design of two specific primers flanking the region of interest. Frequently, in the study of spontaneous mutations, one is interested in amplifying the segments flanking the target gene to identify sequence characteristics that might elucidate the mechanism responsible for the mutational event. Two methods have been developed to amplify specific unknown double-stranded DNA as long as sequence information is available at one extremity : inverse-PCR (IPCR) and single-specific primer-PCR (SSP-PCR).

The IPCR technique inverts the position of the known core sequence with that of the unknown flanking DNA. This is achieved by hydrolyzing genomic DNA with a restriction endonuclease that cleaves outside the segment of known sequence. The digested DNA is diluted and ligated under conditions that favour the formation of monomeric circles (Collins and Weissman, 1984). If the DNA is digested a second time with a restriction endonuclease that cleaves just once within the known sequence, the circle opens to form a linear molecule with the unknown sequences now

flanked by the known core sequence. These products can be used in conventional PCR to amplify the unknown sequences with oligonucleotide primers homologous to the known sequence, but situated in opposite directions (Triglia et al., 1988). Inverse PCR depends on the availability of restriction sites that will generate an appropriately sized fragment for amplification by current PCR techniques.

The SSP-PCR scheme involves the ligation of restricted chromosomal DNA fragments to linearized cloning vectors, such that the region to be amplified is flanked by the vector and a region of known chromosomal sequence. Amplification is carried out with a generic primer that hybridizes to the vector and a specific primer that matches the chromosomal sequence. Although a variety of fragments will ligate to the vector, the sequence specific primer ensures that only the desired fragment will be amplified exponentially (Shyamala and Ames, 1989).

SSP-PCR has two major advantages over IPCR. First, a wide range of restriction enzymes may be used to generate DNA fragments whose lengths lie within the limits of current PCR techniques. Second, SSP-PCR is ideal for genome walking in one direction by progressive amplification into an unknown region, whereas IPCR is better suited for amplifying only directly adjoining sequences.

D. THE KT CELL LINE

To study the molecular nature of spontaneous mutations in human cells, our laboratory has constructed a novel cell line containing an exogenously introduced viral gene. A plasmid carrying the herpes simplex virus type 1 (HSV-1) thymidine kinase (*tk*) gene (pSV2*neo*KT) was used to stably transfect human AK143 tk deficient (*tk*$^-$) cells. A clone containing a single integrated copy of the viral gene was isolated and expanded to form the KT cell line (Goring and DuBow, 1985)(Figure 2).

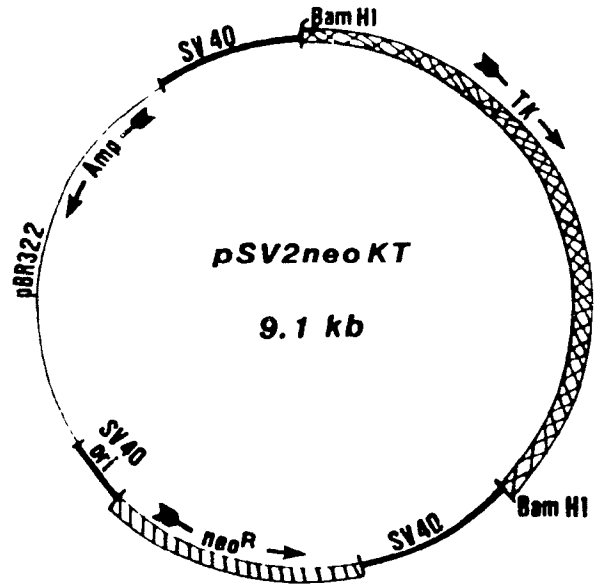
The neomycin (*neo*) gene that codes for G418 resistance in mammalian cells is located next to the *tk* gene on the chromosomally integrated pSV2*neo*KT in the KT cell line. Each gene is under the control of its own promoter. Therefore, it is possible to screen out gross chromosomal alterations and non-dysjunctions by maintaining a selective pressure with G418 for a functional *neo* gene during the selection for spontaneous mutations in the closely-linked *tk* gene.

Spontaneous mutations within the *tk* target gene can be selected with nucleotide analogs such as trifluorothymidine (TFT), acyclovir (ACV) and 9-(1,3-dihydroxy-2-propoxymethyl)-guanine (DHPG), specifically designed for drug therapy in herpes virus infections. These nucleotide analogs are selectively phosphorylated by the viral thymidine kinase and incorporated in the DNA. The analogs act as chain terminators for DNA elongation or cause base mispair-

FIGURE 2. **Structure of pSV2neoKT and restriction map of the integration site.*** The plasmid pSV2neoKT contains the pBR322 origin of replication and coding region for ampicillin resistance, the SV40 origin of replication, the neomycin resistance gene under constitutive control of the SV40 promoter and the HSV-1 thymidine kinase gene which is transcribed from its own promoter (Goring et al., 1987).

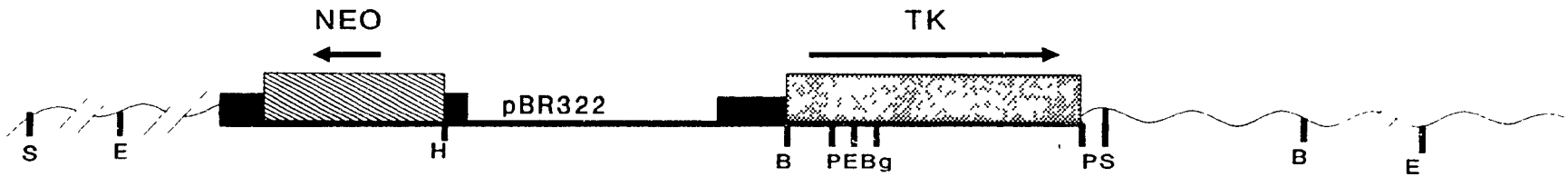
A restriction map of the integrated form of pSV2neoKT in the AK143tk$^-$ cells is shown at the bottom. Black boxes represent SV40 sequences and the broken line represents intervening sequences that are not shown. Restriction endonuclease sites are marked as follows : B: BamHI; Bg: BglII; E: EcoRI; H: HindIII; P: PvuII; and S: SacI.

* modified from Goring et al., 1987.



Transfected
into
143 TK⁻ Cells

Integrated pSV2neoKT



1.0 kb

cellular DNA

press a functional thymidine kinase. In the KT cell line, cells containing a mutation in the tk gene are selected by their inability to phosphorylate and incorporate the nucleotide analogs, thereby surviving the drug treatment. The frequency of spontaneous mutations isolated in the tk gene was~1 per 10^4 cells per generation with TFT and~1 per 10^3 cells per generation with ACV. Although a similar effect has been observed at the mouse and CHO aprt loci (Tischfield et al., 1982; Bradley and Letovanec, 1982), these frequencies are higher than expected for endogenous genes (10^{-6} to 10^{-8} per cell per generation). The enhanced mutation frequency of the HSV-1 tk gene may be due to the exogenous origin of the gene (Razzaque et al., 1984; Ashman and Davidson, 1984), or to the mutagenic effects of the drugs used in selection (Bilimoria et al., 1986). The higher frequency of ACV-resistant clones (ACV^R) versus TFT-resistant clones (TFT^R) may be due to the lower potency of the drug (Nahmias et al., 1981) : ~43% of ACV^R clones were found to be sensitive to treatment with TFT (Goring et al., 1987).

Structural analysis of mutants recovered from selection with either TFT or ACV revealed that most involved changes below the level of Southern blot detection (point mutations or alterations of <50 bp) and many of these reverted to a wild type phenotype at a high rate (10^{-3} - 10^{-2} per cell per generation). Theoretically, exposure to a combination of drugs would select for less frequently occuring mutational

events, such as large deletions and complex rearrangements that would render a cell resistant to both ACV and TFT. Two out of twelve mutants (16.7%) selected with a combination of ACV and TFT exhibited detectable rearrangements or deletions, whereas with single drug selections, 57 mutants had to be screened to find two with detectable alterations (3.5%). It was not possible to recover mutants resistant to DHPG because of a cytotoxic effect associated with this drug(Goring and DuBow, 1985). However, a single clone containing a complete deletion of the tk gene was isolated by briefly exposing the cells to DHPG (32 hours) prior to selection with a second drug, TFT (Goring et al., 1987).

The high reversion rates observed at the chromosomally located HSV-1 tk gene suggested that epigenetic events may be involved in the regulation of gene expression at this locus. To determine whether base-specific modifications, such as methylation, were involved in the modulation of tk expression, both the wild-type (tk^+) and mutant (tk^-) cell lines were digested with methylation-sensitive restriction endonucleases and analysed by Southern blotting. Restriction patterns generated with the methylation-sensitive restriction endonucleases AvaI and SmaI produced the fragments predicted according to the sequence. Furthermore, no detectable differences were observed between digests with the methylation sensitive HpaII and methylation insensitive MspI restriction enzymes, which cleave at the same DNA

recognition site 5'CCGG3'. These results suggest that extensive methylation is not involved in the modulation of tk expression, although site-specific modification of a sequence that is not recognized by these restriction endonucleases has been observed in one case using a similar system (Tasseron-deJong et al., 1989).

Structural analysis of two deletion mutants revealed that, although the right deletion termini located within host DNA, were not identical, their restriction maps suggested the presence of a potentially repetitive sequence organization (Goring et al., 1987).

E. RATIONALE FOR THESIS

The results obtained with the KT cell line suggest that a variety of mechanisms may be involved in the regulation of gene expression at the tk locus. The objective of this research was to determine the influence of host DNA sequence organization on the frequencies and types of mutations isolated in the tk gene. The work presented in the following thesis describes the methods developed in an attempt to clone the cellular sequences flanking the integrated pSV2neoKT.

The first set of experiments describe the plasmid rescue technique designed to recover the entire pSV2neoKT plasmid with some of the adjoining host sequences. A second set of experiments describe a number of cloning strategies used to isolate the junction sites of the integrated plasmid

as two separate clones. All of the recovered plasmids were smaller than expected, and a large number of recombinant plasmids contained deletions of the cloned genomic DNA. It is possible that restriction modification systems present in certain strains of E. coli played a role in generating the deletions observed. It is also possible that the nature of the sequence surrounding the integrated plasmid prevented the isolation of this segment of DNA in bacteria.

The third set of experiments attempted to amplify the cellular DNA adjacent to the tk gene using three different applications of the polymerase chain reaction. A modification of the anchored-PCR (A-PCR) (Frohman et al., 1988) technique used a primer matching the 3' tk sequences to generate single-stranded DNA extending into the adjoining host sequences. A homopolymer poly(dG) segment was then added to the 3' ends of the single-stranded products by terminal transferase, enabling the exponential amplification of the cellular sequences using a tk sequence-specific primer and a poly(dC) primer.

In the second technique, single-specific-primer PCR (SSP-PCR) (Shyamala and Ames, 1989), cellular DNA fragments are ligated to a linearized vector and amplified using a tk sequence-specific primer and a generic primer that matches the vector sequences.

In the third method, inverse PCR (IPCR) (Triglia et al., 1988), the tk sequences and the adjoining cellular DNA

are rearranged such that the unknown host DNA is flanked by fragments of the tk gene. Therefore, the host DNA may be amplified with two tk-specific primers.

Although these experiments did not succeed in isolating the tk gene and the adjoining chromosomal DNA, the results are valuable in directing further research. Several modifications are suggested to improve the techniques.

II. MATERIALS AND METHODS

A. The KT Cell Line

Construction of both the plasmid pSV2neoKT and the KT cell line have been described previously (Goring and DuBow, 1985). Briefly, pSV2neoKT was derived by cloning the HSV-1 tk gene into the single BamHI restriction site on the plasmid pSV2neo (Southern and Berg, 1982). The recombinant plasmid was transfected into human AK143tk^- cells and transfectants were selected with HAT medium and G418 (Littlefield, 1964). A colony with a single integrated copy of the plasmid was identified and expanded to form what has been called the KT cell line.

B. Enzymes and Conditions for DNA Hydrolysis

All restriction endonucleases were purchased from Bethesda Research Laboratories Inc. (BRL), New England Biolabs (NEB) or Pharmacia. DNA was hydrolyzed at 37°C in 75 mM NaCl, 6 mM 2-mercaptoethanol, 6 mM $MgCl_2$, 6 mM Tris-HCl (pH 7.5), and 125 μg/ml Bovine Serum Albumin (BSA). Cellular DNA was incubated for 4 hours to achieve complete digestion, whereas pure plasmid required only 1.5 hours.

C. Isolation of Plasmid DNA

The plasmids pSV2neoKT, pUC118, pUC119 and pBR322 (Table 1) were isolated according to the alkali lysis method described in Maniatis et al. (1982), and purified by centrifugation in cesium chloride-ethidium bromide density

TABLE 1. **Escherichia coli strains and plasmids**

Plasmids	Characteristics	Origin
pSV2neoKT	Amp^R, Neo^R, Tk^+	Goring and DuBow, 1985
pUC118, pUC119	Amp^R	Vieira and Messing, 1987
pBR322	Amp^R, Tet^R	Bolivar et al., 1977a, 1977b

Strain	Genotype	Origin
NM522	thi, rpsL, hsdS(r^-,m^+), supE44, Δ(lac-proAB), /F', traD36, proAB, $lacI^qZ$ ΔM15.	Gough and Murray, 1983
DH1	F^-, gyrA96, thi, hsd R^-M^+, supE44, recA1. $McrA^+$, $McrB^+$	Hanahan, 1983
HB101	F^-, hsdS20(r_B^-,m_B^-), recA13, leuB6, ara-14, proA2, lacY1, galK2, rpsL20, xyl-5, mtl-1, supE44, λ^-, $McrA^-$, $McrB^-$.	Boyer and Roulland-Dussoix, 1969
JM105	thi, rpsL, endA, sbcB15, hsdR4, Δ(lac-proAB), /F', traD36, proAB, $lacI^qZ$ ΔM15.	Yanisch-Perron et al., 1985
401-B ER1451	endA, gyrA96, thi, hsdR17, supE44, relA1, λ^-, Δ(lac-proAB), /F', traD36, proAB, $lacI^qZ$ ΔM15. $McrA^-$, $McrB^-$.	New England Biolabs, Inc.

was collected as described in Maniatis et al. (1982). Ethidium bromide was extracted from the preparation by mixing with an equal volume of 40X SSC (6 M NaCl, 0.6 M sodium citrate, pH 7.0) saturated with isopropyl alcohol. The phases were separated by centrifugation at 1500xg for 3 minutes at room temperature and retreating the aqueous phase with isopropanol until the pink colour disappeared. The aqueous phase was then dialyzed at 4°C against several changes of 1X TE (10 mM Tris-HCl, 1 mM EDTA, pH 8.0).

D. Preparation of Cellular DNA

The KT cell line was maintained in DMEM (Dulbecco's Modified Eagle's Medium) supplemented with 10% heat inactivated fetal bovine serum (Flow), HAT (0.1 mM hypoxanthine, 0.4 μM aminopterin, 16 μM thymidine) which selects for thymidine kinase activity (Littlefield, 1964), and G418 (Sigma)(400 μg/ml) which selects for the presence of a functional neo gene (Goring and DuBow, 1985). High molecular weight DNA was prepared according to Blin and Stafford (1976).

E. Southern Blotting of KT DNA

To characterize the host DNA surrounding the pSV2neoKT integration site, DNA form the KT cell line (KT DNA), was analyzed by Southern blotting (Southern, 1975) using the rapid hybridization kit developed by Amersham. Ten micro-

gram aliquots of KT DNA were hydrolyzed with a variety of restriction endonucleases and subjected to electrophoresis through a 0.75% horizontal agarose gel in 1X E buffer (40 mM Tris-base, 20 mM acetic acid, 1 mM EDTA). Genomic DNA was denatured by gently shaking the gel in a solution of 1.5 M NaCl, 0.5 M NaOH at room temperature for 30 minutes. The gel was rinsed in deionized water and neutralized by soaking in 1.5 M NaCl, 0.5 M Tris-HCl (pH 7.2) for two 15 minute periods at room temperature. The DNA was transferred from the gel to Hybond-N (Amersham) nylon filters by capillary blotting overnight in 20X SSC buffer. The DNA was then fixed to the filters by baking them for 2-4 hours at 80°C in a vacuum oven. Prehybridization, hybridization and probe labelling were performed according to the procedures outlined in Amersham's rapid hybridization kit.

The pSV2neoKT 3.6kb BamHI fragment containing the tk gene was isolated and labelled as a probe with α-^{32}P dGTP to a specific activity of ~$3X10^8$ cpm/μg. Briefly, pSV2neoKT was hydrolyzed with BamHI and subjected to electrophoresis through a 0.5% low melting point agarose gel (Sea Plaque - Mandel Scientific Company) in 1X E buffer. The gel was stained with ethidium bromide to visualize the DNA and the agarose segment containing the 3.6kb fragment was removed and incubated at 65°C for 15-20 minutes in 0.1 volume of 5M NaCl, 0.1M EDTA, pH 7.5. The DNA was extracted twice with 1X TE buffer-saturated phenol, once with ether, and then

precipitated by adding 2.5 volumes of 100% ethanol, incubating at -70°C for 1 hour and subjected to centrifugation at 10,000xg for 20 minutes at 4°C. The supernatant fluid was discarded and the DNA pellet was briefly dried in a vacuum dessicator and resuspended in 30 μl of 1X TE. The DNA was further purified through an Elutip-d column (Schleicher and Schuell, Inc.) according to the manufacturer's instructions but using 1.5 M NaCl during the elution step. The DNA was precipitated with ethanol a second time by adding 2.5 volumes ethanol, setting in a dry ice ethanol bath for 20 minutes and subjecting to centrifugation for 30 minutes at 10,000xg at 4°C. The fragment was then labelled with α-^{32}P dGTP (ICN) to a specific activity of 3 X 10^8 cpm/μg.

