MOLECULAR AND CELLULAR BIOLOGY, Nov. 1985, p. 3149–3156
0270-7306/85/113149-08$02.00/0

Vol. 5, No. 11

Molecular Cloning and Structural Analysis of Murine Thymidine Kinase Genomic and cDNA Sequences

PIN-FANG LIN, HOWARD B. LIEBERMAN, DONG-BOR YEH, TIAN XU, SHOU-YUAN ZHAO,† AND FRANK H. RUDDLE*

Department of Biology, Yale University, New Haven, Connecticut 06511

Received 29 July 1985/Accepted 9 August 1985

Two functional cytosolic thymidine kinase (*tk*) cDNA clones were isolated from a mouse L-cell library. An RNA blot analysis indicated that one of these clones contains a nearly full-length *tk* sequence and that LTK$^-$ cells contain little or no TK message. The nucleotide sequences of both clones were determined, and the functional mouse *tk* cDNA contains 1,156 base pairs. An analysis of the sequence implied that there is an untranslated 32-nucleotide region at the 5′ end of the mRNA, followed by an open reading frame of 699 nucleotides. The 3′ untranslated region is 422 nucleotides long. Thus, the gene codes for a protein containing 233 amino acids, with a molecular weight of 25,873. A comparison of the coding sequences of the mouse *tk* cDNA with the human and chicken *tk* genes revealed about 86 and 70% homology, respectively. We also isolated the *tk* gene from a mouse C57BL/10J cosmid library. The structural organization was determined by restriction mapping, Southern blotting, and heteroduplex analysis of the cloned sequences, in combination with a mouse *tk* cDNA. The *tk* gene spans approximately 11 kilobases and contains at least five introns. Southern blot analysis revealed that this gene is deleted in mouse LTK$^-$ cells, consistent with the inability of these cells to synthesize TK message. This analysis also showed that *tk*-related sequences are present in the genomes of several mouse strains, as well as in LTK$^-$ cells. These segments may represent pseudogenes.

Thymidine kinase (TK; EC 2.7.2.21) catalyzes the phosphorylation of thymidine to thymidine 5′-monophosphate, which is subsequently used for DNA synthesis. Two major forms of TK have been identified in animal cells, one localized in mitrochondria and the other in the cytosol (16). The activities of these two forms of enzyme are regulated differentially during progression through the cell cycle and in association with the replicative state of the cell. Cytosolic TK activity is high in rapidly proliferating cells, and it peaks during the S phase of the cell cycle. Little activity is detectable in cells resting at confluence. The activity of cytosolic TK also increases after infection by simian virus 40, adenovirus 12, or Epstein-Barr virus (16). In addition, terminal differentiation of cells is associated with reduced expression (1, 25). In contrast, the activity of the mitochondrial TK is relatively constant. Interest in TK has been further stimulated by the finding that the levels of the mitrochondrial and cytosolic enzyme activities in neoplastic tissue from patients with a variety of lymphoproliferative disorders are useful indicators for predicting either tumor behavior or the clinical course of patients (9).

Our goal has been to resolve the molecular mechanism that underlies the regulatory pattern of TK expression associated with proliferation and also with cellular differentiation. cDNA sequences encoding human TK and chicken TK have been isolated (5, 24; Lin, Zhao, and Ruddle, unpublished data). In addition, *tk* genomic sequences have recently been cloned from chicken (29), human (4, 19, 22), and Chinese hamster (20) cells, and they appear to retain the proper regulatory function (4, 20, 25). Unfortunately, the genomic sequences were cloned after DNA transfection, which may introduce DNA rearrangements (6, 30). In fact, the differences among the restriction maps of the human genomic *tk* sequence reported by three laboratories (4, 19, 22) may be the result of the DNA transfection artifact. As an initial approach to understanding the mechanism of *tk* control, and to define regions of the gene responsible for regulation, we isolated the functional mouse *tk* genomic sequence directly from a mouse cosmid library (a gift from G. Widera and R. Flavell) and also two functional *tk* cDNAs from a library prepared from the mRNA of mouse L cells (a gift from H. Okayama). In this report, we describe the nucleotide sequence of the mouse cytosolic *tk* cDNA and the amino acid sequence deduced from it. We also characterized the structure of the mouse *tk* gene and identified *tk*-related sequences in the mouse genome. In addition, we used the *tk* cDNA, in combination with Northern and Southern blot analysis, to show that a nonrevertible mouse LTK$^-$ cell population contains little or no cytosolic TK message and a deletion of cytosolic tk DNA sequences.

MATERIALS AND METHODS

Screening of cDNA library. A library of cDNAs prepared from mouse L cells and inserted in the expression vector pcD was provided by H. Okayama, National Institutes of Health, Bethesda, Md. The cDNAs were used to transform competent *Escherichia coli* DH1 to ampicillin resistance as described by Hanahan (13). A total of 2×10^5 colonies were screened by the high-density colony screening method (14). Nitrocellulose filters were prehybridized for 16 h at 42°C in a buffer containing 5× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate), 4× Denhardt solution (1× Denhardt solution is 0.02% bovine serum albumin [pentax fraction V], 0.02% Ficoll, and 0.02% polyvinylpyrrolidone), 10 mM Tris hydrochloride (pH 7.5), 0.1% sodium dodecyl sulfate (SDS), 50 μg of sonicated salmon sperm DNA per ml, and 35%

* Corresponding author.
† Present address: Institute of Genetics, Fudan University, Shanghai, Peoples Republic of China.

formamide. Hybridization was carried out in the same buffer containing 5×10^5 cpm of the ^{32}P-labeled *Pst*I-*Pvu*II fragments (0.4 and 0.5 kilobase pair [kb]) from the cDNA insertion of human *tk* clone pHtk9 (Lin et al., unpublished data) per ml. The filters remained in the hybridization solution for 48 h at 42°C. The hybridized filters were then washed three times for a total of 30 min in 2× SSC containing 0.1% SDS at room temperature, once for 10 min in 0.1× SSC containing 0.1% SDS at 49°C, and twice for a total of 10 min in 0.1× SSC at room temperature. Hybridizing colonies were picked, replated, and rescreened twice.

Screening of cosmid library. A cosmid library prepared from C57BL/10J mouse liver DNA and inserted in vector pTCF (12) was provided by G. Widera and R. Flavell, Biogen, Cambridge, Mass. This library was constructed by using the method of Grosveld et al. (11), and *E. coli* ED8767 was used as the host. A total of 5×10^5 colonies were screened by using a gel-purified 1.4-kb *Bam*HI fragment from the cDNA insertion of human *tk* plasmid pHtk9 as a probe (Lin et al., unpublished data). Nitrocellulose filters were prehybridized at 42°C for 5 h in a buffer containing 35% formamide, 5× SSC, 4× Denhardt solution, 10 mM Tris hydrochloride (pH 7.5), 0.1% SDS, and 50 μg of sonicated salmon sperm DNA per ml. Hybridization was carried out for 48 h at 42°C in the same buffer containing 7×10^5 cpm of the ^{32}P-labeled probe per ml (31). The hybridized filters were washed three times for a total of 1 h in 2× SSC containing 0.5% SDS at room temperature, three times for a total of 45 min in 0.1× SSC containing 0.5% SDS at 48°C, and twice for a total of 10 min in 0.1× SSC at room temperature. Hybridizing colonies were picked, replated, and rescreened twice.

DNA sequencing. The dideoxy chain termination method of Sanger et al. (32) was used, with the modifications described by Biggin et al. (3). Four shotgun libraries (from each of the *Bam*HI cDNA insertions of pMtk4 and pMtk9), which were generated by partial *Sau*3AI, *Alu*I, *Hae*III, or *Msp*I digestion, were prepared in M13 vectors mp8, mp9, mp18, and mp19 (27). By using this strategy, 80% of the sequence was determined on both strands, and more than 95% of the sequence was determined from at least two independent M13 clones. Sequence data were compiled and analyzed by using the computer systems of Conrad and Mount (8) and Sege et al. (33).

Northern blot analysis. Poly(A)$^+$ mRNA from mouse LTK$^-$ cells was isolated by using procedures described previously (21). RNA samples were prepared and fractionated in an agarose gel containing formaldehyde, essentially by the procedures described by Maniatis et al. (23). The only major modification was that 1× gel-running buffer contained 0.02 M morpholinopropanesulfonic acid (pH 7.0), 5 mM sodium acetate, and 0.1 mM EDTA (pH 8.0). After RNA fractionation, the gel was soaked for 10 min in 0.01 M sodium phosphate (pH 7.0), and the RNA was transferred to nitrocellulose. The blot was then baked for 2 h at 80°C and soaked at 65°C for a minimum of 4 h in a prehybridization solution containing 6× SSC, 50 mM sodium phosphate (pH 7.0), 10× Denhardt solution, and 100 μg of sonicated salmon sperm DNA per ml. The blot was then soaked at the same temperature for 1 to 2 days in a hybridization solution containing 6× SSC, 20 mM sodium phosphate (pH 7.0), 2× Denhardt solution, 100 μg of sonicated salmon sperm DNA per ml, 10% sodium dextran sulfate, and 2×10^6 to 6×10^6 cpm of ^{32}P-labeled pMtk4-1 (specific activity, 1×10^8 to 3×10^8 cpm/μg) per ml. The blot was air dried and allowed to expose X-ray film.

Heteroduplex analysis. Hybridization was performed in a solution containing 70 or 40% formamide, 0.3 M NaCl, 10 mM Tris hydrochloride (pH 8.5), and 0.1 mM EDTA. Genomic DNA and cDNA (each at a concentration of 2 μg/ml) were denatured at 68°C for 15 min and annealed by cooling to 30°C in 30 min. The hybridization mixture was then adjusted to contain final concentrations of either 64 or 30% formamide, 100 mM Tris hydrochloride (pH 8.5), 10 mM EDTA, 40 μg of cytochrome *c* (horse heart type III; Sigma Chemical Co., St. Louis, Mo.) per ml, 0.3 μg of single-stranded ϕX174 DNA per ml, and 0.3 μg of nicked double-stranded pACYC184 DNA per ml (7) before being spread onto double-distilled water. Spreads were transferred to Parlodion-coated grids, stained with uranyl acetate, and shadowed with platinum-palladium. The grids were examined with a Phillips model 300 electron microscope. The sizes of introns and exons were measured by projecting the negatives in a device which utilized a Hewlett-Packard model 9825 calculator-digitizer and programs provided by Jack Griffith.

Southern blot analysis. For Southern blot (34) analysis of *tk* sequences, enzyme-restricted total genomic DNA (20 μg) or cloned cosmid DNA (0.1 μg) was fractionated by electrophoresis through a horizontal 0.8% agarose gel and transferred to nitrocellulose filter paper. The 1.3-kb *Bam*HI fragment of pMtk4 was used as a probe. The blot was prehybridized for approximately 4 h at 65°C in buffer containing 6× SSC, 0.05 M sodium phosphate (pH 7.0), 10× Denhardt solution, and 100 μg of denatured salmon sperm DNA per ml of solution. Hybridization was carried out for approximately 24 h at 65°C in a solution containing 6× SSC, 0.02 M sodium phosphate (pH 7.0), 2 × Denhardt solution, 10% sodium dextran sulfate, 100 μg of denatured salmon sperm DNA per ml, and 2×10^6 to 6×10^6 cpm of labeled probe (specific activity, 1×10^8 to 3×10^8 cpm) per ml. The filters were then washed three times for a total of 1 h at room temperature in 2× SSC containing 0.1% SDS, three to four times for 15 min at 65°C in 0.1× SSC containing 0.1% SDS, and finally three times for 5 min at room temperature in 0.1× SSC. The filters were then allowed to expose X-ray film.

RESULTS

Isolation of functional mouse *tk* cDNAs. We previously isolated human cytoplasmic *tk* genomic clones λtk18 and λtk53 and identified two unique *tk* DNA fragments, a 1.5-kb *Xho*I-*Eco*RI fragment from λtk18 and a 1.25-kb *Hin*dIII-*Kpn*I fragment from λtk53 (22). Using these two DNA segments as probes, we isolated a human *tk* cDNA clone, pHtk9, from the human cDNA library of Okayama and Berg (28; Lin et al., unpublished data). We used the cDNA insertion within pHtk9 to screen a mouse L-cell cDNA library constructed in the mammalian expression vector pcD (a gift from H. Okayama). Two positive clones were retrieved after 200,000 colonies were screened. One of these clones contained a 1.1-kb cDNA insertion, and the other contained a 1.0-kb cDNA insertion; these clones were designated pMtk4 and pMtk9.