F. Plasmid Rescue and Transformation

KT genomic DNA was digested in a volume of 0.5 ml with SacI, a restriction endonuclease that cleaves outside the integrated pSV2neoKT sequences, producing a fragment which contains both the integrated plasmid and the flanking cellular sequences (Figure 2). The DNA was then purified through an Elutip-d column and precipitated with ethanol as described above. The DNA pellet was dried for 10 minutes under vacuum, resuspended in 1X TE and ligated under conditions that promote cyclization (Anderson et al., 1982). Each ligation reaction contained 1 μg of SacI hydrolyzed KT DNA, 30 units of T4 DNA ligase (BRL), and linker ligation buffer (0.06 M Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 15 mM DTT,

0.75 mM ATP, 1 mM spermidine and 50 μg/ml gelatin) in a total volume of 250 μl. The mixture was incubated overnight at 16°C and the reaction was inactivated the following day by incubation for 10 minutes at 65°C.

Competent Escherichia coli (E. coli) cells were prepared by a modification of the calcium chloride procedure developed by Mandel and Higa (1970). Approximately 0.5 ml of a 10 ml overnight culture of E. coli strain NM522 (Table 1) was used to inoculate 20 ml of LB broth (Miller, 1972). The cells were grown to mid-log phase (for NM522, OD_{550} ~0.3) and subjected to centrifugation at 4000xg for 7.5 minutes at 4°C. The supernatant fluid was discarded and the cells were resuspended in 12 ml of 10 mM $MgSO_4$, 2 mM Hepes (pH 7.5). The suspension was incubated on ice for 20 minutes, subjected to centrifugation as before, and the pellet was resuspended in 12 ml of 50 mM $CaCl_2$, 2 mM Hepes (pH 7.5). Following 25 minutes on ice, the suspension was subjected to centrifugation a third time and resuspended in 2 ml of 50 mM $CaCl_2$, 2 mM Hepes (pH 7.5), yielding competent cells.

For each transformation, half of the ligation reaction (125 μl or 0.5 μg of DNA) and 275 μl of 50 mM $CaCl_2$, 2 mM Hepes (pH 7.5) was added to a 0.4 ml aliquot of competent E. coli cells. The transformation mixture was chilled on ice for 25 minutes, incubated at 37°C for 3 minutes and then combined with 1.5 ml of LB broth and incubated at 32°C with

shaking for 1.5 hours. Approximately 0.2 ml of transformed cells were spread onto selective media containing either 50 μg/ml kanamycin (to select for the presence of the neo gene) or 50 μg/ml ampicillin (to select for the presence of the amp gene). Plates were incubated at 37°C for 12-16 hours and transformants were mastered the following day as patches in a 1 cm square grid pattern onto selective media containing either kanamycin or ampicillin.

To analyze the size of the plasmid DNA, 3/4 of each mastered patch of clones was transferred, using a sterile toothpick, to a 1.5 ml conical centrifuge tube containing 50 μl of "cracking" buffer (50 mM Tris-HCl (pH 8.0), 1.0% sodium dodecyl sulfate (SDS), 2 mM EDTA, 0.4 M sucrose, 0.01% bromophenol blue). The cells were allowed to lyse at room temperature for 15 minutes and the nucleic acids ware recovered in the supernatant fluid following a 15 minute centrifugation at 4°C and 10,000xg. The supernatant fluid was immediately subjected to electrophoresis through a 0.5% agarose gel and the DNA was visualized with ethidium bromide staining and photography under U.V. illumination (260nm) with Polaroid type 59 film.

G. Plasmid Cloning and Colony Hybridization

Several strategies were attempted to clone the cellular sequences flanking the integrated pSV2neoKT. A summary of the plasmids and restriction enzymes used is presented in Table 2, and a restriction map of the KT genomic fragments

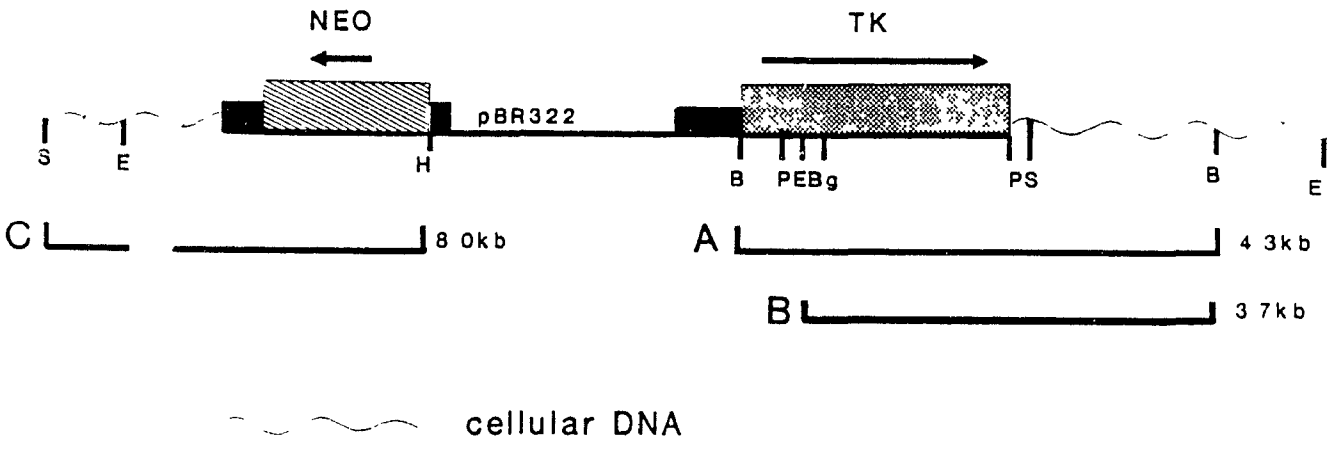
to be cloned is presented at the bottom.

Plasmid cloning vectors prepared by restriction endonuclease cleavage at a single site (Table 2, Strategy I) were dephosphorylated in order to reduce the frequency of self-ligation. The terminal 5' phosphates were removed by treatment with calf intestinal alkaline phosphatase (CIAP) (DuPont). Digested DNA was extracted once with phenol, twice with ether and then precipitated with 0.1 volumes of 2.5 M ammonium acetate and 2.5 volumes 100% ethanol. The DNA pellet was resuspended in 50 μl of 10 mM Tris-HCl (pH 10.0) and incubated at 37^{o}C for 30 minutes with 15 units of CIAP The reaction was then incubated at 65^{o}C for 5 minutes and the DNA was extracted, ethanol precipitated, and resuspended in 50 μl of 1X TE buffer (pH 7.6).

Plasmid cloning vectors prepared by cleavage at two sites with two restriction endonucleases (Table 2, Strategies 2-4) were size fractionated on a 0.5% agarose gel in order to isolate the larger vector band and thus prevent religation of the digested fragments to each other. If a low melting point (65^{o}C) agarose gel was used, the DNA was isolated as previously described and purified through an Elutip-d column. If agarose with a conventional melting point (100^{o}C) was used, the gel slice containing the larger vector band was removed, placed in an Elutrap device (Schleicher and Schuell) with 1X E buffer and subjected to a 100 volt current overnight. The DNA migrating out of the

TABLE 2. **Strategies designed to clone the genomic sequences flanking the pSV2neoKT integration site.** The diagram at the bottom of the table indicates the size and location of the genomic DNA fragments that were isolated for insertion into the plasmid vectors. The black boxes represent SV40 sequences located in the pSV2neoKT plasmid. Intervening sequences are represented by broken lines. Restriction endonuclease cleavage sites are represented by the following letters, Bg: BglII; B: BamHI; E: EcoRI; H: HindIII; P: PvuII; and S: SacI.

Strategy	Vector	Insert (see diagram below)	E. coli Strain
I	pUC119 hydrolyzed with BamHI	A	NM522
II	pUC119 hydrolyzed with EcoRI and BamHI	B	NM522 DH1 401-B ER1451
III	pBR322 hydrolyzed with EcoRI and BamHI	B	NM522 DH1 401-B ER1451
IV	pUC119 hydrolyzed with HindIII and SacI	C	HB101



1X E buffer. The current was reversed the following day for 20 seconds and the DNA was recovered form the chamber with a pasteur pipette and precipitated at -20°C overnight.

KT genomic DNA was prepared for cloning into the plasmid vectors by first hydrolyzing with the appropriate enzymes (Table 2) and then size fractionating the DNA on a 0.5% low melting point agarose gel. The size of the KT genomic DNA fragments was previously determined by Southern blot analysis (Figure 3 and D. Goring's M.Sc. thesis manuscript), and is shown in Figure 2. Both the right junction site, containing the tk gene and flanking cellular sequences, and the left juntion site, containing the neo gene and flanking cellular sequences, were isolated for cloning. A diffuse band corresponding to the molecular weight predicted for each fragment was removed, isolated and purified through an Elutip-d column as previously described.

Ligations of hydrolyzed KT genomic DNA to plasmid vectors were performed in a total volume of 20 μl at 16°C for 12-16 hours in linker ligation buffer and with 3 units of T4 DNA ligase. The insert:vector ratio was varied from 1:1 - 3:1 by weight, but the total amount of DNA per ligation did not exceed 1 μg.

Transformation of recombinant plasmids was performed as described for the plasmid rescue procedure with the following modifications : **1.** Strain 401-B ER1451 was grown to OD_{550} -0.6, **2.** The entire ligation reaction (20 μl) plus 180 μl

of 50 mM $CaCl_2$, 2 mM Hepes (pH 7.5) were added to a single 0.2 ml aliquot of competent E. coli cells and **3.** The cells transformed with recombinant pBR322-based plasmids were plated on selective media containing 50 μg/ml tetracycline.

To identify a transformant containing the cloned junction sequences, colonies were hybridized in situ with the α-^{32}P labelled 3.6 kb HSV-1 BamHI fragment from pSV2neoKT containing the tk gene. The probe was isolated as described for Southern blotting but was labelled using an alternate nick-translation procedure (Rigby et al., 1977). Approximately 1 μg of DNA was labelled with 4 μl each of α-^{32}P dGTP and α-^{32}P dATP (ICN, specific activity = 3000 Ci/mM) in nick translation buffer (50 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 0.05 μg/ml BSA) supplemented with 0.25 M $MgCl_2$ and 3 μl each of 250 μM dCTP and TTP (Sigma) in a total volume of 20 μl. Following a 10 minute incubation at 42°C, 5 units of DNA polymerase I (Pharmacia) and 3 μl of a 1/1000 dilution of DNAseI (1 mg DNAseI in 1 ml 10 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$) were added and the mixture was then incubated at 15°C for 1 hour. A single phenol:chloroform:isoamyl alcohol (24:24:1 by volume) extraction was performed. Following centrifugation, an equal volume of 50% glycerol was added to the aqueous phase.

Unincorporated nucleotides were separated from the labelled probe by passing the mixture through a Sephadex G-50 (Pharmacia) spin-column prepared in a disposable 1 ml syringe. The bottom of the 1ml disposable syringe was

plugged with a small amount of sterile glass wool. Using a pasteur pipette, the syringe was filled with a slurry of Sephadex G-50 and placed in a 15 ml disposable plastic tube. Following a 3 minute centrifugation (1600xg) in a bench-top centrifuge (in a swinging bucket rotor) at room temperature, more resin was added and recentrifuged until the volume of the packed column reached 0.9 ml. The DNA sample was added to the column, which was placed in a new disposable tube and subjected to a 3 minute centrifugation. Specific activity was determined by liquid scintillation counting of a 1 μl sample of the eluted DNA spotted on a 1-inch Whatman 3MM paper. The paper was dried for 10 minutes at 65^{o}C and counted in 4 ml of Econofluor scintillation fluid (DuPont). Specific activities ranged from $7X10^{6}$ - $7X10^{7}$ cpm/μg.

Colony hybridization was performed according to the procedure described in Maniatis et al., (1982). Briefly, nitrocellulose discs (Bio-Rad) were asymmetrically marked with a blue, Bic, ball point pen and placed on the surface of an agar petri plate supporting transformed colonies. The petri dish was similarly marked to maintain an orientation reference. The filters were peeled from the agar medium amd placed colony side up on Whatman 3MM paper saturated with denaturing solution (0.5 M NaOH, 1.5 M NaCl). After 10 minutes, the filters were peeled off and air dried for 15 minutes. They were then placed, colony side up, on Whatman 3MM paper saturated with neutralizing solution (0.5 M Tris-

HCl (pH 7.5), 1.5 M NaCl). After 5 minutes, the filters were peeled off and air dried a second time for 15 minutes. To remove cell debris and other lipids, the filters were dipped in chloroform, air dried for 10 minutes, placed on Whatman 3MM paper saturated with 0.3 M NaCl for 3 minutes and air dried for 30 minutes (Grunstein and Hogness, 1975). The filters were baked in a vacuum oven sandwiched between dry Whatman 3MM paper at 80°C for 2 hours. The petri dishes were returned to the 32°C incubator overnight in order to allow colony regrowth.

A maximum of seven filters were sealed in a single plastic freezer bag and incubated with shaking at 55°C for 6 hours in ~40 ml of prehybridization solution (5X Denhardt solution [100X Denhardt solution contains 2% w/v BSA, 2% w/v Ficoll 400, 2% w/v polyvinyl pyrrolidone], 5X SSC, 0.5% w/v SDS, 1 mM EDTA, 25 μM ATP, 5 μg/ml *E. coli* DNA). This solu tion was then replaced with ~30 ml of hybridization solution (1X Denhardt solution, 5X SSC, 0.5% w/v SDS, 1 mM EDTA, 25 μM ATP and 5 μg/ml *E. coli* DNA) to which the radioactively labelled 3.6 kb *Bam*HI probe (denatured at 100°C for 10 minutes) was added at a concentration of ~10 ng/ml (~10^7 cpm/μg). The filters were shaken overnight at 55°C and then washed for thre 30 minute periods room temperature in 0.5X SSC, 0.2% w/v SDS. After air drying for 30 minutes, the filters were placed on Whatman 3MM paper and their orientations were marked with radioactive ink (Maniatis *et al.*, 1982). The filters were then wrapped in Saran-Wrap and

exposed on Kodak XAR-5 X-ray film at -70°C with DuPont Cronex intensifying screens.

The approximate sizes of the recombinant plasmids were determined with the "cracking" procedure described for plasmid rescue analysis. To further characterize the size of each cloned fragment, plasmid DNA was isolated according to a modification of the alkali extraction procedure (Morelles, 1989) and hydrolyzed with the appropriate restriction enzymes to release the insert. The digests were then subjected to electrophoresis through a 0.5% agarose gel and the DNA was visualized with ethidium bromide staining and U.V. illumination. A photograph was taken with Polaroid type 57 film.

H. Polymerase Chain Reaction

The polymerase chain reaction (PCR) was carried out in a total volume of 50 μl, containing 0.5-1.0 μg of KT genomic DNA, 67 mM Tris-HCl (pH 8.8), 6.7 mM $MgCl_2$, 16.6 mM $(NH_4)_2SO_4$, 10 mM 2-mercaptoethanol, 10% dimethyl sulfoxide, 1.25 mM of each dNTP (Sigma), 25 pmoles of the specified oligonucleotide primers and 2.5 units of Taq polymerase (Perkin Elmer Cetus Corp.). Taq polymerase was added last, after the PCR mixture was incubated at 95°C to inactivate any enzyme inhibitors. The 7 specific oligonucleotide primers were prepared and purified by the Regional DNA Synthesis Laboratory at the University of Calgary (FS1, 5'-

GGACGAACTAAACCTGACTACG-3'; FS2, 5'-TATTGGTCACCACGGCCGAGTTT-3'; A, 5'-TGGCGGTGTCCCGGGAAGAAATAT-3'; A', 5'-AAGGCATGCC-CATTGTTATCTGGG-3'; B, 5'-TATCGTCGACGTACCCGAGCCGAT-3'; B', 5'-TTAGCCTCCCCCATCTCCCGGGCAAA-3'; FSP, 5'-CGCCAGGGTTTTCC-CAGTCACGAC-3'). The first six primers are specific for sequences within the tk gene. FSP is the M13 universal forward sequencing primer which matches sequences within the pUC118 and pUC119 vectors. The oligodeoxycytidilic acid $d(pC)_{22}$ was ordered from Sigma.

In order to reduce evaporation, the PCR reaction mixture was overlayed with ~50 μl of mineral oil. The amplification reactions were carried out for 30 cycles in the Hypercell Biologicals PTC-100 programmable thermal cycler. Each cycle consisted of three steps: **1.** denaturation for 1 minute at 94°C **2.** annealing of the primer to the template for 2 minutes at 55°C and **3.** elongation from the 3'OH of the primer for 3 minutes at 72°C. The final elongation step was extended to 8 minutes. To recover the entire reaction volume, 100 μl of chloroform was added to the reaction following amplification. The aqueous phase containing the DNA floated on the chloroform-oil mixture and was easily collected. To analyze the amplification products, 5-10 μl of the reaction mix was subjected to electrophoresis through either 0.75% or 1.0% agarose gels and visualized as described.

H.1. Anchored-PCR

A modification of the RACE (rapid amplification of cDNA ends) protocol of Frohman et al. (1988) was developed to attempt to amplify the unknown host sequences flanking the integrated tk gene. The first PCR reaction was carried out as previously described, except only one oligonucleotide primer was added, either FS1 or FS2. The entire reaction mix was recovered as described, and excess primer was removed with either a Sephacryl-300 (S-300) column (Pharmacia) or a Centricon-100 spin filter (Amicon Corp.). A 4.5 ml S-300 column was built in a 5 ml disposable pipette and packed at room temperature at 30 ml/cm^{2}hr. The column was washed with 6 ml of 4 M potassium acetate (pH 7.5), pretreated with 0.5 ml of a solution containing 25 μM ATP, 100 μg/ml BSA, and 50 μg calf thymus DNA and equilibrated with 40 ml of equilibration buffer (500 mM potassium acetate (pH 7.5), 10 mM Tris-HCl (pH 7.5), 1 mM EDTA). The DNA was diluted to 100 μl with diethyl pyrocarbonate-treated water, mixed with 100 μl of 50% glycerol and then loaded onto the column. Once the solution had entered the column, 250 μl of equilibration buffer was added and 10 μl fractions were collected. The Centricon-100 spin filter was used according to the manufacturer's instructions.