To establish whether these two clones contained functional *tk* sequences, we tested their ability to confer upon mouse LTK$^-$ cells the TK$^+$ phenotype by DNA transfection and subsequent growth in medium containing hypoxanthine, aminopterin, and thymidine. The TK$^+$ transformation frequency for pMtk4 was about 8,000 colonies per μg of DNA. The frequency for pMtk9 was low, and we obtained only one TK$^+$ transfectant from 102 μg of plasmid DNA in four

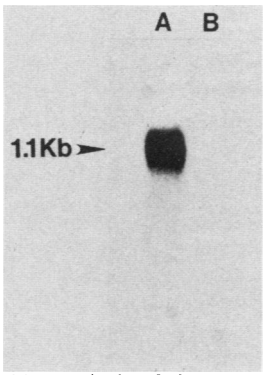

FIG. 1. Northern blot analysis of *tk* mRNA. A 5-μg portion of poly(A)$^+$ RNA was fractionated by electrophoresis in a 1% agarose gel containing 2.2 M formaldehyde. RNA was transferred to nitrocellulose and hybridized with pMtk4. Molecular size was determined by the mobility of rRNA. Lane A, Mouse Ehrlich ascites tumor cells; lane B, mouse LTK$^-$ cells.

experiments. Since both of our cDNA clones can confer the TK$^+$ phenotype upon LTK$^-$ cells, which are deficient in cytoplasmic TK but not in mitochondrial TK, we reasoned that pMtk4 and pMtk9 encode the cytoplasmic form of the enzyme. Further evidence was obtained by demonstrating on polyacrylamide gels (10) that pMtk4 DNA transfection produced TK$^+$ transformants with the cytoplasmic form of TK enzyme activity (data not shown). The TK enzyme activity in the transformant generated by pMtk9 DNA transfection was not detectable by the method which we used, although the transformant evidently has enough TK activity to survive in medium containing hypoxanthine, aminopterin, and thymidine. Southern blot analysis revealed that this pMtk9 transformant contains DNA sequences homologous to *tk* in addition to those found in the original LTK$^-$ cell line (data not shown), indicating that this line is not a revertant or a contaminant. We concluded that pMtk4 contains a functional cytoplasmic *tk* cDNA and that pMtk9 retains an incomplete cytoplasmic *tk* cDNA.

Identification of murine cytoplasmic TK mRNA. To determine whether pMtk4 contained a nearly full-length cDNA insertion and also to study the expression of TK mRNA in normal mouse cells and LTK$^-$ mouse cells (a widely used cell line in somatic cell genetics), poly(A)$^+$ RNA was fractionated in denaturing formaldehyde gels and analyzed by Northern blotting (Fig. 1). Using pMtk4 as a probe, we detected only one band in the RNA from wild-type mouse cells. No band was observed in the lane containing 5 μg of poly(A)$^+$ RNA from LTK$^-$ cells. The size of the apparent cytoplasmic TK mRNA was about 1.1 kb, approximately the same size as the human and Chinese hamster message (4, 20, 21).

The cDNA insertion within pMtk4 is about the same size as the TK mRNA to which it hybridizes, suggesting that it is close to full length. The lack of a corresponding band in LTK$^-$ poly(A)$^+$ RNA further confirms that pMtk4 codes for the cytoplasmic form of TK. The RNA used was intact, since the same RNA blot probed with a human hypoxanthine phosphoribosyltransferase cDNA clone (15) revealed a hybridizing message in both tracks (data not shown).

Nucleotide sequence analysis of the mouse *tk* cDNAs. The sequencing strategy used and a partial restriction map of the cDNA insertions in pMtk4 and pMtk9 are shown in Fig. 2. For more than 95% of each cDNA, the sequence was confirmed by analysis of two independent M13 clones. The

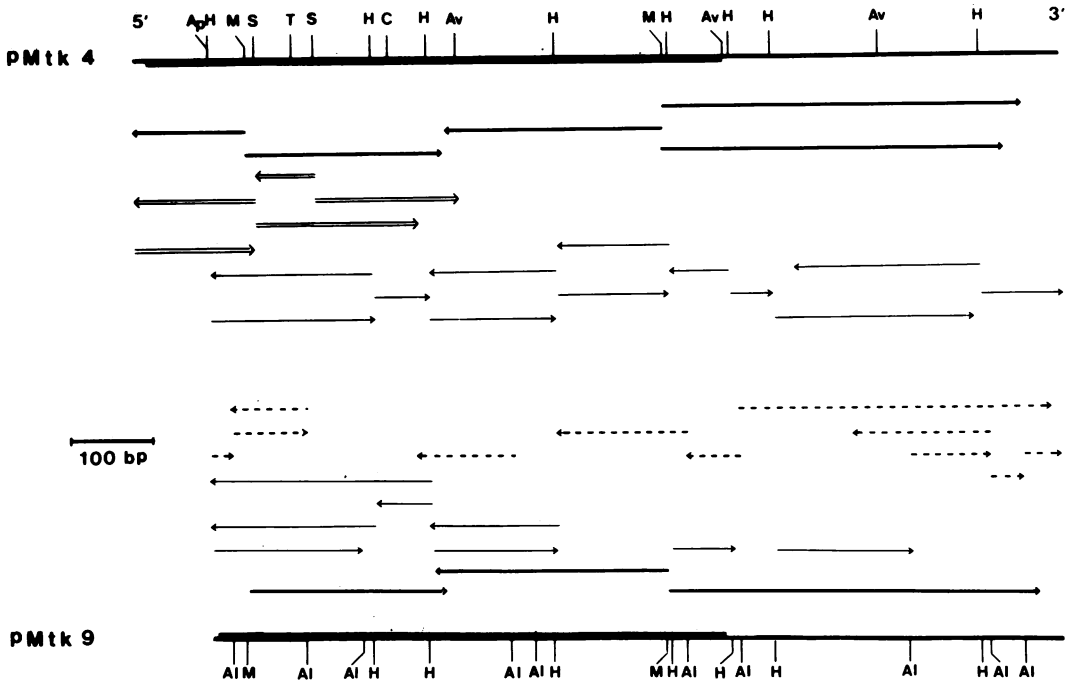

FIG. 2. Sequencing strategy used for the mouse *tk* cDNA insertions in pMtk4 and pMtk9. The thick line represents the coding region. The arrows indicate the directions and extents of sequence determination. M13 clones were constructed from partial *Sau*3AI (⇒), *Msp*I (⟶), *Alu*I (– – →), and *Hae*III (→) digests of the cDNA insertions excised by *Bam*HI. *Apa*I, *Ava*II, *Cla*I, and *Taq*I sites on pMtK4 are also indicated. Al, *Alu*I; Ap, *Apa*I; Av, *Ava*II; C, *Cla*I; H, *Hae*III; M, *Msp*I; S, *Sau*3AI; T, *Taq*I.

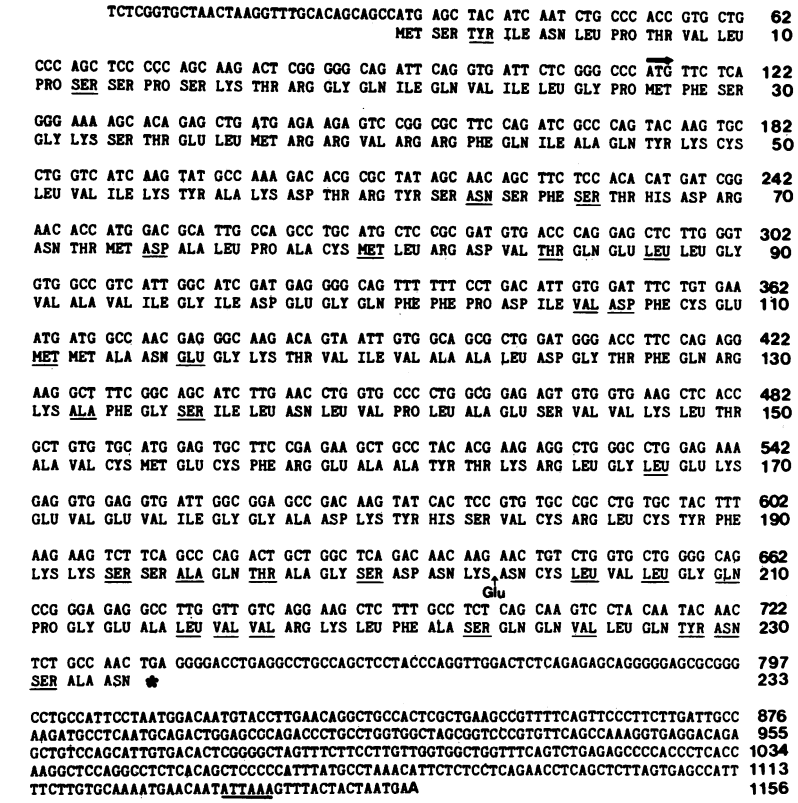

FIG. 3. Nucleotide and deduced amino acid sequences of the mouse cytoplasmic *tk* cDNA in pMtk4. The underlined amino acids differ from those of the previously published human TK sequence (5). The additional glutamic acid within the human TK protein is indicated. The reading frame in pMtk9 may start on nucleotide 114 of pMtk4, which is indicated by an arrow. The putative polyadenylation signal (ATTAAA) is also underlined.

sites for *Cla*I and *Apa*I restriction enzyme cleavage were located previously by standard mapping techniques and were also predicted by the sequence. The complete nucleotide and deduced amino acid sequences of the mouse *tk* cDNA in pMtk4 are shown in Fig. 3. The insertion is 1,156 nucleotides long, not including the polydeoxyadenylate tail at the 3′ end and the polydeoxyguanylate tail at the 5′ end. We presumed that the first methionine (ATG) codon, located at the 33rd nucleotide from the 5′ end of the cDNA, is the translation initiation site. Two other open reading frames resulted in very short polypeptide sequences.

The reading frame remains open for 699 base pairs, closing with an opal (TGA) terminator. The 233 amino acid residues specified by this sequence comprise a polypeptide with a molecular weight of 25,873. The 3′ untranslated region of the cDNA is 422 nucleotides long, with a putative polyadenylation signal (ATTAAA) 22 nucleotides upstream from the poly(A) sequence.

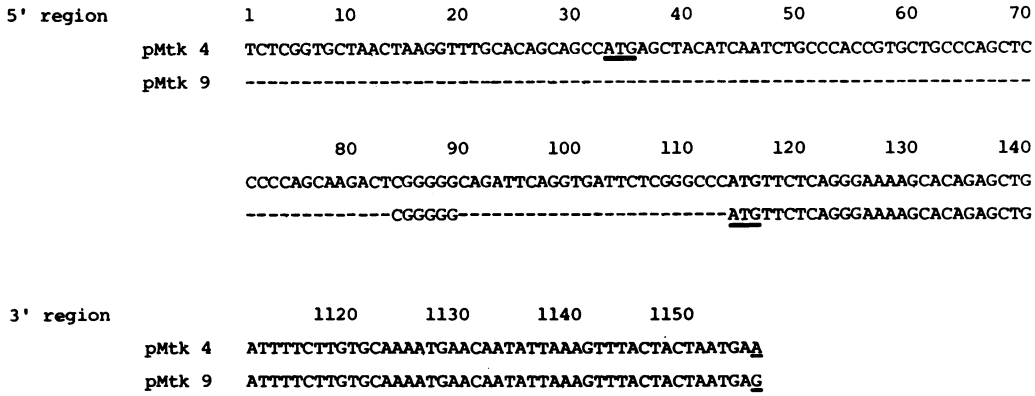

FIG. 4. Comparison of the 5′ and 3′ regions of mouse *tk* cDNAs within pMtk4 and pMtk9. The dashed lines indicate the deleted sequences in pMtk9. The possible translation starting sites (ATG) of pMtk4 and pMtk9 are underlined. The different last bases of these two cDNA clones are also indicated.