Both the column and the filter were tested with radioactively labelled DNA fragments to determine their ability to effectively separate 1.0 kb fragments from primers 24

nucleotides (nt) long. The plasmid pBR322 was hydrolyzed with BamHI and PstI (producing a 1.1 kb and a 3.2 kb fragment) and labelled with 4 μl of α-^{32}P dGTP in Klenow buffer (7 mM Tris-HCl (pH 7.5), 7 mM $MgCl_2$, 50 mM NaCl, 1 mM DTT), 5 μl each of dATP,dCTP and TTP and 5 units of Klenow DNA polymerase I (BRL). The reaction was incubated for 60 minutes at room temperature. The oligonucleotide primer FS1 (0.2 μg) was labelled at its 3' end with 3 μl of gamma-^{32}P-ATP (5,000 Ci/mM)(Amersham) in kinase buffer (70 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$) and 1 μl (5 units) of T4 polynucleotide kinase (Pharmacia). The reaction was incubated at 37^{o}C for 60 minutes and inactivated by incubation at 65^{o}C for 10 minutes. The pBR322 and FS1 labelled products were first mixed and then separated with either the S-300 column or the Centricon-100 filter. Samples from the S-300 fractions were subjected to electrophoresis through a 10% polyacrylamide gel which was immediately exposed to Kodak XAR-5 X-ray film for 20 minutes. Samples of the Centricon-100 retentate and flow-through were subjected to electrophoresis on both 12% and 3.5% polyacrylamide gels which were also immediately exposed to Kodak XAR-5 X-ray film for 20 minutes.

Once the excess primers were removed, the PCR products were concentrated to 10 μl using Speedvac centrifugation. Homopolymer dG tracts were added to the 3' ends of amplified DNA in a 20 μl total volume with 3 mM dGTP, 4 μl of 10X DNA tailing buffer (BRL) and 15 units of terminal deoxynucleotidyl transferase (TdT) (BRL) (Frohman et al., 1988). The

reaction was incubated for 10 minutes at 37°C and inactivated by incubation at 65°C for 15 minutes. To monitor the reaction, homopolymer dG tracts were added to 0.2 μg of FS1 labelled with gamma-^{32}P-ATP. The products were subjected to electrophoresis through a 12% polyacrylamide gel and immediately exposed to Kodak XAR-5 X-ray film for 20 minutes.

Following the addition of a homopolymer dG tract, the products were diluted 1/250 in 1X TE buffer and 10 μl was amplified exponentially by PCR using the poly dC oligonucleotide primer with either FS1 or FS2. The PCR products were recovered with a chloroform extraction, purified with a phenol extraction, precipitated with ethanol and resuspended in 40 μl of 1X TE buffer. Aliquots of 15 μl were hydrolyzed with either BamHI or SacI, and the products were subjected to electrophoresis on 1.3% agarose gels.

Agarose gels from various stages of the entire procedure were denatured in 0.5 M NaOH and 1.5 M NaCl at room temperature for two 15 minute periods, rinsed in deionized water, neutralised in 0.02 M NaOH and 1 M ammonium acetate at room temperature for two 15 minute periods, rinsed in deionized water and Southern blotted to Hybond-N filters with the liquid trapped in the gel itself as the only source of transfer buffer (Smith and Summers, 1980). The DNA was transferred to the filter over a 16 hour period and fixed under vacuum at 80°C for 3 hours. The filters were prehybridized overnight at 37°C in 6X SSC, 1X Denhardt solution,

0.5% w/v SDS, 100 μg/ml E. coli DNA and 0.05% w/v sodium pyrophosphate. Kinased oligonucleotide primers FS1 or FS2 were prepared as previously described, denatured for 10 minutes at 100°C, cooled on ice for 2 minutes and used as probes. Hybridization was carried out overnight at 37°C in 6X SSC, 1X Denhardt, 20 μg/ml E. coli DNA and 0.05% w/v sodium pyrophosphate. Filters were washed the following day for two 30 minute periods at 37°C in 6X SSC and 0.05% w/v sodium pyrophosphate, air dryed for 30 minutes and exposed to Agfa CURIX RP-1 X-ray film at -70°C with DuPont intensifying Cronex screens.

H.2. Single-specific-primer-PCR

Plasmid vector pUC118 was digested with EcoRI and SacI or EcoRI and BamHI and vector pUC119 was digested with PstI and SacI or PstI and BamHI. The large vector fragments were isolated from a low melting point agarose gel as previously described. KT and AK143 genomic DNA was digested with BglII and SacI or EcoRI and BamHI. Genomic DNA was ligated to the vectors at the SacI or BamHI sites in a total volume of 10 μl, containing 1X linker ligation buffer, 2 units of T4 DNA ligase, 0.2 μg of vector DNA and insert DNA at concentrations ranging from 0.6 μg/ml - 6 μg/ml. As a control, 5 ng of pSV2neoKT digested with BamHI and HindIII was added to the AK143-pUC118 BamHI ligation. Reactions were carried out at 16°C overnight and inactivated at 65°C for 10 minutes. Amplification with PCR was performed as previously described

with 2 μl of the ligation product as template and FS1 and FSP as the oligonucleotide primers. The products (15 μl) were subjected to electrophoresis through both 1% agarose gels and 8% polyacrylamide gels and visualized with ethidium bromide. Southern blots of agarose gels were performed as described for previous PCR gels and hybridized with kinased primer FS2.

H.3. Inverse-PCR

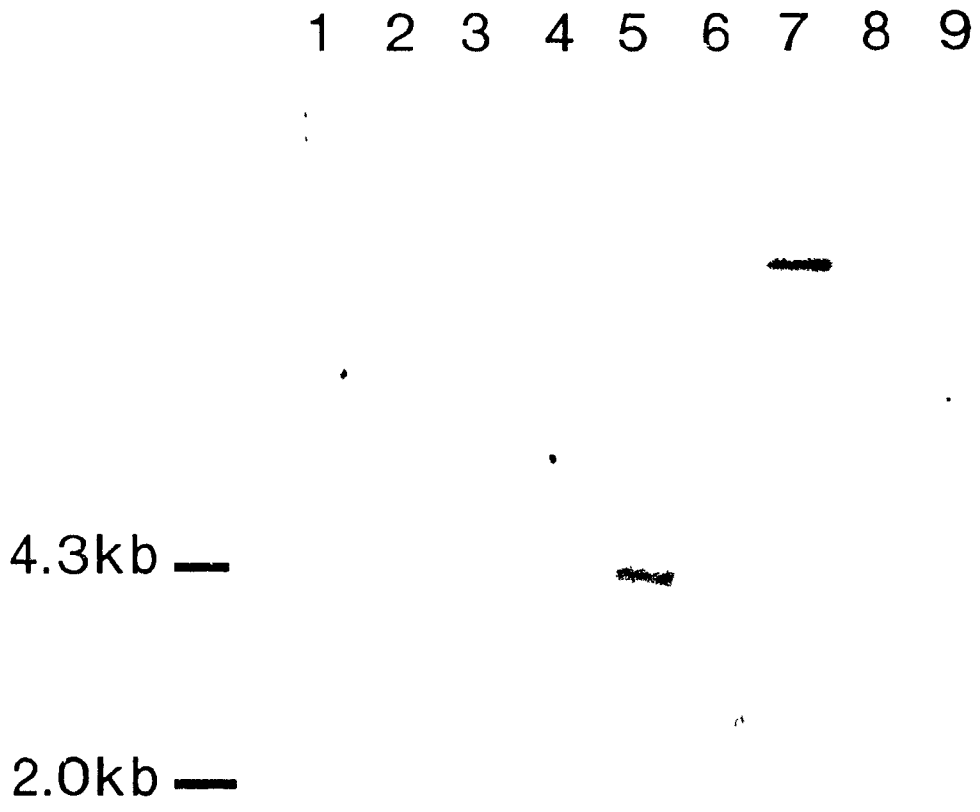
KT genomic DNA was digested with BamHI and 1.5 μg of DNA was ligated under conditions that promote cyclization - that is, in a total volume of 650 μl with linker ligation buffer and 30 units of T4 DNA ligase (Triglia et al, 1989). The ligation reaction was carried out at 16°C overnight and inactivated at 65°C for 10 minutes. A 50 μl aliquot was incubated at 94°C for either 5, 10, 20, or 30 minutes, chilled on ice for 5 minutes and then 25 μl was used as template in a PCR reaction with FS1 and A' as the oligonucleotide primers. The products (10 μl) were subjected to electrophoresis through a 0.5% agarose gel, Southern blotted as described for previous PCR gels and hybridized with kinased primer FS2 as the probe.

III. RESULTS

A. Southern Blot Analysis of the KT Cell Line

An initial restriction map of the pSV2neoKT integration site was previously prepared (Goring and DuBow, 1985; see Figure 2). During integration, a short stretch of plasmid sequence was lost at the 3' end of the 3.6 kb BamHI fragment. Southern blot analysis of PvuII cleaved KT DNA identified the expected 2.02 kb fragment, indicating that the entire tk coding region was intact. To further characterize the cellular host DNA flanking the tk gene, DNA from the KT cell line was hydrolyzed with a variety of restriction enzymes and analyzed by Southern blotting (Figure 3). These results determined the positions of both the SacI and EcoRI sites in the 3' flanking DNA. The restriction endonuclease SacI does not cleave pSV2neoKT sequences but Southern blot analysis of DNA from the KT cell line (Figure 3, lane 3: BglII/SacI; lane 6: BamHI/SacI; lane 8: EcoRI/SacI), places the SacI restriction site next to the tk PvuII site, suggesting that the length of plasmid sequence retained after PvuII is between 50-300 bp. Comparison of the single EcoRI and double EcoRI/SacI digests (Figure 3, lanes 8 and 9, respectively) indicates that the EcoRI site at nucleotide 45570 in the tk gene sequence (McGeoch et al., 1988) is missing and the next one is ~6 kb into the host sequences (Figure 2).

FIGURE 3. **Southern blot analysis of the pSV2neoKT integration site.** DNA from the KT cell line was digested with the following restriction endonucleases : lanes 1. PvuII 2. BglII 3.BglII and SacI 4.BamHI and EcoRI 5. BamHI 6. BamHI and SacI 7. SacI 8. EcoRI and SacI and 9. EcoRI. The 3.6kb BamHI fragment of pSV2neoKT, containing the tk gene, was labelled with α-^{32}P and used as a probe.


1 2 3 4 5 6 7 8 9
4.3kb
2.0kb

B. Plasmid Rescue

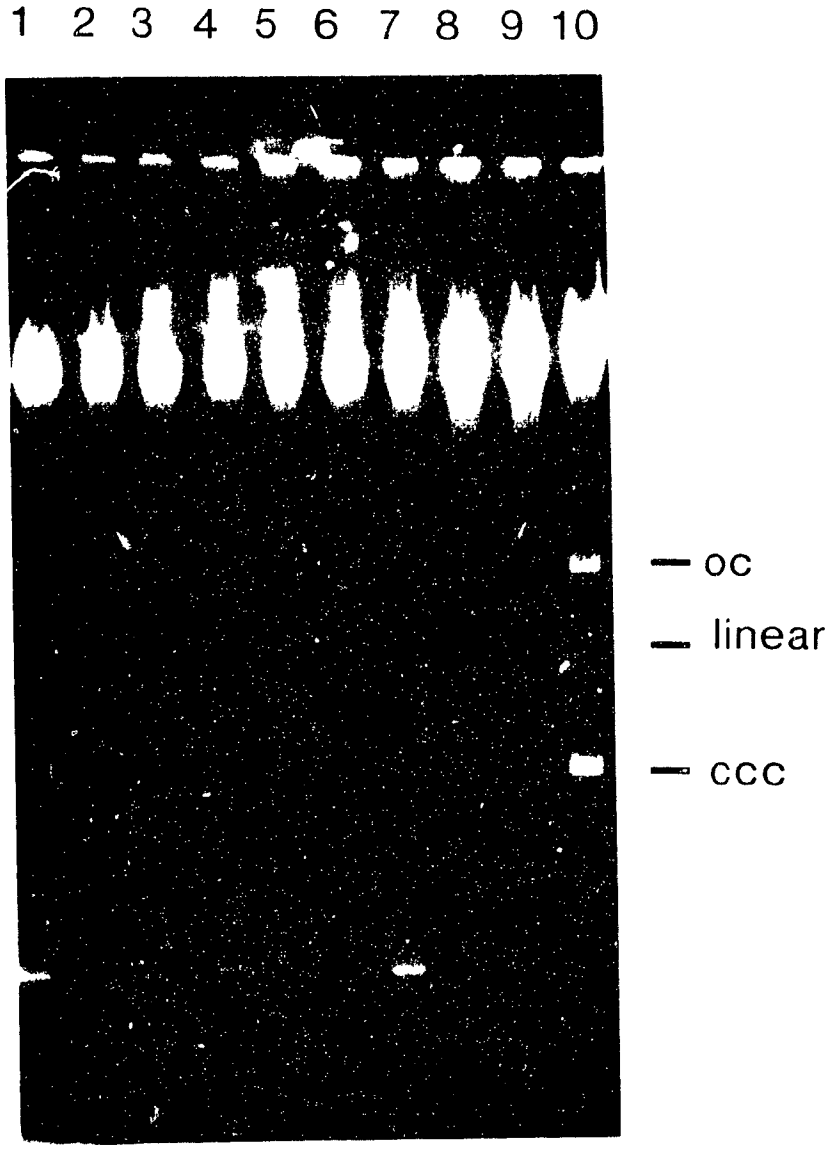
Six separate rescue experiments were performed with SacI cleaved KT genomic DNA in an effort to recover the integrated pSV2neoKT and its flanking host sequences as an autonomously replicating plasmid. When transformed into E. coli, the desired SacI fragment was expected to impart both kanamycin and ampicillin resistance through the neo and bla genes respectively. Although transformants were never isolated from selective media containing kanamycin, each transformation yielded 3 to 4 ampicillin-resistant colonies. These colonies did not grow on kanamycin selective media, suggesting that the neomycin resistance gene was either mutated or deleted during the recovery process. To determine the size of the rescued plasmids, the cells were "cracked", the nucleic acids were isolated from the supernatant fluid and subjected to electrophoresis through an agarose gel. Figure 4 represents 9 of the 20 transformants analyzed in this manner. All of the rescued plasmids migrated to a lower molecular weight position than that of the original plasmid pSV2neoKT (Figure 4, compare lanes 1-9 with lane 10), indicating that sequences were deleted from the SacI fragment during the rescue procedure.

C. Plasmid Cloning

A summary of the strategies employed to clone the integration junction sites on separate fragments are pre-

FIGURE 4. **Plasmids isolated from ampicillin resistant transformants obtained with the plasmid rescue procedure.** Each plasmid represented in lanes 1-9 was isolated from a separate agar petri plate following transformation of SacI digested and self-ligated KT genomic DNA. Lane 10: uncut pSV2neoKT DNA, which was transformed and isolated as described for the plasmids in lanes 1-9.

The open circle (OC), linear and covalently closed circular (CCC) forms of the uncut pSV2neoKT plasmid are marked on the right hand side of the photograph. The DNA present in every lane above the OC form is E. coli genomic DNA.


1 2 3 4 5 6 7 8 9 10
oc
linear
ccc

sented in Table 2. Initially (Strategy I), a single restriction endonuclease, BamHI, was used to clone the right junction site spanning the tk gene and its flanking cellular sequences into the plasmid pUC119. However, colony hybridization analysis of recombinant plasmids from 14 separate trials failed to identify the clone of interest (data not shown).

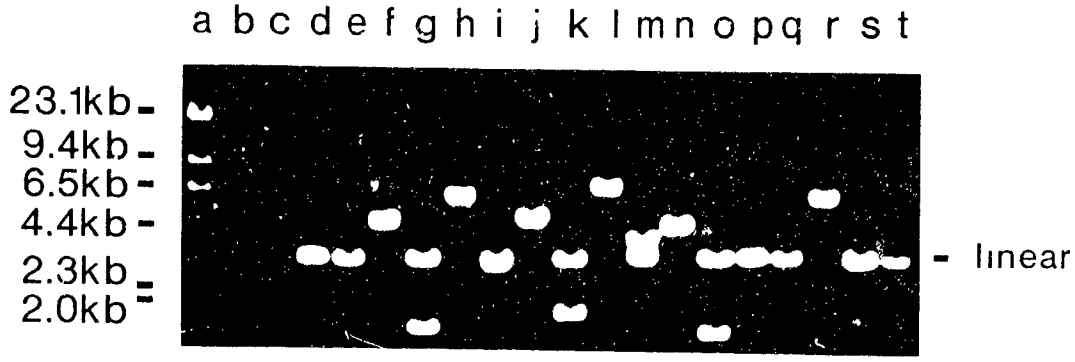
Subsequent experiments (Table 2, Strategies II-IV) used two restriction endonucleases to produce appropriately sized KT genomic DNA fragments for cloning. The use of two enzymes served to facilitate restriction identification of the clones by predetermining the orientation of the inserts. In addition, the influence of cell type was explored by transforming the recombinants into different strain backgrounds. In Strategy II (table 2), cloning with EcoRI and BamHI cleaved DNA also failed to identify the clone of interest.