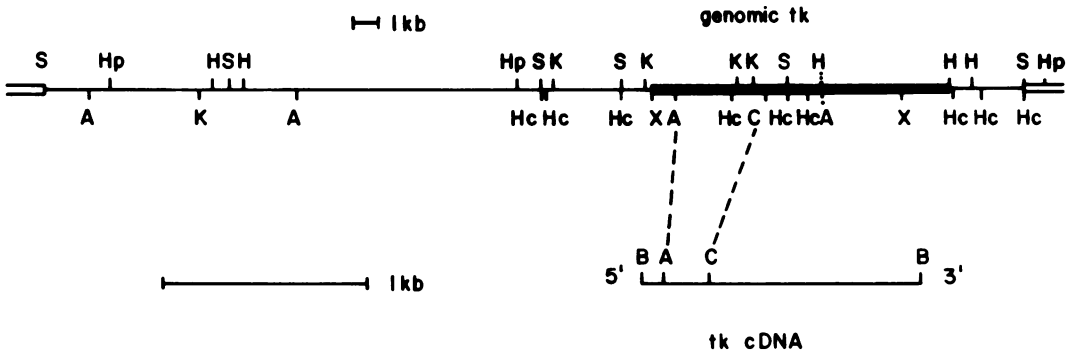

FIG. 5. Restriction map of pMtk116 DNA and preliminary localization of the *tk* coding sequence. The figure shows a restriction map of *tk* cosmid pMtk116 and a simplified map of the *tk* cDNA insertion within pMtk4. The thick solid line indicates the outer boundaries of the region within which hybridization occurs between the cloned genomic DNA and the cDNA probe. The open bars indicate the vector sequence. The *Cla*I site and one of the *Apa*I sites were assigned precisely to corresponding cDNA locations, as indicated. A, *Apa*I; B, *Bam*HI; C, *Cla*I; Hc, *Hin*cII; H, *Hin*dIII; Hp, *Hpa*I; K, *Kpn*I; S, *Sal*I; X, *Xho*I. Not all *Apa*I sites are mapped. *Hin*cII sites are only mapped within the *Hpa*I fragment which contains the *tk* coding region. The dotted line indicates that the exact site is unclear.

A direct comparison of the nucleotide sequences of pMtk4 and pMtk9 showed that differences reside at the 3′ and 5′ ends of the *tk* insertion. As shown in Fig. 4, *tk* cDNA in pMtk9 contains two 5′-end deletions between nucleotides 1 and 83 and between nucleotides 90 and 113, resulting in a sequence 107 nucleotides shorter than pMtk4. In addition, there is a single base change at the last nucleotide of the cDNA insertion.

Isolation of functional *tk* genomic clones. Using the 1.4-kb *Bam*HI insertion of human *tk* cDNA pHtk9 as a probe, we screened a cosmid library made from mouse C57BL/10J liver DNA and obtained four hybridizing clones. These clones were designated pMtk116, pMtk322, pMtk535, and pMtk536.

To establish whether these clones were carrying functional *tk* sequences, mouse LTK$^-$ cells were transfected by the cosmid DNAs and subsequently challenged with medium containing hypoxanthine, aminopterin, and thymidine. Only pMtk116 and pMtk536 were able to confer TK activity (resistance to hypoxanthine, aminopterin, and thymidine) upon LTK$^-$ cells. The transformation frequency was about three TK$^+$ colonies per μg of DNA. Since LTK$^-$ cells are deficient in the cytoplasmic TK, our result suggests that these two cloned sequences encode the functional cytoplasmic TK enzyme. In support of this conclusion, we detected cytoplasmic TK enzyme activity in extracts of LTK$^-$ cells transfected by either pMtk116 or pMtk536 (data not shown). Only cosmid pMtk116 was used for further study because several subclones of pMtk536 showed dissimilar restriction enzyme digestion patterns.

Characterization of the mouse *tk* genomic sequence. We constructed a restriction map of cosmid pMtk116 by using a series of single and double enzyme digests. To measure the gene size and to locate coding sequences, the *tk* cDNA was hybridized to restriction fragments of the genomic cosmid DNA (Fig. 5). The hybridizing genomic restriction fragments extended over an 11-kb region, which therefore corresponded to the approximate size of the gene. All cDNAs constructed by the method described by Okayama and Berg (28) are directionally oriented. By studying the hybridization

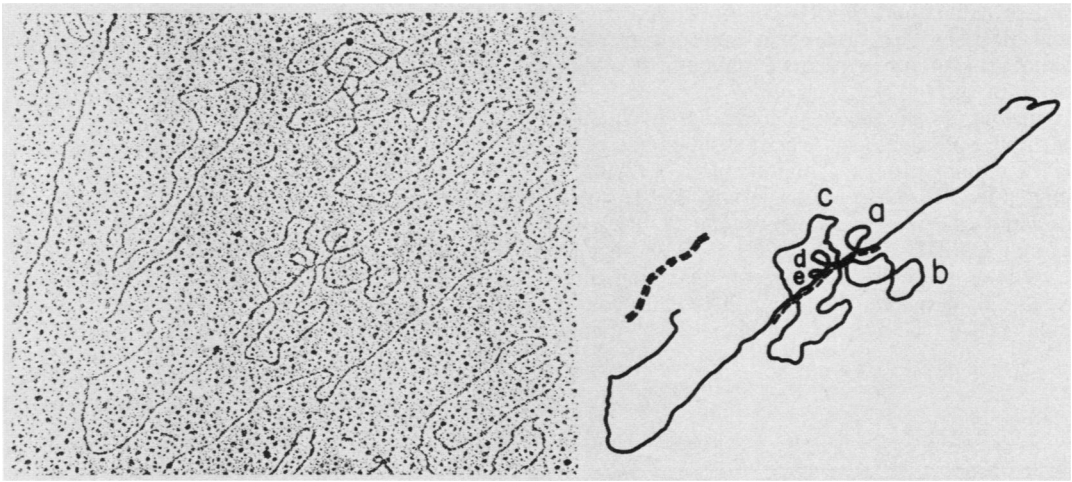

FIG. 6. Electron micrograph of heteroduplexes formed by hybridization of *tk* cDNA clone pMtk4 with *tk* genomic DNA of pMtk116. The cDNA was excised from pMtk4 as a 1.3-kb *Bam*HI fragment. The *tk* genomic DNA was isolated from cosmid pMtk116 as a 22-kb *Hpa*I fragment. The dashed line indicates cDNA insertion DNA. The narrow solid line represents single-stranded genomic DNA. The letters designate introns. Magnification, ×36,644.

TABLE 1. Sizes of exons and introns in the mouse tk gene[a]

Sequence	Size (kb)[b]
Exons	
1	0.12 ± 0.01 (8)
2	0.11 ± 0.02 (13)
3	0.10 ± 0.02 (10)
4	0.10 ± 0.02 (12)
5	0.12 ± 0.02 (16)
6	0.57 ± 0.05 (7)
Introns	
a	0.89 ± 0.11 (16)
b	2.55 ± 0.21 (14)
c	5.29 ± 0.22 (8)
d	0.49 ± 0.05 (15)
e	0.23 ± 0.03 (14)

[a] The exon-intron structure was determined by heteroduplex analysis.
[b] Mean ± standard deviation. In parentheses are the numbers of independent determinations.

patterns of restricted genomic DNA with the cDNA, we were able to assign the *Apa*I and *Cla*I sites, located at the 5′ end of the cDNA, to corresponding sites on the genomic sequence, as shown in Fig. 5. Our results suggest that the promoter region is either within or flanking the 5′-end *Apa*I-*Xho*I fragment in the genomic sequence.

Mapping of exons and introns in the *tk* gene. To determine the exon-intron structure of the murine *tk* gene, we hybridized the 1.3-kb *Bam*HI insertion of cDNA clone pMtk4 to the 22-kb *Hpa*I fragment of the genomic sequence coding for TK. The heteroduplexes formed were examined by electron microscopy, and the lengths of hybridized DNA and unhybridized genomic loops were measured by using a program designed by J. Griffith. As shown in Fig. 6, the *tk* gene contains a minmum of six exons and five introns. Exons less than 50 base pairs (bp) long may not have been detected by this method. Table 1 shows the sizes of the exons and introns. These results also confirmed our Southern blot data which indicated that the *tk* gene is approximately 11 kb long. The gene structure determined from the measurements is shown in Fig. 7. The orientation of the cDNA with respect to the genomic sequence, as previously judged by Southern blotting and restriction enzyme analysis (Fig. 5), was confirmed by examining heteroduplexes formed between the *Hpa*I *tk* genomic fragment and *Hin*dIII-cut, linearized, intact cDNA plasmid pMtk4. The first exon and intron were defined by locating the vector segments flanking the 5′ end of the cDNA (data not shown).

***tk*-Specific sequences in the mouse genome.** To investigate whether the structure of the mouse *tk* gene isolated had been altered during the cloning process, genomic blots of *Hin*dIII-digested cosmid pMtk116 and various mouse DNAs were probed with pMtk4 cDNA. As shown in Fig. 8, the two fragments (4.7 and 23 kb) that were expected to hybridize to pMtk4 on the basis of the restriction map of the cloned gene (Fig. 5) were found in normal genomic DNA but not in LTK$^-$ genomic DNA. This result implies that cosmid pMtk116 contains a near-normal *tk* gene structure and that no gross artifactual rearrangements occurred during molecular cloning. It also revealed that the structural gene for cytoplasmic TK is deleted in mouse LTK$^-$ cells.

Interestingly, in addition to these two functional *tk*-containing DNA segments, we detected two other *tk*-related sequences, 3.7 and 9.3 kb long, in mouse genomic DNA. These sequences are present in the DNAs of all four inbred mouse strains studied, including strains C57BL/10J, C3H/HeJ, A/HeJ, and BALB/cJ (not all data shown). The presence of these *tk*-related DNA sequences in mouse LTK$^-$ cells suggests that they are not functional cytosolic *tk* genes. We also believe that they do not represent mitochondrial *tk* gene sequences for reasons described in the Discussion. These bands could represent fragments of other *tk*-related genes or pseudogenes.

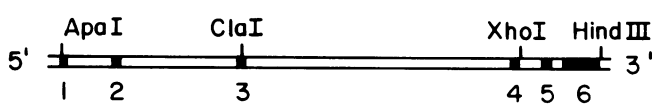

FIG. 7. Exons and introns in the murine *tk* gene. The numbers designate exons.

DISCUSSION

We isolated two functional mouse cytoplasmic *tk* cDNAs. An analysis of the nucleotide sequences of these cDNAs implied that the TK mRNA has a 5′ untranslated region of at least 32 bp, a coding region of 699 bp, and a 3′ untranslated region of 422 bp. The nucleotide sequence of the cDNA within pMtk4 predicts a TK protein of 233 amino acids, with a molecular weight of 25,873. This closely agrees with the sizes predicted for the human TK protein (234 amino acids) and chicken TK protein (224 amino acids) (5, 24). As determined by sedimentation coefficient measurements in glycerol gradients, the catalytically active murine TK has an estimated molecular weight of 86,400 (17), suggesting that it functions as a multimer.

A comparison of the nucleotide sequences of mouse,

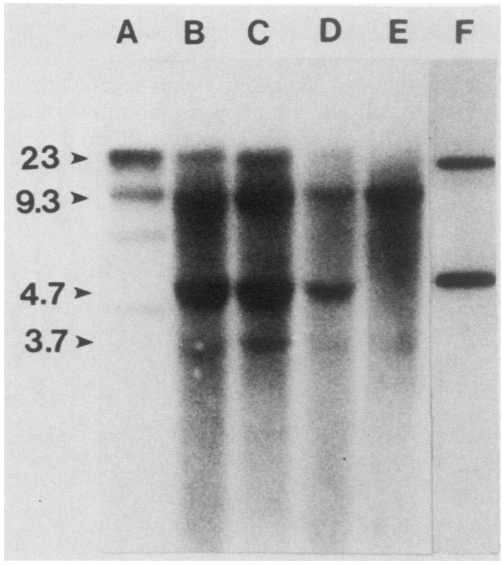

FIG. 8. Distribution of *tk*-specific sequences in *Hin*dIII-digested DNAs from various mouse strains. Samples of genomic mouse DNA (20 μg each) and *tk* cosmid DNA (0.1 μg) were digested to completion with *Hin*dIII, fractionated by electrophoresis through a 0.8% agarose gel, transferred to nitrocellulose, and hybridized with the *Bam*HI insertion of mouse *tk* cDNA pMtk4. Lane A, Phage λ DNA; lane B, C57BL/10J; lane C, C3H/HeJ; lane D, L929; lane E, LTK$^-$; lane F, pMtk116. Cosmid pMtk116 was derived from a C57BL/10J mouse. L929 and LTK$^-$ cells originated from a C3H mouse. L929 expresses cytosolic TK activity, and LTK$^-$ is deficient in that activity.

human (5), and chicken (18, 24) *tk* genes reveals a striking conservation. The amino acid differences between human and mouse TK proteins are indicated in Fig. 3. The human TK enzyme has one extra amino acid (glutamic acid) encoded between nucleotides 203 and 204 (Fig. 3). The level of homology of amino acid residues between human and mouse TK proteins is about 87%, and the coding nucleotide sequence shows 86% homology. Mouse TK and chicken TK share 69% of their amino acid residues and 70% of their coding nucleotides. It is interesting to note that although the major amino acid divergence between mouse TK and human TK occurs at the carboxy terminus, the mouse and chicken proteins differ at both amino and carboxy termini. Little homology was found between mouse TK and herpes simplex viral TK. The extent and position of the TK homologies among these species closely reflect their evolutionary relationships.