To determine the structure of the recombinant plasmids, transformants were randomly chosen and screened with the cracking procedure. A representative group was chosen for mini-prep and restriction endonuclease analysis (Figure 5, panel A). Digestion with EcoRI and BamHI (Figure 5, panel A, lanes c, e, g, i, k, m, o, q, s) released the inserted DNA fragment from the original linear vector. All of the E. coli DH1 transformants analyzed contained inserts that were smaller in molecular weight than the size-selected genomic DNA (3.5 to 4.0 kb) (Figure 5, panel A, lanes b-g). The same was observed in approximately half of both NM522 and

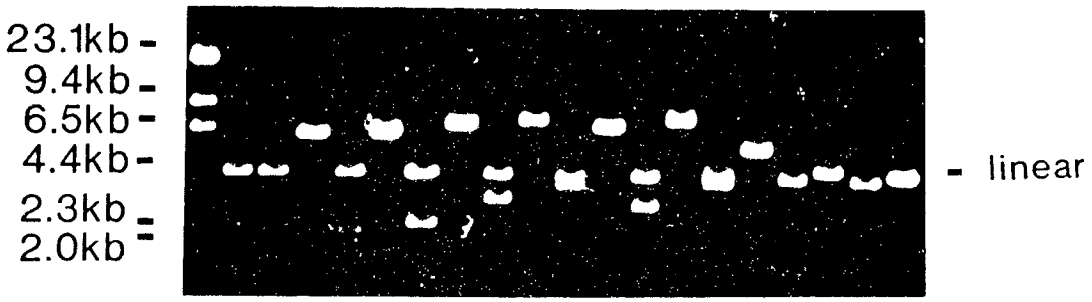
FIGURE 5. **Restriction enzyme analysis of recombinant plasmids containing cloned genomic DNA from the KT cell line.** Panel A: Genomic DNA from the KT cell line was cloned into the pUC119 vector. Lane a: marker, lambda DNA cleaved with HindIII. Lanes b, d, f, h, j, l, n, p, and r: recombinant plasmids hydrolyzed with BamHI. Lanes c, e, g, i, k, m, o, q, and s: recombinant plasmids hydrolyzed with EcoRI and BamHI. Lane t: pUC119 hydrolyzed with BamHI. Lanes b and c represent the same recombinant plasmid as do lanes d and e, f and g, h and i, j and k, l and m, n and o, p and q, and r and s. Recombinants were transformed either into E. coli strain DH1 (lanes b-g), NM522 (lanes f-i) or 401-B ER1451 (lanes j-s).

Panel B: same as for panel A except the cloning vector used is pBR322; lane t: pBR322 hydrolyzed with BamHI; recombinants were transformed either into DH1 (lanes b-e), NM522 (lanes f-i), or 401-B ER1451 (lanes j-s).

A


a b c d e f g h i j k l m n o p q r s t
23.1kb
9.4kb
6.5kb
4.4kb
2.3kb
2.0kb
linear

B


a b c d e f g h i j k l m n o p q r s t
23.1kb
9.4kb
6.5kb
4.4kb
2.3kb
2.0kb
linear

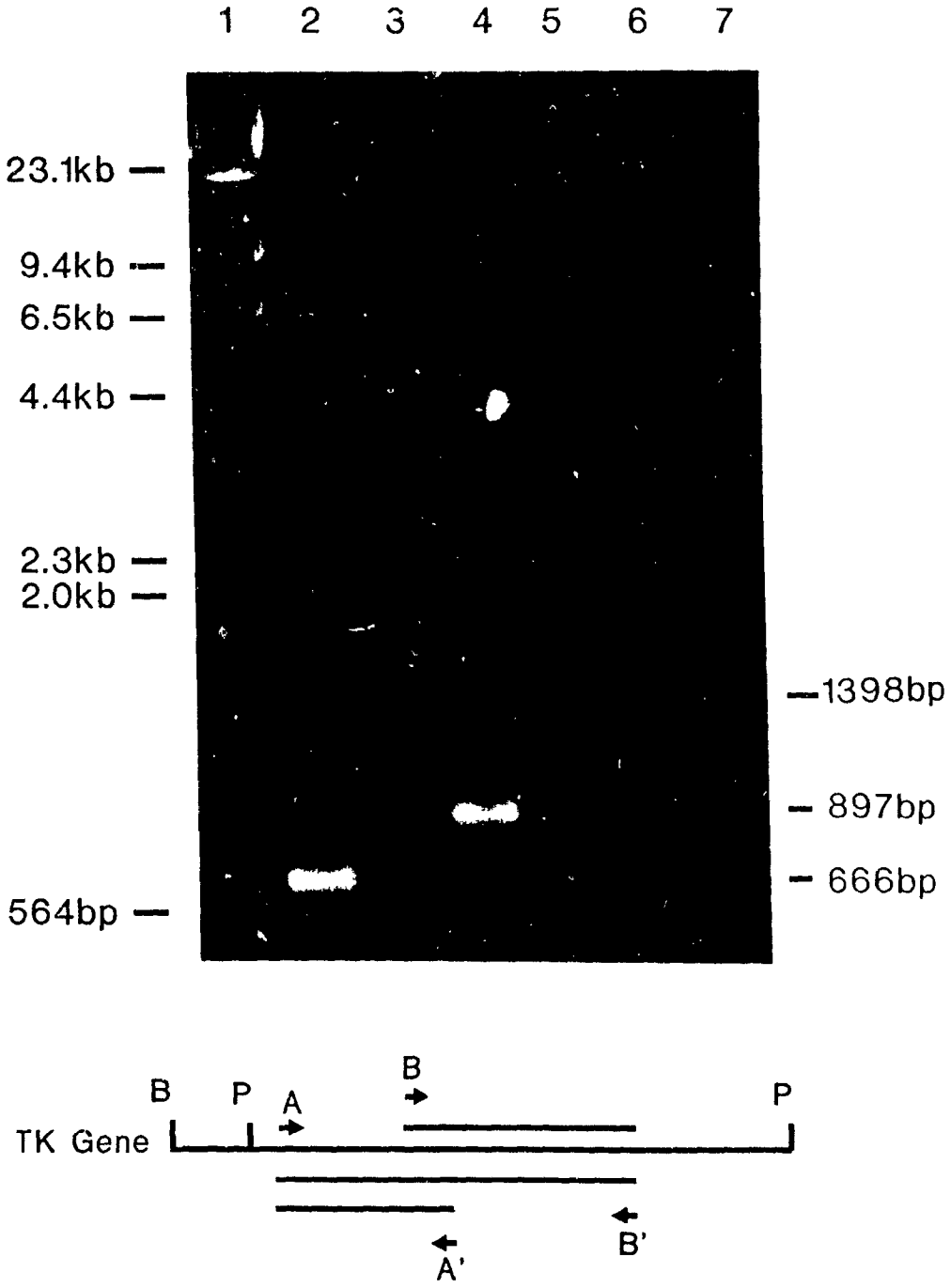
401-B ER1451 transformants - the other half contained inserts of the expected size (Figure 5, panel A, lanes f-s; other data not snown). Similar results were obtained with strategy III by cloning into pBR322 hydrolyzed with EcoRI and BamHI which resulted in inactivation of the tet gene and identification of recombinants by their acquired sensitivity to tetracycline (Maniatis et al., 1982) (figure 5, panel B). Cloning of the left junction site containing the neo gene (Strategy IV) did not produce any kanamycin resistant transformants and those expressing a resistance to ampicillin also contained deletions in the inserted sequences (data not shown).

D. Polymerase Chain Reaction

The ultimate objective in employing the polymerase chain reaction was to amplify the unknown genomic sequences flanking the integrated pSV2neoKT. However, it was first necessary to determine the optimal conditions required for PCR amplification of KT genomic DNA. The oligonucleotide primer pairs A-A', B-B' and A-B' were shown to specifically anneal to the integrated tk sequences and to amplify the expected fragments (666 bp, 897 bp, and 1398 bp) under the conditions specified in materials and methods (Figure 6). Amplification of AK143 DNA (the KT parental cell line which lacks the integrated pSV2neoKT), was used as a negative control (Figure 6, lanes 2, 4, and 6). In subsequent PCR

FIGURE 6. **Amplification of sequences within the tk gene using the polymerase chain reaction.** Lane 1: lambda DNA hydrolyzed with HindIII. Lanes 2, 4, and 6: KT genomic DNA. Lanes 3, 5, and 7: AK143 tk^- genomic DNA. Lanes 2 and 3: amplification of a 666bp sequence with primers A and A'. Lanes 4 and 5: amplification of a 897bp sequence with primers B and B'. Lanes 6 and 7: amplification of a 1398bp sequence with primers A and B'.

A schematic representation of the region amplified is represented at the bottom. Primers are represented by arrows. Restriction endonucleases are represented by B: BamHI and P: PvuII.


1
2
3
4
5
6
7
23.1kb
9.4kb
6.5kb
4.4kb
2.3kb
2.0kb
564bp
1398bp
897bp
666bp
B
P
A
B
P
TK Gene
A'
B'

experiments, either A-A' or B-B' amplification of KT genomic DNA was included as a positive control to verify each PCR reaction.

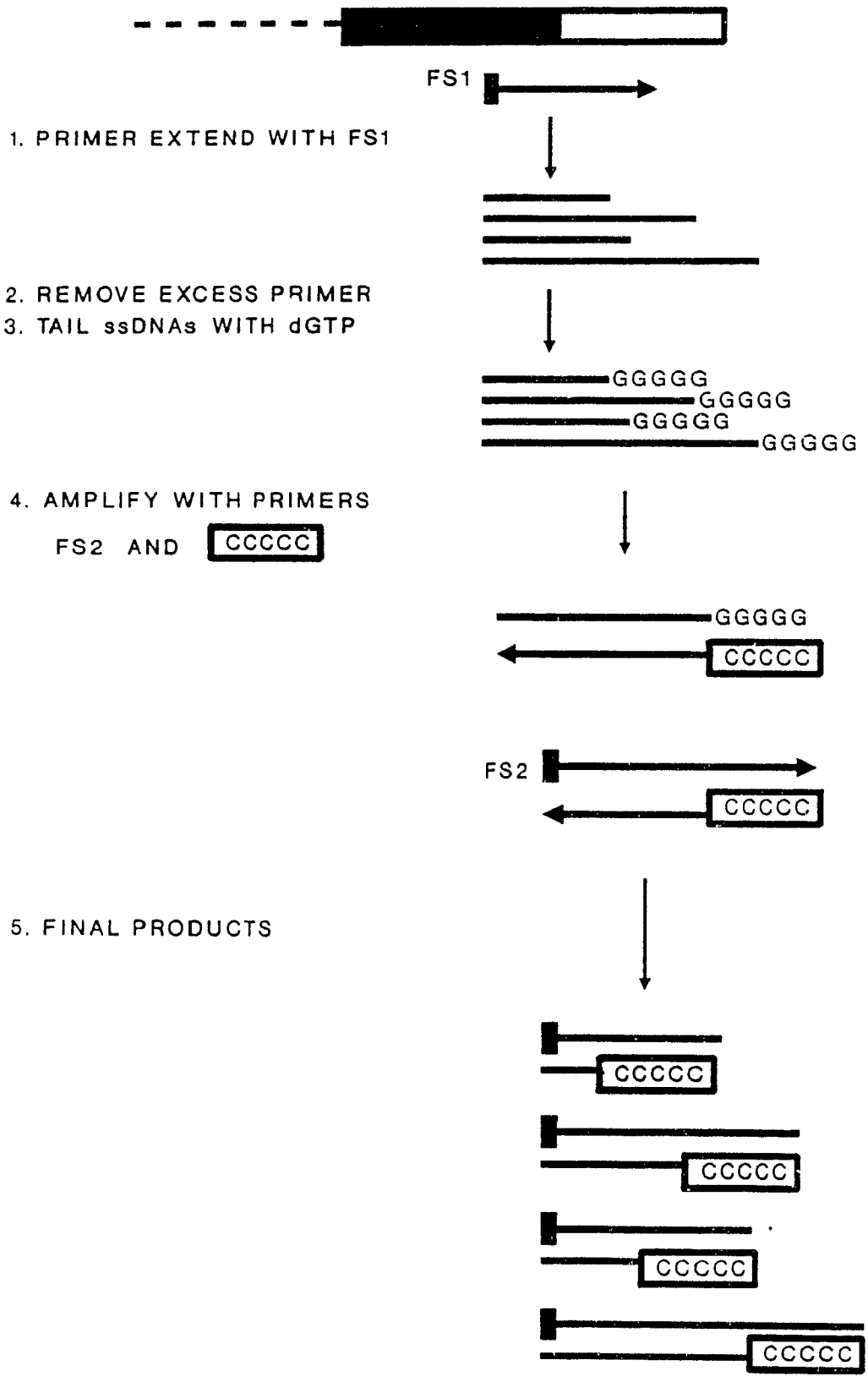
D.1. Anchored-PCR

Conventional PCR techniques employ two sequence-specific primers to bracket the stretch of DNA targeted for amplification. This requirement for sequence information at both extremities has limited the application of PCR to regions of known sequence. In a modification of PCR, Frohman et al. (1988) used the polyA sequence at the 3' end of mRNAs to design a poly(dT) oligonucleotide primer which was then used with reverse transcriptase to amplify cDNA between the unknown 3' end and a known sequence within the mRNA. Unknown sequences at the 5' end of the mRNA were similarly amplified by first producing a cDNA with reverse transcriptase using a primer matching a sequence within the mRNA and then adding a homopolymer dATP sequence to the 3' end of the cDNA.

One of the procedures we designed to amplify the unknown sequences flanking the tk gene is outlined in Figure 7 and was based on the protocol developed by Frohman et al. (1988). In the first step, called primer extension, PCR is performed with a single primer - a sequence-specific primer matching a region of known DNA within the tk gene. Elongation with Taq polymerase is expected to produce ssDNAs that

FIGURE 7. **Schematic representation of the anchored-PCR technique used to amplify the KT genomic sequences adjoining the integrated tk gene.** The tk gene is represented by the black box; the adjoining genomic sequences are represented by the open box; pBR322 sequences adjacent to the tk gene are represented by a dashed line; direction of elongation is depicted by the arrows.

In the first step, primer extension, the FS1 primer (▮) is used to amplify ssDNA (straight lines) spanning the tk gene and flanking genomic sequences. A poly(dG) "tail" is added in the third step, and the products are then amplified in the fourth step with the FS2 (▮) and poly(dC) (CCCCC) primers. Variations of this method utilized either the FS1 primer in both the first and fourth steps, or the FS2 primer in both the first and fourth steps.


FS1
1. PRIMER EXTEND WITH FS1
2. REMOVE EXCESS PRIMER
3. TAIL ssDNAs WITH dGTP
GGGGG
GGGGG
GGGGG
GGGGG
4. AMPLIFY WITH PRIMERS
FS2 AND CCCCC
GGGGG
CCCCC
FS2
CCCCC
5. FINAL PRODUCTS
CCCCC
CCCCC
CCCCC
CCCCC

extend from the 3' end of the primer into the adjacent cellular sequences. Because the region amplified is not bracketed by a second primer, the elongation products will not accumulate exponentially and they will not be of uniform length. The diffuse bands observed in lanes 2 and 3 of Figure 8 represent ssDNA products obtained from primer extension of KT DNA with the FS1 and FS2 primers, respectively. It is possible that some of these products are due to non-specific binding of the primers and do not contain the desired sequence.

It is critical at this point to remove the excess primers and dNTPs from the reaction mixture to prevent their interference in subsequent procedures. Initially, a 4 ml Sephacryl-300 column was constructed and assayed by testing its ability to separate radioactively labelled 1 kb fragments from smaller 24 nt primers. Fragments of ~1 kb were chosen because this is the minimal length we wished to recover from the PCR amplification. Fractions analyzed on a 10% polyacrylamide gel (Figure 9) showed that the majority of the 1 kb population could be recovered in fractions 18-33 without primer contamination, since the latter eluted from the column only after fraction 36. Although the S-300 column was successful, it required the incorporation of radioactive nucleotides in the PCR reactions in order to monitor the progress of the elution. Furthermore, the fractions could not be routinely analyzed by gel electropho-

FIGURE 8. **Primer-extension of KT genomic DNA using either the FS1 or FS2 primers.** Lane 1: positive control, amplification of tk sequences with primers B and B'. Lane 2: primer extension with the FS1 primer. Lane 3: primer extension with the FS2 primer. Lanes 4: negative control, PCR reaction as for lane 2 but with no template DNA. Lane 5: negative control, PCR as for lane 3 but with no template DNA. Lane 6: marker, lambda DNA hydrolyzed with HindIII. The 4.4 kb and 564 bp bands are not visible in this photograph.