The truncated nucleotide sequence of pMtk9, compared with pMtk4, may be derived either from incomplete polymerization by reverse transcriptase due to secondary structures in TK mRNA or from differential splicing of the original TK message. One explanation for the fact that pMtk9 expresses weak TK activity without the first 27 amino acids is that the second methionine codon at the corresponding nucleotide position of pMtk4 (position 114) (Fig. 3 and 4) or the ATG codon in the simian virus 40 sequence of vector pcD is used as a translation starting site. Alternatively, detectable activity could also have been due to the low frequency with which the plasmid recombined with sequences in the LTK$^-$ genome capable of restoring *tk* function. The low transforming activity of pMtk9 seems to favor the latter hypothesis.

In this report, we also describe a functional genomic sequence that encodes the mouse cytosolic TK isolated directly from a mouse C57BL/10J cosmid library. The structure of the gene is near normal, and no gross rearrangements were generated during the cloning process. The physical map of the *tk* gene, based on a heteroduplex analysis between the genomic sequence and a cDNA clone derived from mouse L cells, reveals the presence of at least five intervening sequences and six exons, spanning at least 11 kb. The largest intron encompasses a 5.3-kb region. The longest exon contains 0.57 kb of DNA and encodes the carboxyl-terminal domain of the polypeptide.

TABLE 2. Lengths of protein-coding sequences in mouse and chicken *tk* exons

Species	Exon	Length (kb)[a]
Mouse	1	0.09
	2	0.11
	3	0.10
	4	0.10
	5	0.12
	6	0.15
Chicken[b]	1	0.066
	2	0.033
	3	0.114
	4	0.093
	5	0.090
	6	0.120
	7	0.156

[a] The lengths of the protein-coding sequences in the first and sixth exons of mouse *tk* were calculated by subtracting the untranslated nucleotides (see Fig. 3) from the lengths of the exons (see Table 1).

[b] Data from references 18 and 24.

The chicken *tk* gene consists of six introns and seven exons, as judged by a nucleotide sequence analysis of genomic and cDNA copies of the gene (18, 24). A comparison of the sizes of the exons within the mouse and chicken *tk* genes (Table 2) revealed a surprising conservation of the length of the protein-coding regions comprising the five exons near the 3' end. The protein-coding region of the first exon in the mouse *tk* gene is 0.09 kb long, whereas the protein-coding domains of the first and second exons of the chicken *tk* are 0.066 and 0.033 kb long, respectively. Since 0.05-kb exons approach the limits of detection by heteroduplex analysis, it is possible that the first mouse exon could ultimately be resolved into two exons. A direct sequence analysis of the genomic DNA will be required to answer this question.

Mouse LTK$^-$ cells, which are defective in cytoplasmic TK but not mitochondrial TK, are widely used in somatic cell genetics due to their nonrevertible and selectable phenotype (2, 26). The availability of cloned mouse *tk* cDNA sequences enabled us to demonstrate the absence of TK message in these cells, as well as the absence of cytoplasmic *tk* gene sequences. These experiments also established the existence of *tk*-related sequences in various inbred mouse strains, but the identities of these fragments have yet to be defined. They are not likely to be sequences of the mitochondrial *tk* gene because no RNA sequences homologous to the cytosolic *tk* cDNA are detectable in the poly(A)$^+$ RNA of LTK$^-$ cells (which express mitochondrial TK) by Northern blot analysis, even under low-stringency hybridization conditions (data not shown). Therefore, the strong intensity of these bands in the genomic blot of LTK$^-$ DNA, together with the undetectable signal in Northern blots, excludes the possibility that they are of mitochondrial *tk* origin. These sequences must be isolated and further characterized to determine whether they are truly pseudogenes.

In summary, we isolated and characterized cDNA and genomic clones encoding mouse cytosolic TK. The structural information obtained from these studies, as well as from other studies, should facilitate the identification of control elements that regulate TK expression during proliferation and differentiation.

ACKNOWLEDGMENTS

We thank H. Okayama and P. Berg for providing their human and mouse cDNA expression libraries and G. Widera and R. Flavell for providing the mouse cosmid library. We also thank P. Howard-Flanders and A. McClelland for advice concerning computer analysis of DNA sequences; W. McGinnis for the use of his IBM-XT computer; B. Piekos, P. Howard-Flanders, E. Cassuto, J. Flory, and Keith MisKimins for advice concerning heteroduplex analysis; H. Samanta for the gift of Ehrlich ascites tumor poly(A)$^+$ RNA; and P. Barker for critical reading of the manuscript. We thank M. Newman and V. Salerno for technical assistance, S. Pafka for photography, and M. Siniscalchi and M. Reger for typing the manuscript.

This research was supported by Public Health Service grant GM09966 from the National Institutes of Health. During part of this study, H.B.L. was supported by Public Health Service postdoctoral training grant 5-T32-GMO7439 from the National Institutes of Health. D.-B.Y. was partially supported by the Ministry of Education, Taiwan, Republic of China.

ADDENDUM IN PROOF

DNA sequence analysis has resolved the first exon into two protein-coding regions. Therefore, the mouse *tk* gene contains at least six introns and seven exons.