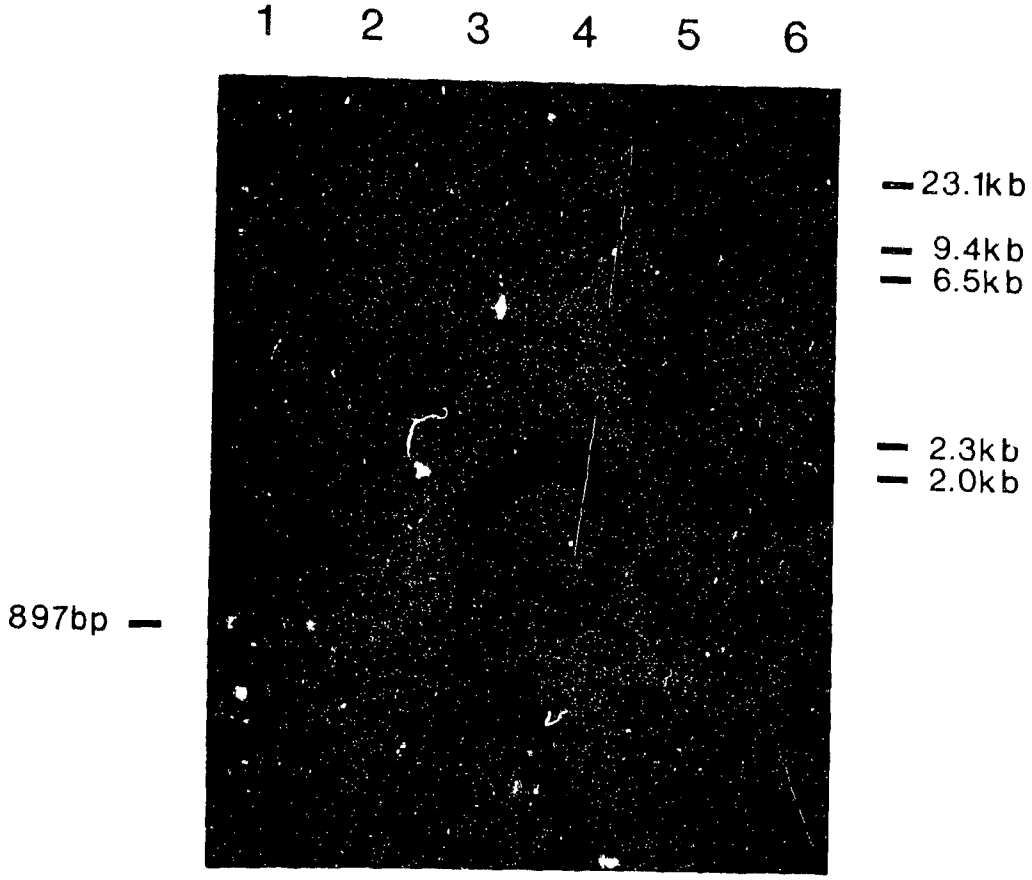
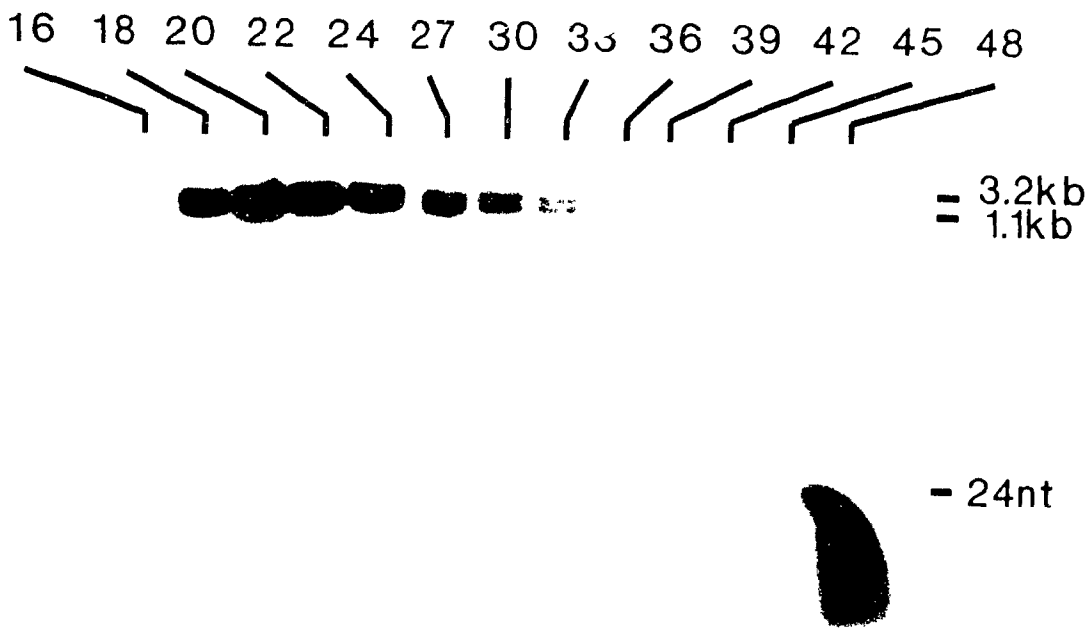

1
2
3
4
5
6
23.1kb
9.4kb
6.5kb
2.3kb
2.0kb
897bp

FIGURE 9. **Separation of 1.1 kb fragments from 24 nt primers using a 4 ml Sephacryl-300 column.** The 3.4 kb and 1.1 kb fragments were generated by hydrolyzing pBR322 with BamHI and PstI. The 24 nt primer is FS1.


FRACTIONS
16 18 20 22 24 27 30 36 39 42 45 48
= 3.2kb
= 1.1kb
- 24nt

resis because of the insufficient quantity of ssDNA products. The S-300 column was replaced with a simpler, faster, and equally efficient Centricon-100 spin column. Its ability to separate 1 kb fragments from 24 nt primers is shown in Figure 10. Only the larger DNA fragments (>1.1 kb) were recovered in the retentate (Figure 10,panel A. lane 2 and panel B, lane 1) whereas the 24 nt primers were found in the flow through (Figure 10, panel A, lane 1 and panel B, lane 2).

Once the primer-extended products were separated from excess primers and dNTPs, a poly(dG) "tail" was added to the 3' end of the ssDNA by using dGTP and terminal deoxynucleotidyl transferase (Figure 11). The tailing reaction was monitored by treating a radioactively labelled oligonucleotide primer under the same conditions. Evidence for the addition of a homopolymer "tail" to the 3' end of the primer was indicated by its increase in size from 24 nt (Figure 11, lane 1) to ~45 nucleotides (Figure 11, lane 2). The length of the homopolymer "tail" was expected to increase over time but the products of a 2, 5, 10 and 15 minute reactions exhibited similar sizes (Figure 11, lanes 2-5). These results indicate that a 10 minute reaction is sufficient to add a 20 to 25 nucleotide poly(dG) tail to the 3' ends of ssDNA (Frohman et al., 1988).

In the final step, a sequence-specific primer and a poly(dC) primer are used in a PCR reaction to exponentially

FIGURE 10. **Separation of 1.1 kb fragments from 24 nt primers using a Centricon-100 spin column.** The 3.2 kb and 1.1 kb fragments were generated by hydrolyzing pBR322 with BamHI and PstI. Panel A: 12% polyacrylamide gel. Lane 1: Centricon-100 flow through sample. Lane 2: Centricon-100 retentate sample.

Panel B: 3.5% polyacrylamide gel. Lane 1: Centricon-100 retentate sample. Lane 2: Centricon-100 flow through sample. Lane 3: FS1 primer.

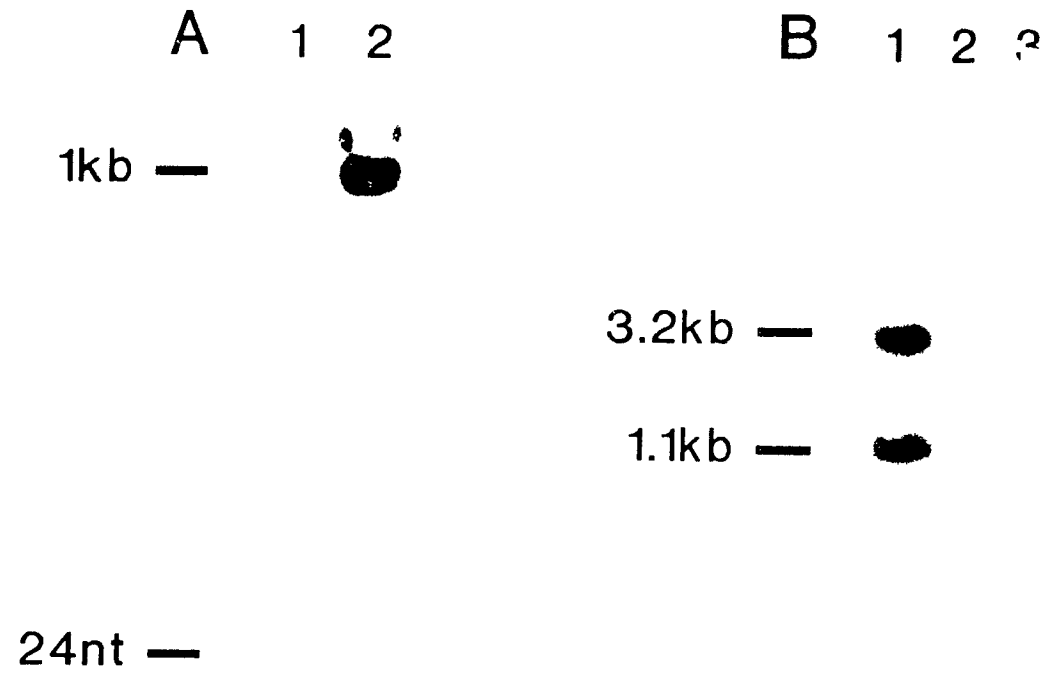
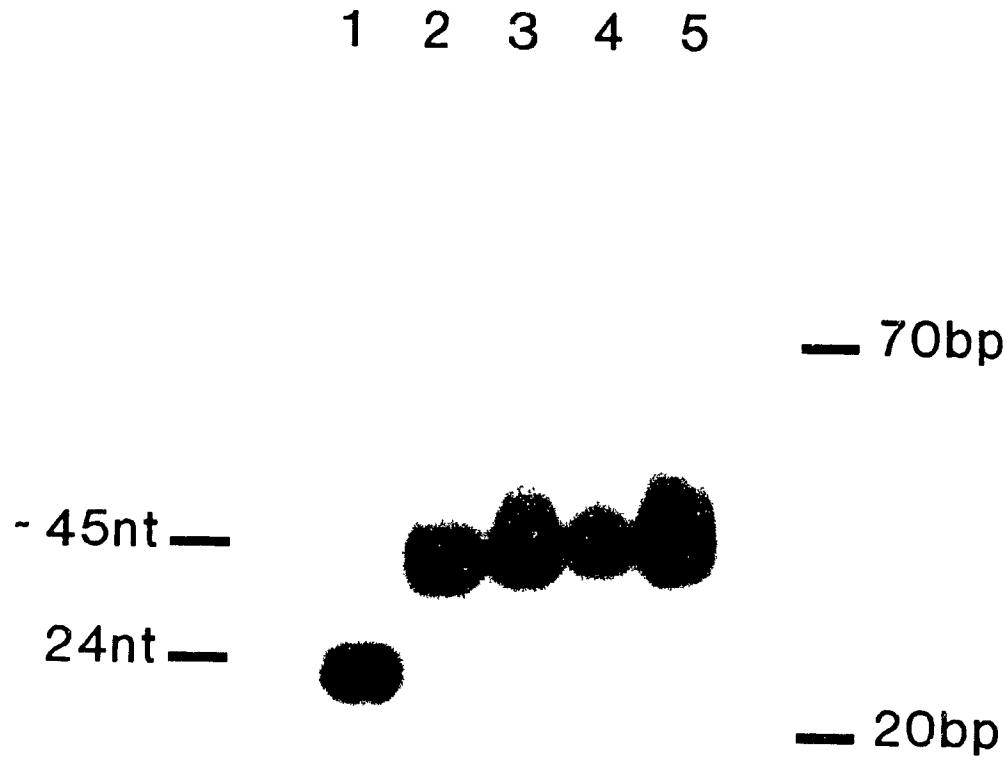

A
1
2
1kb
B
1
2
3
3.2kb
1.1kb
24nt

FIGURE 11. **Addition of dGTP nucleotides tc the 3' end of the 24 nucleotide primer FS1.** Lane 1: FS1. Lanes 2-5: products recovered following a 2, 5, 10, and 15 minute incubation time of FS1 with dGTPs and terminal deoxynucleotidyl transferase. Positions corresponding to the migration of 20 bp and 70 bp fragments are shown to the right of the photograph and the size of the bands are shown to the left of the photograph.


1 2 3 4 5
70bp
45nt
24nt
20bp

amplify the poly(dG) "tailed" fragments (Figure 7). If the initial primer extension was performed with FS1, the second PCR reaction could be carried out with the sequence-specific primer FS2. The FS2 primer binds to the tk sequences previously amplified with FS1 and thus increases the specificity and efficiency of amplification, whereas FS1 binds to all amplified ssDNAs present, including any generated through non-specific annealing during primer extension. The results of one such trial are presented in Figure 12, lane 2. As expected, the products varied in length and produced a visible smear on an agarose gel. Certain regions appeared brighter and may represent clusters of fragments that were preferentially amplified either because of their length, or because of termination signals or secondary structues present in the DNA flanking the tk gene that may limit the extent of Taq polymerase elongation. Similar patterns were observed in repeated trials. The experiment was also carried out using either FS1 (Figure 12, lane 1) or FS2 (Figure 12, lane 3) as the sequence-specific primer in both PCR steps. Under these conditions, specific and non-specific ssDNA products from the primer extension step are amplified in the final PCR step.

To determine whether the band represented in Figure 12, lane 1 contained the desired sequence, the gel was Southern blotted and hybridized with the kinased FS2 primer as probe. No signal was observed, indicating that the majority of the

FIGURE 12. **Exponential amplification of ssDNAs produced by the anchored-PCR technique.** Lane 1: ssDNA produced by primer extension with the FS1 primer and amplified exponentially with FS1 and poly(dC); Lane 2: ssDNA produced by primer extension with the FS1 primer and amplified exponentially with FS2 and poly(dC); Lane 3: ssDNA produced by primer extension with the FS2 primer and amplified exponentially with FS2 and poly(dC); Lane 4: negative control, PCR reaction with no DNA template; Lane 5: positive control, PCR reaction with primers B and B'; Lane 6: marker, lambda DNA hydrolyzed with HindIII.

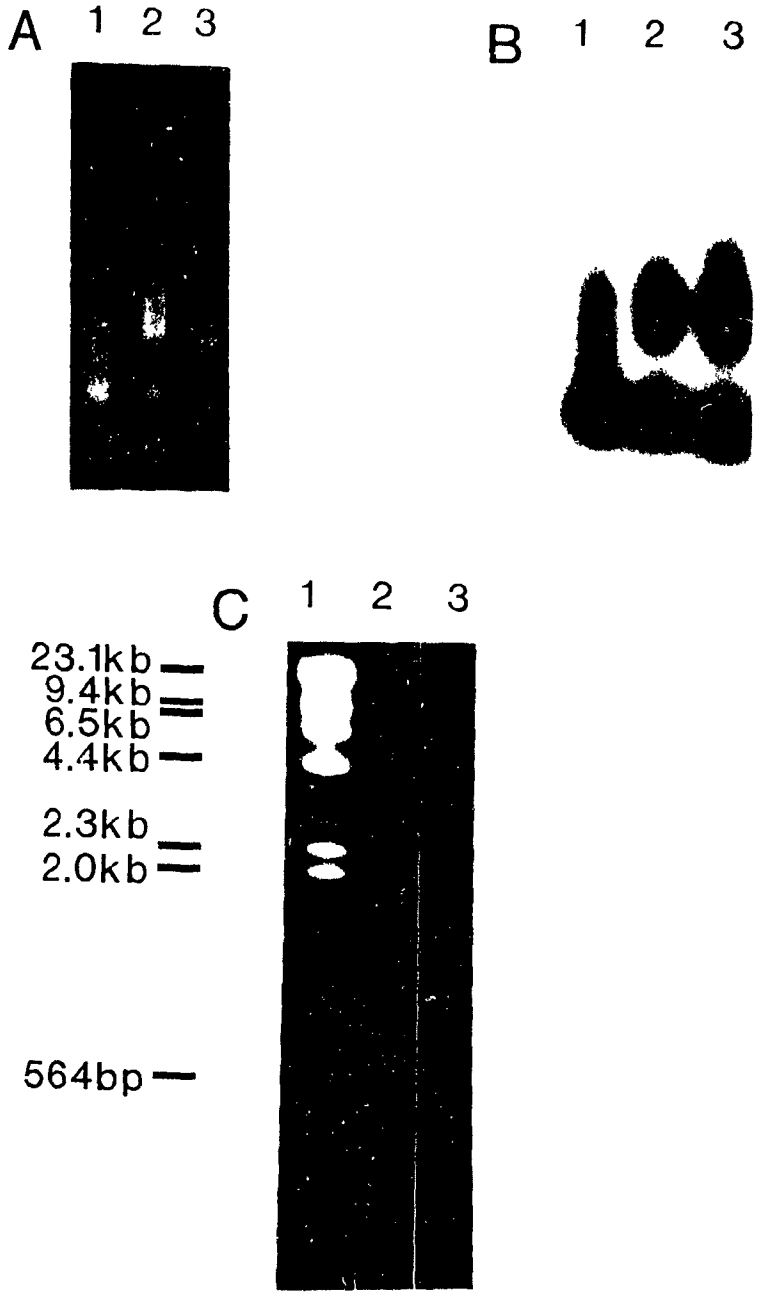

1 2 3 4 5 6
23.1kb
9.4kb
6.5kb
4.4kb
2.3kb
2.0kb
897bp
564bp

DNA was not specific for the tk sequences (data not shown). It is more difficult to design a probe to determine the specificity of DNA amplified with FS2 because there may be as few as 22 bp of tk sequence between FS2 and the 3' host DNA sequences (as determined by Southern blot analysis) (see Figures 2 and 3). Therefore an alternate approach was employed. If Taq polymerase elongated past the SacI or BamHI sites within the flanking cellular DNA, then digestion of the PCR products with one of these enzymes should produce an increased number of fragments of uniform length that are bounded by FS2 and either SacI or BamHI. PCR products from FS2 amplifications were cleaved with either SacI or BamHI, and the results are presented in Figure 13. Experiments using FS2 in both the primer extension and exponential amplification steps are presented in panel A of Figure 13. The SacI digest is inconclusive - no single band of ~45 to 150 bp can be distinguished by an increased intensity over the background (Figure 13, panel A, lane 2). Although the BamHI digest appeared to produce a band of the expected 1.7 kb (Figure 13, panel A, lane 1), when subjected to electrophoresis through a 1.3% agarose gel, the band proved to be only 1.3kb in length (data not shown). To detect any bands that may not be visible in the photograph, the gel was Southern blotted and hybridized with FS2 as probe but no other bands could be distinguished (Figure 13, panel B). In the FS1 primer-extended and FS2 amplified PCR products, a

FIGURE 13. **Hydrolysis of DNA amplified with the anchored-PCR technique.** Panel A: PCR products were produced by using FS2 in both the primer extension and exponential amplification steps. Lane 1: PCR products hydrolyzed with BamHI; Lane 2: PCR products hydrolyzed with SacI; Lane 3: uncut PCR products.