LITERATURE CITED

1. **Adler, R., and B. R. McAuslan.** 1974. Expression of thymidine kinase variants is a function of the replicative state of cells. Cell **2**:113–117.
2. **Attardi, B., and G. Attardi.** 1972. Persistance of thymidine kinase activity in mitochondria of a thymidine kinase-deficient derivative of mouse L cells. Proc. Natl. Acad. Sci. USA **69**:2874–2878.
3. **Biggin, M. D., T. J. Gibson, and G. F. Hong.** 1983. Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA **80**:3963–3965.
4. **Bradshaw, H. D., Jr.** 1983. Molecular cloning and cell cycle-specific regulation of a functional human thymidine kinase gene. Proc. Natl. Acad. Sci. USA **80**:5588–5591.
5. **Bradshaw, H. D., Jr., and P. L. Deininger.** 1984. Human thymidine kinase gene: molecular cloning and nucleotide sequence of a cDNA expressible in mammalian cells. Mol. Cell. Biol. **4**:2316–2320.
6. **Calos, M. P., J. S. Lebkowski, and M. R. Botchan.** 1983. High mutation frequency in DNA transfected into mammalian cells. Proc. Natl. Acad. Sci. USA **80**:3015–3019.
7. **Chang, A. C. Y., and S. N. Cohen.** 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. J. Bacteriol. **134**:1141–1156.
8. **Conrad, B., and D. W. Mount.** 1982. Microcomputer programs for DNA sequence analysis. Nucleic Acids Res. **10**:31–38.
9. **Ellims, P. H., M. B. Von der Weyden, and G. Medley.** 1981. Thymidine kinase isozymes in human malignant lymphoma. Cancer Res. **41**:691–695.
10. **Goff, S. P., and P. Berg.** 1979. Construction, preparation, and expression of SV40 recombinant genomes containing the *Escherichia coli* gene for thymidine kinase and a *Saccharomyces cerevisiae* gene for tyrosine transfer RNA. J. Mol. Biol. **133**:359–383.
11. **Grosveld, F. G., H. H. M. Dahl, E. deBoer, and R. A. Flavell.** 1981. Isolation of β-globin-related genes from a human cosmid library. Gene **13**:227–337.
12. **Grosveld, F. G., T. Lund, E. J. Murray, A. L. Mellor, H. H. M. Dahl, and R. A. Flavell.** 1982. The construction of cosmid libraries which can be used to transform eukaryotic cells. Nucleic Acids Res. **10**:6715–6732.
13. **Hanahan, D.** 1983. Studies on transformation of *E. coli* with plasmids. J. Mol. Biol. **166**:557–580.
14. **Hanahan, D., and M. Meselson.** 1980. A protocol for high density screening of plasmids in χ1776. Gene **10**:63–67.
15. **Jolly, D. J., H. Okayama, P. Berg, A. C. Esty, D. Filpula, P. Bohlen, G. G. Johnson, J. E. Shively, T. Hunkapillar, and T. Friedman.** 1983. Isolation and characterization of a full-length expressible cDNA for human hypoxanthine phosphoribosyltransferase. Proc. Natl. Acad. Sci. USA **80**:477–481.
16. **Kit, S.** 1976. Thymidine kinase, DNA synthesis, and cancer. Mol. Cell. Biochem. **11**:161–182.
17. **Kit, S., W. C. Leung, D. Trkula, and G. Jorgensen.** 1974. Gel electrophoresis and isoelectric focusing of mitrochondrial and viral-induced thymidine kinase. Int. J. Cancer **13**:203–218.
18. **Kwoh, T. J., and J. A. Engels.** 1984. The nucleotide sequence of the chicken thymidine kinase gene and the relationship of its predicted polypeptide to that of the vaccinia virus thymidine kinase. Nucleic Acids Res. **12**:3959–3971.
19. **Lau, Y.-F., and Y.-W. Kan.** 1984. Direct isolation of the functional human thymidine kinase gene with a cosmid shuttle vector. Proc. Natl. Acad. Sci. USA **81**:414–418.
20. **Lewis, J. A., K. Shimizu, and D. Zipser.** 1983. Isolation and preliminary characterization of the Chinese hamster thymidine kinase gene. Mol. Cell. Biol. **3**:1815–1823.
21. **Lin, P. F., M. Yamaizumi, P. D. Murphy, A. Egg, and F. H. Ruddle.** 1982. Partial purification and characterization of the mRNA for human thymidine kinase and hypoxanthine-guanine phosphoribosyl transferase. Proc. Natl. Acad. Sci. USA **79**:4290–4294.
22. **Lin, P. F., S.-Y. Zhao, and F. H. Ruddle.** 1983. Genomic cloning and preliminary characterization of the human thymidine kinase gene. Proc. Natl. Acad. Sci. USA **80**:6528–6532.
23. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning, a laboratory manual, p. 202–203. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
24. **Merrill, G. F., R. E. Harland, M. Groudine, and S. L. McKnight.** 1984. Genetic and physical analysis of the chicken *tk* gene. Mol. Cell. Biol. **4**:1769–1776.
25. **Merrill, G. F., S. D. Hauschka, and S. L. McKnight.** 1984. TK enzyme expression in differentiating muscle cells is regulated through an internal segment of the cellular *tk* gene. Mol. Cell. Biol. **4**:1777–1784.
26. **Migeon, B. R., S. W. Smith, and L. L. Leddy.** 1969. The nature of thymidine kinase in the human-mouse hybrid cell. Biochem. Genet. **3**:583–590.
27. **Norrander, J., T. Kempe, and J. Messing.** 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene **26**:101–106.
28. **Okayama, H., and P. Berg.** 1983. A cDNA cloning vector that permits expression of cDNA inserts in mammalian cells. Mol. Cell. Biol. **3**:280–289.
29. **Perucho, M., D. Hanahan, L. Lipsich, and M. Wigler.** 1980. Isolation of the chicken thymidine kinase gene by plasmid rescue. Nature (London) **285**:207–210.
30. **Razzaque, A., H. Mizusawa, and M. M. Seidman.** 1983. Rearrangement and mutagenesis of a shuttle vector plasmid after passage in mammalian cells. Proc. Natl. Acad. Sci. USA **80**:3010–3014.
31. **Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. J. Berg.** 1977. Labeling deoxyribonucleic acid to high specific activity *in vitro* with DNA polymerase I. J. Mol. Biol. **113**:237–251.
32. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
33. **Sege, R. D., D. Söll, F. H. Ruddle, and C. Queen.** 1981. A conversational system for the analysis of nucleic acid sequences. Nucleic Acids Res. **9**:437–444.
34. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.