Panel B: Southern blot of gel shown in Panel A. The FS2 primer labelled at the 3' end with gamma-^{32}P was used as a probe.

Panel C: PCR products were produced by using FS1 in the primer extension step and FS2 in the exponential amplification step. Lane 1: marker, lambda DNA hydrolyzed with HindIII; Lane 2: PCR products hydrolyzed with SacI; Lane 3: PCR products hydrolyzed with BamHI.


A
1 2 3
B
1 2 3
C
1 2 3
23.1kb
9.4kb
6.5kb
4.4kb
2.3kb
2.0kb
564bp

band of the expected 1.7 kb was observed (Figure 13, panel C, lane 3).

D.2. Single-Specific-Primer-PCR

A second method used to amplify the unknown sequences flanking the tk gene followed the single-specific-primer (SSP)-PCR scheme developed by Shyamala and Ames (1989). KT genomic DNA was cleaved with a combination of enzymes (BglII and SacI or EccRI and BamHI), ligated to the cloning vector pUC118 or pUC119, and the specific ligated fragment was amplified with a sequence specific primer, FS1, and a generic primer specific for the vector, FSP (M13 universal forward sequencing primer) (Figure 14). Although a variety of fragments are expected to ligate to the vector DNA, specificity is imparted by FS1 and only the fragment containing the tk gene will accumulate in a PCR reaction. Although a range of chromosomal DNA concentrations was used, a specific band was not detected, either by ethidium bromide staining (Figure 15, lanes 2-6) or by Southern blotting and hybridization with FS2 (data not shown)(note - only one of the ligation combinations is shown). However, the merit of the procedure is indicated by its ability to specifically amplify the 1.1 kb fragment containing the tk gene from a ligation mix containing 5 ng of digested pSV2neoKT and 6 μg/ml of digested AK143 DNA as inserts plus 0.2 μg of pUC118 DNA

FIGURE 14. **Schematic representation of amplification of KT genomic sequences adjoining the tk gene using the single-sequence-specific-PCR technique.** Four different ligations were performed and are presented as A, B, C, and D. The primers used in the PCR reaction are indicated by arrows:FS1 and FSP. Restriction endonuclease cleavage sites are represented as follows, B: BamHI; Bg: BglII; E: EcoRI; P: PstI; and SacI. The regions to be amplified are bracketed and the size of the expected fragment is indicated.

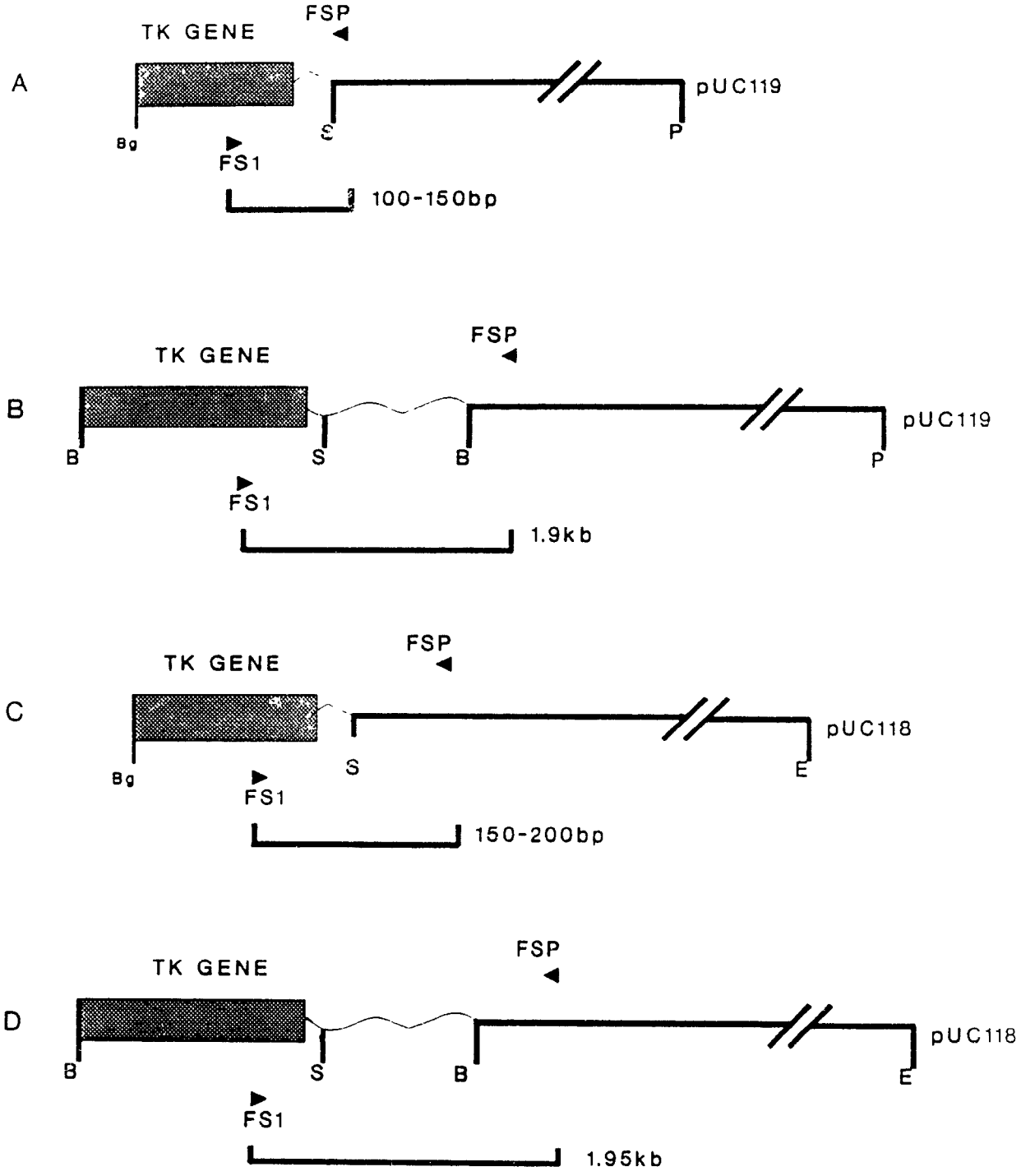
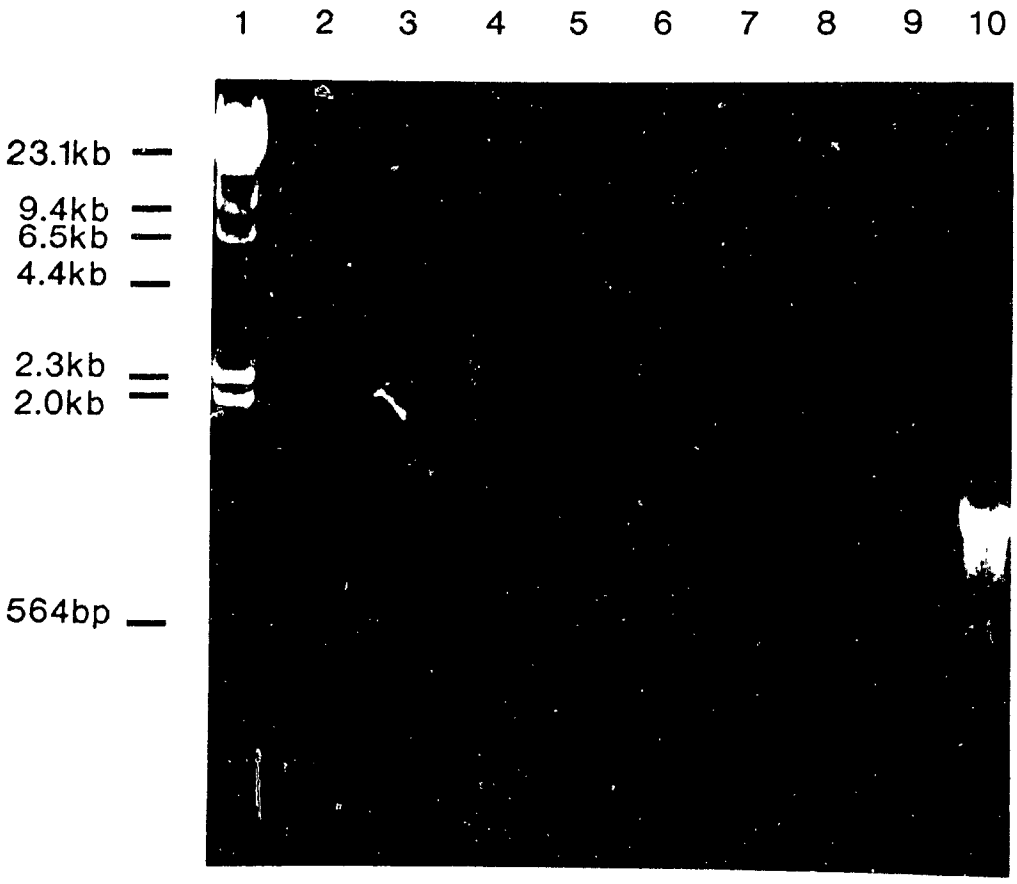

A
TK GENE
FSP
pUC119
Bg
FS1
100-150bp
B
TK GENE
FSP
pUC119
B
S
B
P
FS1
1.9kb
C
TK GENE
FSP
pUC118
Bg
S
E
FS1
150-200bp
D
TK GENE
FSP
pUC118
B
S
B
E
FS1
1.95kb
KT genomic DNA

FIGURE 15. **Amplification of KT genomic sequences adjoining the tk gene using the single-sequence-specific-PCR technique.** Lane 1: marker, lambda DNA hydrolyzed with HindIII; Lanes 2-5: exponential amplification of genomic DNA from the KT cell line that was digested with PstI and SacI and ligated to the pUC119 vector linearized with BglII and SacI; the primers used are FS1 and FSP; the concentration of KT genomic DNA used in the ligation reaction was 6.0 μg (Lane 2). 4.0 μg (Lane 3), 2.5 μg (Lane 4), 1.0 μg (Lane 5), and 0.06 μg (Lane 6). Lane 7: positive control, amplification of genomic DNA from the KT cell line with primers B and B' (897 bp); Lane 8: negative control, amplification as for lanes 2-6 but with no template DNA; Lane 9: negative control, amplification of genomic DNA from the AK143 cell line that was prepared as for the genomic DNA in lanes 2-6 and ligated to pUC119; amplification with primers FS1 and FSP; Lane 10: positive control, amplification of pSV2neoKT sequences ligated to pUC119 with the FS1 and FSP primers.

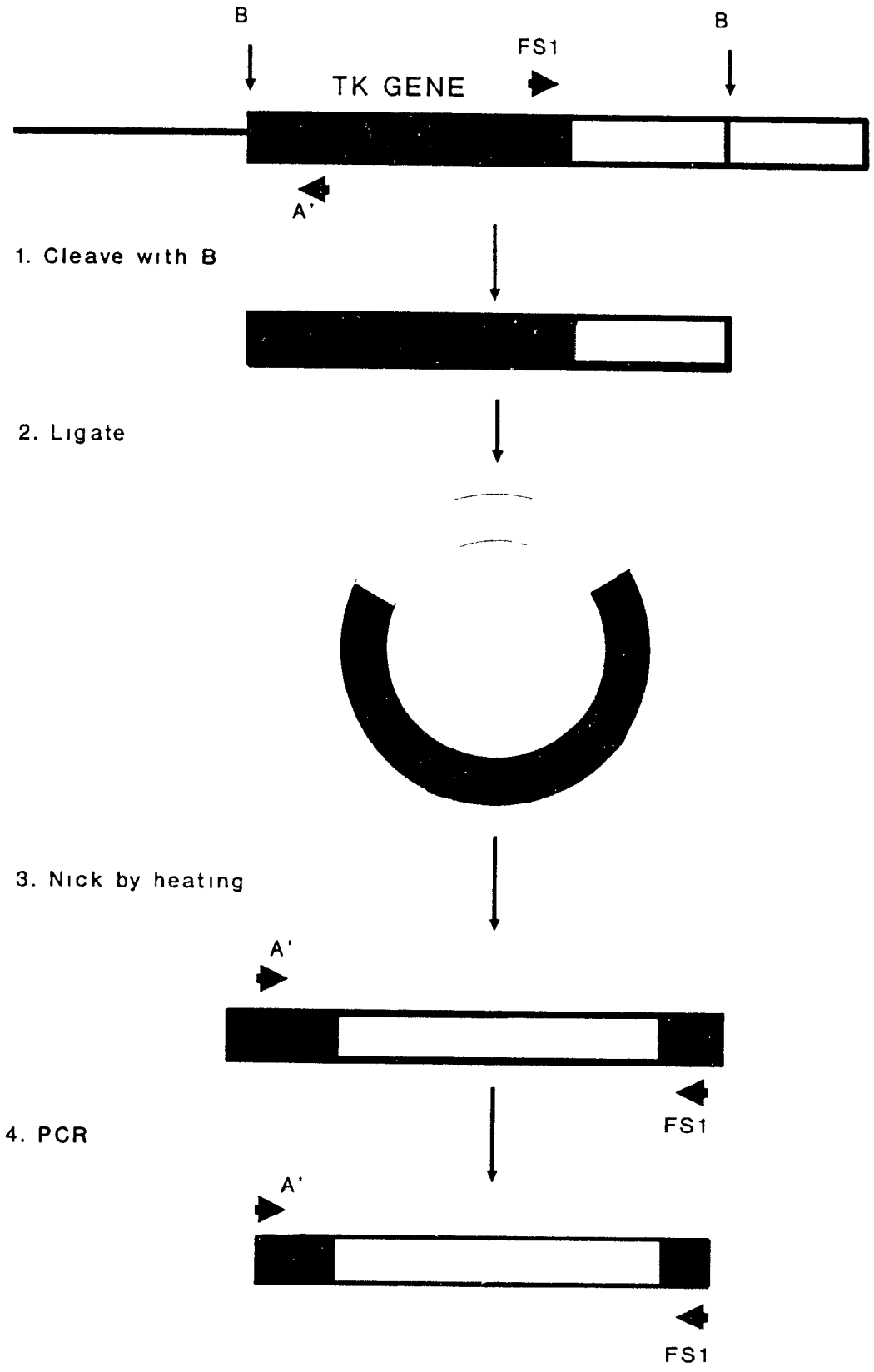

1
2
3
4
5
6
7
8
9
10
23.1kb
9.4kb
6.5kb
4.4kb
2.3kb
2.0kb
564bp

as vector (Figure 15, lane 10).

D.3. Inverse-PCR

The inverse polymerase chain reaction (IPCR) technique developed by Triglia *et al*. (1988) was also attempted. KT genomic DNA was cleaved with *Bam*HI, ligated under conditions that favour circularization and then amplified in a PCR reaction with the primers FS1 and A' (Figure 16). However, visualization of the products on an agarose gel and subsequent hybridization of the Southern blot with the FS2 probe failed to identify the expected 3 kb fragment (data not shown).

FIGURE 16. **Schematic representation of the inverse-PCR technique used to amplify the genomic sequences adjoining the integrated pSV2neoKT plasmid.** The black box represents the tk gene and the open box represents the adjoining genomic sequences. The BamHI restriction endonuclease cleavage sites are represented by the letter B. The primers used for amplification are indicated by arrows and labelled FS1 and A'.


B
B
FS1
TK GENE
A'
1. Cleave with B
2. Ligate
3. Nick by heating
A'
FS1
4. PCR
A'
FS1

IV. DISCUSSION

The KT cell line was developed to study the molecular nature of spontaneous mutations at a single chromosomal location in human AK143 cells. Antiviral drugs were used to select for mutations in the exogenously introduced HSV-1 tk gene. When a single drug was used as the selective agent, the system's mutation frequency was 2-4 orders of magnitude higher than the endogenous rate (10^{-3}-10^{-4} versus 10^{-6}-10^{-8}) and most of the mutants exhibited a very high reversion frequency to the wild type phenotype (Goring et al., 1987). Theoretically, exposure to a combination of drugs results in the selection of less frequently occuring mutational events, such as large deletions and complex rearrangements, that render the cell resistant to a number of drugs. When used in combination, the drugs were shown to enrich for non-reverting mutant genotypes (Goring et al., 1987).

Furthermore, restriction enzyme analysis of two deletion mutants suggested that the 3' ends of the deletions terminated in potentially repetitive sequences in the host DNA (Goring et al., 1987). The objective of this thesis was to clone and characterize the cellular DNA flanking the wild type tk gene in order to determine the influence of host sequence organization on the frequency and types of mutations isolated.

Southern blot analysis of DNA from the KT cell line

revealed that the SacI, BamHI and EcoRI restriction endonuclease cleavage pattern previously observed in the analysis of two deletion mutants was also present in the cellular DNA adjoining the wild type tk gene (Goring et al., 1987). The positions of other restriction endonuclease sites were not identical, indicating that the host sequences analyzed represent different regions of the host genome. The presence of the consecutive restriction motif SacI - BamHI - EcoRI suggests that the cellular sequences directly adjoining both the tk gene and the 3' ends of the two deletion mutants may share a repetitive sequence organization.

Repetitive DNA elements are favoured target sites for the integration of exogenous DNA (Kato et al., 1986) and have been associated with genome instabilities, rearrangements and deletions (Butner and Lo, 1986; Wahls et al., 1990; Myerowitz and Hogikyan, 1987). If pSV2neoKT has integrated in a repetitive region of the genome, the frequency of detectable deletions and rearrangements is expected to be higher than that observed in non-repetitive regions (Wahls et al., 1990). However, in the KT cell line, the recovery of such mutants may be limited or even prevented by the constant selection pressure that is maintained for a functional neomycin gene.

Attempts to isolate the cellular DNA flanking the tk gene, either by plasmid rescue or by separately cloning each junction site as two smaller fragments, were not successful. Plasmid rescue of the integrated pSV2neoKT with the flanking

host DNA requires a functional bacterial origin of replication, but this sequence may have mutated either during transfection (Hauser et al., 1986) or while integrated in the mammalian genome. Plasmids expressing a resistance to ampicillin were successfully recovered from the SacI hydrolyzed KT genomic DNA. indicating that the bacterial origin of replication was intact. However, the plasmids isolated were smaller than even the original pSV2neoKT plasmid and they failed to grow on kanamycin selective media, suggesting that deletions occurred during the rescue procedure which included the neomycin resistance gene. It is possible that sequences in the host DNA were involved in directing the deletion events but the plasmids were not further characterized because the accuracy of sequence information obtained for the DNA flanking the integration site could not be determined.

Initial experiments designed to clone the 4.3 kb BamHI fragment containing the right junction site failed, after 14 separate trials, to produce a positive transformant (see figure accompanying Table 2). Additional experiments with a second vector and a variety of E. coli strains with different genetic backgrounds were also carried out but did not succeed in cloning the desired fragment. Analysis of the transformants revealed that a large number of recombinant plasmids contained smaller inserts than expected from the size fractionated KT genomic DNA which was used for cloning.

The deletions were found with equal frequency in pUC119 and pBR322 recombinants but a disproportionately higher number were found in plasmids transformed into E. coli strain DH1. Although the recA mutation in DH1 should reduce the frequency of deletions that are due to recombination events, deletions of cloned mammalian DNA have been documented in recA hosts (Nikaido et al., 1981; Taub et al., 1983). It has been suggested that E. coli hosts containing additional mutations in the recB, recC, and sbcB genes may be required to propagate regions of the human genome that contain inverted repeats (Leach and Stahl, 1983; Wyman et al., 1985) and palindromic structures (Nader et al., 1985). The presence of inverted repeats, dyad symmetries or repetitive DNA elements (Jeffreys et al., 1985; Wahls et al., 1990) in the DNA flanking the integrated pSV2neoKT may have contributed to the difficulties encountered in cloning this region of the human genome.

Furthermore, restriction of methylated sequences by the McrA and McrB modification systems may have played a major role in generating the deletions observed in both the plasmid rescue and cloning experiments. Although the integrated tk gene was shown not to be highly methylated (Goring et al., 1987), it is possible that methylated residues in the adjoining cellular sequences were degraded by the Mcr modification system. Of the five E. coli strains used in these experiments only one, 401-B ER1451, is known to be deficient for both the McrA and McrB restriction systems (see Table

1).

The polymerase chain reaction is an attractive alternative to the repetitive cloning and transformation required to isolate the single copy tk gene from the human genome. In addition to amplifying single copy sequences by a factor of 2^{30}, the PCR technique can be used to directly sequence amplified DNA fragments that are difficult to clone.

DNA segments within the single copy tk gene were successfully amplified with PCR although numerous non-specific products were obtained with the amplification of the longer fragments (1.4 kb). However, cellular DNA surrounding the integrated pSV2neoKT cannot be amplified by conventional PCR techniques because the sequence is not known. The method we modified to amplify cellular DNA surrounding the integrated pSV2neoKT is based on a protocol published by Frohman et al.(1988). Our technique can be divided into three major steps : 1. unidirectional primer extension, 2. homopolymer tailing and 3. exponential amplification (Figure 7). The primer extension step is critical because the entire procedure depends on the production of specific ssDNA. Unfortunately, it is difficult to check the specificity of the ssDNA at this stage. First, there may be as few as 100 bp of amplified tk sequences therefore specifically amplified products may not hybridize well to a 3.6 kb or 2.0 kb tk probe. Second, the ssDNA products are few in number and vary in length therefore a positive hybridization signal

would be faint and smeary which could also be interpreted as non-specific background hybridization. Consequently, the conditions used for primer extension were identical to those used in the conventional PCR amplification of tk sequences. Since the primers used in these two PCR reactions are similar in length and G+C content, they are predicted to anneal under the same conditions with similar specificity.

Between steps one and two, the ssDNA products were purified in order to prevent homopolymer tailing of the primers that are present in excess. Because of their small size and large initial concentration, these tailed primers could serve as potent substrates for the subsequent PCR reaction. Both the purification and homopolymer tailing procedures were tested on control DNA fragments and proved successful. As observed by Frohman et al. (1988), the length of the homopolymer (dG) tail did not exceed 20-25 nucleotides, even though no precautions were taken to limit its length.

The final step, exponential PCR amplification, was performed with one homopolymer primer (poly (dC)) and one primer that matched sequences within the tk gene. When possible, the specific primer was designed to match DNA sequences already amplified by primer extension, thus increasing the specificity of the reaction. As expected, the final PCR products varied in length and produced a visible smear on agarose gels. To determine whether specific amplification had occurred, the DNA products were hydrolyzed with

either SacI or BamHI in an effort to identify the desired sequence by its predicted size. After 5 complete trials of the entire procedure, a band corresponding to the expected molecular weight was identified in the BamHI digest (Figure 13, panel C, lane 3). However, the specificity of the product could not be determined by Southern blot analysis because of the unavailability of an appropriate DNA probe.

DNA amplified using the modified anchored-PCR technique with FS1 alone did not hybridize with the FS2 probe, suggesting that the FS1 primer was not specifically annealing to sequences within the tk gene. It is possible that the FS1 primer requires different PCR conditions in order to specifically amplify DNA from the KT cell line.

The application of two additional PCR techniques, designed to amplify DNA of unknown sequence, was also examined during the course of this research. The first technique, SSP-PCR, utilized the universal forward sequencing primer (FSP) as a generic oligonucleotide to amplify DNA fragments ligated to pUC118 and pUC119 vectors. Although specific amplification of the host DNA flanking the tk gene was not attained, a test amplification of 5 ng of pSV2neoKT was successful. The ratio of chromosomal DNA to vector DNA is critical in ligation reactions and may be the limiting step in the SSP-PCR procedure. It is therefore possible that specific amplification of the chromosomal DNA was prevented not by the PCR reaction itself but by failure to

create the optimal ligation conditions required to obtain the correct ligation products.

The second technique, IPCR, was designed to amplify a 3 kb stretch of DNA which included the surrounding cellular sequences using two tk-specific olionucleotide primers. Failure to observe a product of ~3 kb may be due to the greater difficulty involved in amplifying longer stretches of DNA. However, amplification of DNA fragments up to 6.5 kb has been reported (Schwartz et al., 1990), and protocol modifications may be required to generate the longer PCR products.

Although the original objective to clone and characterize the cellular sequences flanking the integrated pSV2neoKT plasmid was not achieved, the results are valuable in designing modifications to improve the techniques employed.

For example, in order to pursue the plasmid rescue and cloning techniques, it is critical to transform the plasmids into E. coli strains exhibiting both an McrA$^-$ and McrB$^-$ phenotype in order to prevent degradation of methylated genomic DNA. A thymidine kinase deficient E. coli strain may also prove useful in the isolation of the right junction fragment containing the integrated tk gene (Martin et al., 1989). Transformation of the KT genomic fragment containing the tk gene into bacteria may also have been complicated by the potentially repetitive organization of the flanking DNA.

Amplification with the PCR technique provides a promising alternative to the isolation of single copy mammalian genes on bacterial plasmids. Direct sequencing of the product may provide sufficient sequence information to identify the chromosomal location of pSV2*neo*KT through *in situ* hybridization analysis or it may be used to design a host-specific primer for subsequent amplification of adjoining regions. However, the major problem encountered with this procedure has been the inability to positively identify the specifically amplified products. This is primarily due to the fact that, even in the final amplification, the specific products are expected to be of varying lengths. Therefore, it may be advantageous to utilize a third sequence-specific primer as a probe to identify the amplified products generated by the two nested primers. Furthermore, higher yields of longer PCR products may be obtained by adding the T4 gene 32 protein to the PCR reactions in order to stabilize and enhance the activity of *Taq* polymerase (Schwartz *et al.*, 1990).

To analyze the host DNA surrounding *tk*$^-$ mutants on a regular basis, the SSP-PCR or IPCR techniques, which involve fewer manipulations than the anchored-PCR method, may prove more efficient. To develop the SSP-PCR method for our system it is necessary to determine the optimal insert:vector ratio for the ligation reaction. The IPCR technique may be improved by designing a primer to replace

A' that would permit linearization of the circular molecule by cleavage at a specific restriction site (i.e. BglII) rather than by a non-specific 10 minute heat incubation at 65^{o} which was used in the present protocol.

In order to achieve a complete understanding of the frequencies and types of mutations isolated in the KT cell line, the identity of the sequences flanking the integrated plasmid pSV2neoKT remains critical. The assessment of techniques presented in this thesis will prove valuable in directing future research but it is also important to explore other methods of characterizing regions of genomic DNA. Experiments involving nuclease digests (Davies et al., 1982), DNAseI sensitivity assays (Davies et al., 1982) and Southern blots with repetitive DNA probes (for example Alu repetitive elements or satellite sequences) may provide valuable clues in the characterization of the cellular DNA flanking pSV2neoKT and determination of its influence on the generation of tk mutants.

V. REFERENCES

1. Albertini, A.M., Hofer, M., Calos, M.P., and Miller, J.H. (1982). The importance os fhort sequence homologies in the generation of large deletions. Cell 29, 319-328.

2. Albertini, R.J., O'Neill, J.P., Nicklas, J.A., Heintz, N.H., and Kelleher,P.C. (1985). Alterations of the HPRT gene in human in vivo-derived 6-thioguanine resistant T lymphocytes. Nature 316, 369-371.

3. Allgood, N.D., and Silhavy, T.J. (1988). In Genetic recombination. Kucherlapati, R., and Smith, S.R. (ed.) American Society for Microbiology, Washington, D.C.

4. Anderson, R.A., Krakauer, T., and Camerini-Otero, R.D. (1982). DNA mediated gene transfer : recombination between cotransfected DNA sequences and recovery of recombinants in a plasmid. Proc. Natl. Acad. Sci. USA 79, 2748-2752.

5. Applegate, M.L., Moor, M.M., Broder, C.B., Burrell, A., Juhn, G., Kasweck, K.L., Lin, P.-F., Wadhams, A., and Hozier, J.C. (1990). Molecular dissection of mutations at the heterozygous thymidine kinase locus in mouse lymphoma cells. Proc. Natl. Acad. Sci. USA 87, 51-55.

6. Ashman, C.R., and Davidson, R.L. (1984). High sponta-

neous mutation frequency in shuttle vector sequences recovered from mammalian cellular DNA. Mol. Cell. Biol. 4, 2266-2272.

7. Ashman, C.R. and Davidson, R.L. (1985). High spontaneous mutation frequency of BPV shuttle vector. Somat. Cell Mol. Genet. 11, 499-504.

8. Ashman, C.R., and Davidson, R.L. (1987a). Sequence analysis of spontaneous mutations in a shuttle vector gene integrated into mammalian chromosomal DNA. Proc. Natl. Acad. Sci. USA 84, 3354-3358.

9. Ashman, C.R., and Davidson, R.L. (1987b). DNA base sequence changes induced by ethyl methanesulfnate in a chromosomally integrated shuttle vector gene in mouse cells. Som. Cell Mol. Genet. 13, 563-568.

10. Ashman, C.R., Jagadeeswaran, P., and Davidson, R.L. (1986). Efficient recovery and sequencing of mutant genes from mammalian chromosomal DNA. Proc. Natl. Acad. Sci. USA 83, 3356-3360.

11. Bell, G.I., Selby, M.J., and Rutter, W.J. (1982). The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. Nature

295, 31-35.

12. Benton, W.D., and Davis, R.W. (1977). Screening lambda-gt recombinant clones by hubridization to single plaques in situ. Science 196, 180-182.

13. Benzer, S. (1961). On the topography of the genetic fine structure. Proc. Natl. Acad. Sci. USA 47, 403-415.

14. Bilimoria, M.H., and Gupta, S.V. (1986). Comparison of the mutagenic activity of 5-hydroxymethyl-deoxyuridine with 5-substituted 2'-deoxyuridine analogues in the Ames Salmonella/microsome test. Mut. Res. 169, 123-127.

15. Bird, A.P. (1986). CpG-rich islands and the function of DNA methylation. Nature 321, 209-213.

16. Blin, N., and Stafford, D.W. (1976). A general method for isolation of high molecular weight DNA from eukaryotes. Nucleic Acids Res. 3, 2303-2308.

17. Bolivar, F., Rodriguez, R.L., Betlach, M.C., and Boyer, H.W. (1977a). Construction and characterizationof new cloning vehicles. I. Ampicillin-resistant derivatives of the plasmid pMB9. Gene 2, 75-93.

18. Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach,

M.C., Heyneker, H.L., Boyer, H.W., Crosa, J.H., and Falkow, S. (1977b). Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene2, 95-113.

19. Boyer, H.W., and Roulland-Dussoix, D. (1969). A complementation analysis of the restriction and modification of DNA in E. coli. J. Mol. Biol. 41, 459-472.

20. Bradley, W.E.C., and Letonavec, D. (1982). High-frequency nonrandom mutational events at the adenine phosphoribosyl transferase (aprt) locus of sib-selected CHO variants heterozygous for aprt. Somat. Cell Mol. Genet. 8, 51-61.

21. Breitman, M.L., Tsui, L.-C., Buchwald, M., and Siminovitch, L. (1982). Introduction and recovery of a selectable bacterial gene from the genome of mammalian cells. Mol. Cell. Biol. 2, 966-976.

22. Butner, K., and Lo. C.W. (1986). Modulation of tk expression in mouse pericentromeric heterochromatin. Mol. Cell Biol. 6, 4440-4449.

23. Calos, M.P., Lebkowski, J.S., and Botchan, R.B. (1983). High mutation frequency in DNA transfected into mammalian cells. Proc. Natl. Acad. Sci. USA 80, 3015-3019.

24. Chien, A., Edgar, D.B., and Trela, J.M. (1976). Deoxyribonucleic acid polymerase from the extreme thermophile Thermus Aquaticus. J. Bacteriol. 127, 1550-1557.

25. Collins, F.S., and Weissman, S.M. (1984). Directional cloning of DNA fragments at a large distance from an initial probe : a circularization method. Proc. Natl. Acad. Sci. USA 81, 6812-6816.

26. Cox, R., and Masson, W.K. (1978). Do radiation-induced thioguanine-resistant mutants of cultured mammalian cells arise by HPRT gene mutation or X-chromosome rearrangement? Nature 276, 629-630.

27. Davidson, R.L., Brocker, P., and Ashman, C.R. (1988). DNA base sequence changes and sequence specificity of bromodeoxyuridine-induced mutations in mammalian cells. Proc. Natl. Acad. Sci. USA 85, 4406-4410.

28. Davies, R.L., Fuhrer-Krusi, S., and Kucherlapati, R.S. (1982). Modulation of transfected gene expression mediated by changes in chromatin structure. Cell 31, 521-529.

29. deJong, P., Grosovsky, A.J., and Glickman, B.W. (1988). Spectrum of spontaneous mutation at the APRT locus of Chi-

nese hamster ovary cells : An analysis at the DNA sequence level. Proc. Natl. Acad. Sci. USA 85, 3499-3503.

30. Dobrovic, A., Gareau, J.L.P., Ouellette, G., and Bradley, W.E.C. (1988). DNA methylation and genetic inactivation at thymidine kinase locus : two different mechanisms for silencing autosomal genes. Som. Cell Mol. Genet. 14, 55-68.

31. Drinkwater, N.R., and Klinedinst, D.K. (1986). Chemically induced mutagenesis in a shuttle vector with a low-background mutant frequency. Proc. Natl. Acad. Sci. USA 83, 3402-3406.

32. Ehrlich, M., and Wang, R.Y. (1981). 5-Methylcytosine in eukaryotic DNA. Science 212, 1350-1357.

33. Engelke, D.R., Hoener, P.A., and Collins, F.S. (1988). Direct sequencing of enzymatically amplified human genomic DNA . Proc. Natl. Acad. Sci. USA 85, 544-548.

34. Evans, H.H., Mencl, J., Horng, M.-F., Ricanati, M., Sanchez, C., and Hozier, J. (1986). Locus specificity in the mutability of L5178Y mouse lymphoma cells : The role of multilocus lesions. Proc. Natl. Acad. Sci. USA 83, 4379-4383.

35. Farabaugh, P.J., Schmeissner, U., Hofer, M., and Miller, J.H. (1978). Genetic studies of the lac repressor VII. On the molecular nature of spontaneous hotspots in the lacI gene of Escherichia coli. J. Mol. Biol. 126, 847-863.

36. Feiss, F., and Siegele, D.A. (1979). Packaging of the bacteriophage lambda chromosome : Dependence of cos cleavage on chromosome length. Virol. 92, 190-200.

37. Frohman, M.A., Dush, M.K., and Martin, G.R. (1988). Rapid production of full-length cDNAs from rare transcripts : Amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA 85, 8998-9002.

38. Fuscoe, J.G., Fenwick, R.G., Ledbetter, D.H., and Caskey, C.T. (1983). Deletion and amplification of the HGPRT locus in chinese hamster cells. Mol. Cell. Biol. 3, 1086-1096.

39. Gennett, I.N., and Tilly, .. G. (1988). Mapping large spontaneous deletion endpoints in the human HPRT gene. Mut. Res. 201, 149-160.

40. Godbout, R., Dryja, T.P., Squire, J., Gallie, B.L., and Phillips, R.A. (1983). Somatic inactivation of genes on chromosome 13 is a common event in retinoblastoma. Nature

304, 451-453.

41. Goring, D.R. (1985). M.Sc. thesis manuscript. McGill University, Montreal, Canada.

42. Goring, D.R., and DuBow, M.S. (1985). A cytotoxic effect associated with 9-(1,3-dihydroxy-2-propoxymethylG)-guanine is observed during the selection for drug resistant human cells containing a single herpesvirus thymidine kinase gene. Biochem. Biophys. Res. Commun. 133, 195-201.

43. Goring, D.R., Gupta, K., and DuBow, M.S. (1987) Analysis of spontaneous mutations in a chromosomally located HSV-1 thymidine kinase (TK) gene in a human cell line. Somat. Cell Mol. Genet. 13, 47-56.

44. Gough, J.A., and Murray, N.E. (1983). Sequence diversity among related genes for recognition of specific targets in DNA molecules. J. Mol. Biol. 166, 1-19.

45. Grosovsky, A.J., Drobetsky, E.A., deJong, P.J., and, Glickman, B.W. (1986). Southern analysis of genomic alterations in gamma-ray-induced APRT$^-$ hamster cell mutants. Genetics 113, 405-415.

46. Grunstein, M., and Hogness, D.S. (1975). Colony hybri-

dization: A method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. USA 72, 3961-3965.

47. Hanahan, D. (1983). Studies on transformation in E. coli with plasmids. J. Mol. Biol. 166, 557-580.

48. Hauser, J., Levine, A.S., and Dixon, K. (1987). Unique pattern of point mutations arising after gene transfer into mammalian cells. EMBO 6, 63-67.

49 Hirt, B. (1967). Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. 26, 365-369.

50. Jeffreys, A.J., Wilson, V., and Thein, S.L. (1985). Hypervariable "minisatellite" regions in human DNA. Nature 314, 67-73.

51. Jelinek, W.R., and Schmid, C.W. (1982). Repetitive sequences in eukaryotic DNA and their expression. Annu. Rev. Biochem. 51, 813-844.

52. Kato, S., Anderson, R.A., and Camerini-Otero, R.D. (1986). Foreign DNA introduced by calcium phosphate is integrated into repetitive DNA elements of the mouse L cell

genome. Mol. Cell. Biol. 6, 1787-1795.

53. Kobori, J.A., Strauss, E., Minard, K., and Hood, L. (1986). Molecular analysis of the hotspot of recombination in the murine major histocompatibility complex. Science 234, 173-179.

54. Leach, D.R.F., and Stahl, F. (1983). Viability of lambda phages carrying a perfect palindrome in the absence of recombination nucleases. Nature 305, 448-451.

55. Lebkowski, J.S., DuBridge, R.B., Antell, E.A., Greisen, K.S., and Calos, M.P. (1984). Transfected DNA is mutated in monkey, mouse, and human cells. Mol. Cell. Biol. 4, 1951-1960.

56. Lebkowski, J.S., Clancy, S., Miller, J.H., and Calos, M.P. (1985). The lacI shuttle : rapid analysis of the mutagenic specificity of ultraviolet light in human cells. Proc. Natl. Acad. Sci. USA 82, 8606-8610.

57. Lebo, R.V., Chakravarti, A., Buetow, K.H., Cheung, M., Cann, H., Cordell, B., and Goodman, H. (1983). Recombination within and between the human insulin and beta-globin gene loci. Proc. Natl. Acad. Sci. USA 80, 4808-4812.

58. Lehrman, M.A., Schneider, W.J., Sudhof, T.C., Brown, M.S., Goldstein, J.L., and Russel, D.W. (1985). Mutation in LDL receptor : Alu-Alu recombination deletes exons encoding transmembrane and cytoplasmic domains. Science 227, 140-146.

59. Lehrman, M.A., Goldstein, J.L., Russel, D.W., and Brown, M.S. (1987). Duplication of seven exons in LDL receptor gene caused by Alu-Alu recombination in a subject with familial hypercholesterolemia. Cell 48, 827-835.

60. Lewin, B.M. (ed.) (1985) Genes II. John Wiley and Sons, Inc. New York, N.Y.

61. Liber, H.L., Yandell, D.W., and Little, J.B. (1989). A comparison of mutation induction at the tk and hprt loci in human lymphoblastoid cells ; quantitative differences are due to an additional class of mutations at the autosomal tk locus. Mut. Res. 216, 9-17.

62. Littlefield, J.W. (1964). Selection of hybrids from matings of fibroblasts in vitro and their presumed recombinants. Science 145, 709-710.

63. Mandel, M., and Higa, A. (1970). Calcium dependent

bacteriophage DNA infection. J. Mol. Biol. 53, 159-162.

64. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982). Molecular cloning : A laboratory manual. Cold Spring Harbour Laboratory, Cold Spring Harbour, N.Y.

65. Martin, S.L., Aparisio, D.I., and Bandyopadhyay, P.K. (1989). Genetic and biochemical characterization of the thymidine kinase gene from Herpesvirus of turkeys. J. Virol. 63, 2847-2852.

66. McGeoch, D.J., Dalrymple, M.A., Davison, A.J., Doaln, A., Frame, M.C., McNab, D., Perry, L.J., Scott, J.E., and Taylor, P. (1988). The complete DNA sequence of the long unique region in the genome of herpes simplex virus type 1. J. Gen. Virol. 69, 1531-1574.

67. Melton, D.W., Konecki, D.S., Brennard, J., and Caskey, C.T. (1984). Structure, expression and mutation of the hypoxanthine phosphoribosyl transferase gene. Proc. Natl. Acad. Sci. USA 81, 1484-1488.

68. Menck, C.F.M., James, M.R., Gentil, A., and Sarasin, A. (1987). Strategies to analyse mutagenesis in mammalian cells using simian virus 40 or shuttle vectors. J. Cell. Sci. Suppl. 6, 323-331.

69. Miller, J.H. (1972). Experiments in molecular genetics. Cold Spring Harbour Laboratory, Cold Spring Harbour, N.Y.

70. Miller, J.H., Ganem, D., Lu, P., and Schmitz, A. (1977). Genetic studies of the lac repressor I. Correlation of mutational sites with specific amino acid residues : construction of a colinear gene-protein map. J. Mol. Biol. 109, 275-301.

71. Moore, M.M., Clive, D., Hozier, J.C., Howard, B.E., Batson, A.G., Turner, N.T., and Sawyer, J. (1985). Analysis of trifluorothymidine-resistant (TFT^R) mutants of L5178Y/$TK^{+/-}$ mouse lymphoma cells. Mutation Res. 151, 161-174.

72. Moore, M.M., Harrington-Broek, K., Doerr, C.L., and Dearfield, K.L. (1989). Differential mutant quantitation at the mouse lymphoma tk and CHO hgprt loci. Mutagenesis 4, 394-403.

73. Morelles, G. (1989). A plasmid extraction procedure on a mini-prep scale. Focus 11, 7-7.

74. Mulligan, R.C., and Berg, P. (1980). Expression of a bacterial gene in mammalian cells. Science 209, 1422-1427.

75. Myerowitz, R., and Hogikyan, N.D. (1987). A deletion involving Alu sequences in the beta-hexosaminidase alpha-chain gene of French Canadians with Tay Sachs Disease. J. Biol. Chem. 262, 15396-15399.

76. Nader, W.F., Edlind, T.D., Huettermann, A., and Sauer, H.W. (1986). Cloning of Physarum actin sequences in an exonuclease-deficient bacterial host. Proc. Natl. Acad. Sci. USA 82, 2698-2702.

77. Nahmias, A.J., Dowdle, W.R., and Schinazi, R.E. (ed.) (1981). The human herpesviruses. Elsevier North Holland, Inc., New York.

78. Nalbantoglu, J., Goncalves, O., and Meuth, M. (1983). Structure of mutant alleles at the aprt locus of chinese hamster ovary cells. J. Mol. Biol. 67, 575-594.

79. Nalbantoglu, J., Hartley, D., Phear, G., Tear, G., and Meuth, M (1986). Spontaneous deletion formation at the aprt locus of hamster cells : the presence of short sequence homologies and dyad symmetries at deletion termini. EMBO 5, 1199-1204.

80. Nalbantoglu, J., Phear, G., and Meuth, M. (1987). DNA sequence analysis of spontaneous mutations at the aprt locus

of hamster cells. Mol. Cell. Biol. 7, 1445-1449.

81. Nicholls, R.D., Hill, A.V.S., Clegg, J.B., and Higgs, D.R. (1985). Direct cloning of specific genomic DNA sequences in plasmid libraries following fragment enrichment. Nucleic Acids Res. 13, 7569-7578.

82. Nicholls, R.D., Fishel-Godsian, N., and Higgs, D.R. (1987). Recombination at the human alpha-globin gene cluster : sequence features and topological constraints. Cell 49, 369-378.

83. Nikaido, T., Nakai, S., and Honjo, T. (1981). Switch region of immunoglobulin C_{μ} gene is composed of simple tandem repetitive sequences. Nature 393, 845-848.

844. Pardue, M.L., Lowenhaupt, K., Rich, A., and Nordheim, A. (1987). $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$ sequences have evolutionarily conserved chromosomal locations in Drosophila with implications for roles in chromatin structure and function. EMBO J. 6, 1781-1789.

85. Patel, P.I., Framson, P.E., Caskey, C.T., and Chinault, A.C. (1986). Fine structure of the human hypoxanthinephosphoribosyl transferase gene. Mol. Cell. Biol. 6, 393-403.

86. Perucho, M., Hanahan, D., Lipsich, L., and Wigler, M. (1980). Isolation of the chicken thymidine kinase gene by plasmid rescue. Nature 285, 207-210.

87. Phear, G., Armstrong, W., and Meuth, M. (1989). Molecular basis of spontaneous mutation at the aprt locus of hamster cells. J. Mol. Biol. 209, 577-582.

88. Pulak, R.A., and Anderson, P. (1988). Structures of spontaneous deletions in Caenorhabditis elegans. Mol. Cell. Biol. 8, 3748-3754.

89. Raleigh, E.A., and Wilson, G. (1986). Escherichia coli K12 restricts DNA containing 5-methylcytosine. Proc. Natl. Acad. Sci. USA 83, 9070-9074.

90. Raleigh, E.A., Murray, N.E., Revel, H., Blumenthal, R.M., Westaway, D., Reith, A.D., and Rigby, P.W.J. (1988). McrA and McrB restriction phenotypes of some E. coli strains and implications for gene cloning. Nucleic Acids Res. 16, 1563-1575.

91. Razzaque, A., Mizusawa, H., and Seidman, M.M. (1983). Rearrangement and mutagenesis of a shuttle vector plasmid after passage in mammalian cells. Proc. Natl. Acad. Sci. USA 80, 3010-3014.

92. Razzaque, A., Chakrabarti, S., Joffee, S., and Seidman, M. (1984). Mutagenesis of a shuttle vector plasmid in mammalian cells. Mol. Cell. Biol. 4, 435-441.

93. Reuter, G., Giarre, M., Farah, J., Gausz, J., Spierer, A., and Spierer, P. (1990). Dependence of position-effect variegation in Drosophila on dose of a gene encoding an unusual zinc-finger protein. Nature 344, 219-223.

94. Rigby, P.W.J., Dieckmann, M., Rhodes, C., and Berg, P. (1977). Labelling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J. Mol. Biol. 113, 237-251.

95. Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A., and Arnheim, N. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230, 1350-1354.

96. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., and Erlich, H.A. (1988). Primer directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239, 487-491.

97. Sanger, F., Coulson, A.R., Hong, G.F., and Peterson,

G.B. (1982). Nucleotide sequence of bacteriophage lambda DNA. J. Mol. Biol. 162, 729-773.

98. Sarkar, S., Dasgupta, U.B., and Summers, W.G. (1984). Error-prone mutagenesis detected in mammalian cells by a shuttle vector containing the supF gene of Escherichia coli. Mol. Cell. Biol. 4, 2227-2230.

99. Schaaper, R.M., Danforth, B.N., and Glickman, B.W. (1986). Mechanisms of spontaneous mutagenesis : An analysis of the spectrum of spontaneous mutation in the Escherichia coli lacI gene. J. Mol. Biol. 189, 273-284.

100. Schwartz, K., Hansen-Hagge, T., and Bartram, C. (1990). Improved yields of long PCR products using gene 32 protein. Nucleic Acids Res. 18, 1079.

101. Seed, B. (1983). Purification of genomic sequences from bacteriophage libraries by recombination and selection in vivo. Nucleic Acids Res. 11, 2427-2445.

102. Seidman, M.M., Dixon, K., Razzaque, A., Zagursky, R.J., and Berman, M.L. (1985). A shuttle vector plasmid for studying carcinogen-induced point mutations in mammalian cells. Gene 38, 233-237.

103. Shyamala, V., and Ames, G.F-L. (1989). Genome walking

by single-specific-primer polymerase chain reaction : SSP-PCR. Gene 84, 1-8.

104. Smith, G.E., and Summers, M.D. (1980). The bidirectional transfer of DNA and RNA to nitrocellulose or diazobenzyloxymethyl paper. Anal. Biochem. 109, 123-129.

105. Smith, G.R., Kunes, S.M., Schultz, D.W., Taylor, A., and Trinan, K.L. (1981). Structure if chi hotspots of generalized recombination. Cell 24, 429-436.

106. Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by agarose gel electrophoresis. J. Mol. Biol. 98, 503-515.

107. Southern P.J., and Berg, P. (1982). Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. J. Mol. Appl. Genet. 1, 327-341.

108. Stankowski Jr., L.F., and Hsie, A.W. (1986). Quantitative and molecular analysis of radiation-induced mutation in A552 cells. Radiat. Res. 105, 37-48.

109. Steinmetz, M., Stephan, D., and Lindahl, K.F. (1986). Gene organization and recombinational hotspots in the murine

major histocompatibility complex. Cell 44, 895-904.

110. Takada, S., Gotoh, Y., Hayashi, S., Yoshida, M., and Koike, K. (1990). Structural rearrangement of integrated hepatitis B virus DNA as well as cellular flanking DNA is present in chronically infected hepatic tissues. J. Virol. 64, 822-828.

111. Tasseron-deJong, J.G., denDulk, H., van de Putte, P., and Giphart-Gassler, M. (1989). De novo methylation as a major event in the inactivation of transfected herpesvirus thymidine kinase genes in human cells. Biochem. Biophys. Acta 1007, 215-223.

112. Taub, R.A., Hollis, G.F., Hieter, P.A., Korsmeyer, S., Waldman, T.A., and Leder, P. (1983). Vatiable amplification of immunoglobulin lambda light-chain genes in human populations. Nature 304, 172-174.

113. Thacker, J. (1985). The molecular nature of mutations in cultured mammalian cells : A review. Mut. Res. 150, 431-443.

114. Tischfield, J.A., Trill, J.J., Lee, Y.I., Coy, K., and Taylor, M.W. (1982). Genetic instability at the adenine phosphoribosyl transferase locus in mouse L cells. Mol. Cell. Biol. 2, 250-257.

115. Treco, D., Thomas, B., and Arnheim, N. (1985). Recombination hot spot in the human beta-globin gene cluster : meiotic recombination of human DNA fragments in Saccharomyces cerevisiae. Mol. Cell. Biol. 5, 2029-2038.

116. Triglia, T., Peterson, M.G., and Kemp, D.J. (1988). A procedure for in vitro amplification of DNA segments that lie outside the boundaries of known sequences. Nucleic Acids Res. 16, 8186.

117. Vanin, E.F., Henthorn, P.S., Kioussis, D., Grosveld, F., and Smithies, O. (1983). Unexpected relationships between four large deletions in the human beta-globin gene cluster. Cell 35, 701-709.

118. Vieira, J., and Messing, J. (1987). Production of single-stranded plasmid DNA. Methods Enzymol. 153, 3-11.

119. Vogel, F., and Motulsky, A.G. (1986). Human genetics, problems and approaches. 2nd edn., Springer, Berlin, 807.

120. Wahls, W.P., Wallace, L.J., and Moore, P.D. (1990). Hypervariable minisatellite DNA is a hotspot for homologous recombination in human cells. Cell 60, 95-103.

121. White, T.J., Arnheim, N., and Erlich, H.A. (1989). The polymerase chain reaction. Trends Genet. 5, 185-189.

122. Woodcock, D.M., Crowther, P.J., and Diver, W.P. (1987). The majority of methylated deoxycytidines in human DNA are not in the CpG dinucleotide. Biochem. Biophys. Res. Commun. 145, 888-894.

123. Woodcock, D.M., Crowther, P.J., Diver, W.P., Graham, M., Bateman, C., Baker, D.J., and Smith, S.S. (1988). RglB facilitated cloning of highly methylated eukaryotic DNA : the human L1 transposon, plant DNA, and methylated *in vitro* with human DNA methyltransferase. Nucleic Acids Res. 16, 4465-4482.

124. Worton, R.G. (1988). Genetics of Duchenne muscular dystrophy. Annu. Rev. Genet. 22, 601-629.

125. Wyman, A.R., Wolfe, L.B., and Botstein, D. (1985). Propagation of some human DNA sequences in bacteriophage lambda vectors requires mutant *E. coli* hosts. Proc. Natl. Acad. Sci. USA 82, 2880-2884.

126. Yandell, D.W., Dryja, T.P., and Little, J.B. (1986). Somatic mutations at a heterozygous autosomal locus in human

cells occur more frequently by allele loss than by intragenic structural alterations. Somat. Cell Mol. Genet. 12, 255-263.

127. Yang, T.P., Patel, P.I., Chinault, A.C., Stout, J.T., Jackson, L.G., Hildebrand, B.M., and Caskey, C.T. (1984). Molecular evidence for new mutation at the hprt locus in Lesch-Nyhan patients. Nature 310, 412-414.

128. Yanisch-Perron, C., Vieira, J., and Messing, J. (1985). Improved M13 cloning vectors and host strains : Nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33, 103-119.