MOLECULAR AND CELLULAR BIOLOGY, June 1986, p. 1998–2010
0270-7306/86/061998-13$02.00/0

Vol. 6, No. 6

Structure and Expression of the Chinese Hamster Thymidine Kinase Gene

JOHN A. LEWIS†

Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724

Received 15 October 1985/Accepted 21 February 1986

My colleagues and I have cloned a nearly full-length Chinese hamster thymidine kinase (TK) cDNA in a λgt10 vector and characterized this cDNA by nucleotide sequencing. The hamster TK protein is encoded in this cDNA by a 702-base-pair open reading frame which specifies a 25,625-dalton protein closely homologous to the previously described human and chicken TK proteins. Using cDNA nucleotide sequence data in conjunction with sequence data derived from selected subclones of the hamster TK gene recombinant phage λHaTK.5, we have resolved the structure of the TK gene, finding the 1,219 base pairs of the cDNA sequence to be distributed through 11.2 kilobases of genomic DNA in at least seven exon segments. In addition, we have constructed a variety of Chinese hamster TK minigenes and exonuclease III-S1 derivatives of these genes which have permitted us to define the limits of the Chinese hamster TK gene promoter and demonstrate that efficient TK transformation of Ltk$^-$ cells by TK minigenes depends on the presence of both TK intervening sequences and sequences 3′ to the site of mRNA polyadenylation.

The enzyme thymidine kinase (TK) (EC 2.1.7.21), which salvages thymidine for DNA synthesis through an ATP-dependent phosphorylation, is one of a number of enzymes important to nucleotide metabolism whose expression is cell cycle S-phase specific (for a review see Mitchison [59]). The TK activity in synchronously growing in vitro cell cultures, for example, rises sharply as cells enter and progress through S phase and then diminishes as the cells move through the G_2, M, and G_1 phases of the cell division cycle (10, 38, 39, 60, 72, 73). As a reflection of the S phase specificity of TK gene expression, TK activity in asynchronously growing cell cultures is growth phase dependent, being maximal as cultures grow through mid-log phase and minimal as these cultures reach confluence, withdraw from the cell cycle, and accumulate in the G_1/G_0 phase (2, 16, 34, 37, 49, 50, 60, 64). The genetic determinants which govern this pattern of TK gene expression have not yet been precisely defined. These determinants are presumed, however, to be closely linked to TK structural gene sequences since TK gene expression in mouse Ltk$^-$ cells transformed to Tk$^+$ with either total mouse genomic DNA or a cloned human TK gene has been shown to be S phase specific (8, 68).

The recent clonings of the hamster (47), human (7, 44, 48), and chicken (63) TK genes invite the possibility that the genetic determinants which underlie the cell cycle and growth phase dependence of TK gene expression can be localized within segments of these cloned TK genes through a systematic in vitro construction of TK gene mutants which can be analyzed for their pattern of expression after transfection and stable integration into a suitable Tk$^-$ host cell. My colleagues and I have previously reported the isolation and preliminary characterization of the Chinese hamster TK gene, carried in the Chinese hamster genome on chromosome 7 (70). In this report I describe (i) the molecular cloning and nucleotide sequencing of a nearly full-length Chinese hamster TK cDNA, (ii) the nucleotide sequencing of selected subclones of the λHaTK.5 clone which has led us to a resolution of the exon-intron organization of the hamster TK gene, and (iii) the construction of various hamster TK minigenes with which we have analyzed the structural determinants of hamster TK gene expression in mouse Ltk$^-$ cells. In the accompanying manuscript (46), D. A. Matkovich and I describe the growth phase dependence and adenovirus responsiveness of TK gene expression in rat 4 cells transformed to Tk$^+$ with various chimeric hamster TK minigenes.

MATERIALS AND METHODS

Isolation of cytoplasmic poly(A)$^+$ RNA. Cytoplasmic RNA was extracted by the urea-sodium dodecyl sulfate method of Holmes and Bonner (32) from postnuclear supernatants prepared from monolayer cultures of the Chinese hamster ovary cell line A-29 (20) growing asynchronously in mid-log phase, 72 h after an initial seeding at a density of 5×10^5 cells per 100-mm dish. Polyadenylated [poly(A)$^+$] RNA was selected from total cytoplasmic RNA by oligo(dT)-cellulose chromatography essentially as described (45).

cDNA synthesis. First- and second-strand cDNA synthesis was accomplished using avian myeloblastosis virus reverse transcriptase (Life Sciences) essentially as described by Helfman et al. (31). S1 nuclease (Boehringer-Mannheim)-digested double-stranded cDNAs were size fractionated by column chromatography on Sepharose 4B-CL equilibrated in 10 mM Tris hydrochloride (pH 7.9)–300 mM NaCl–5 mM EDTA. cDNAs larger than 500 base pairs (bp) were methylated with 5 U of *Eco*RI methylase (New England BioLabs) per μg under reaction conditions specified by the vendor and ligated at 40 μg/ml with T4 DNA ligase (New England BioLabs) in 50 μl of a buffer of 10 mM Tris hydrochloride (pH 7.6)–10 mM $MgCl_2$–1.0 mM ATP–1 mM spermidine–2 mM dithiothreitol at 4°C for 20 h with an equivalent mass of synthetic 8-mer *Eco*RI linkers (GGAATTCC; Collaborative Research). Linkers were prepared for ligation by phosphorylation with polynucleotide kinase (P.L. Biochemicals) un-

† Present address: Department of Virus and Cell Biology, Merck, Sharp & Dohme Research Laboratories, West Point, PA 19486.

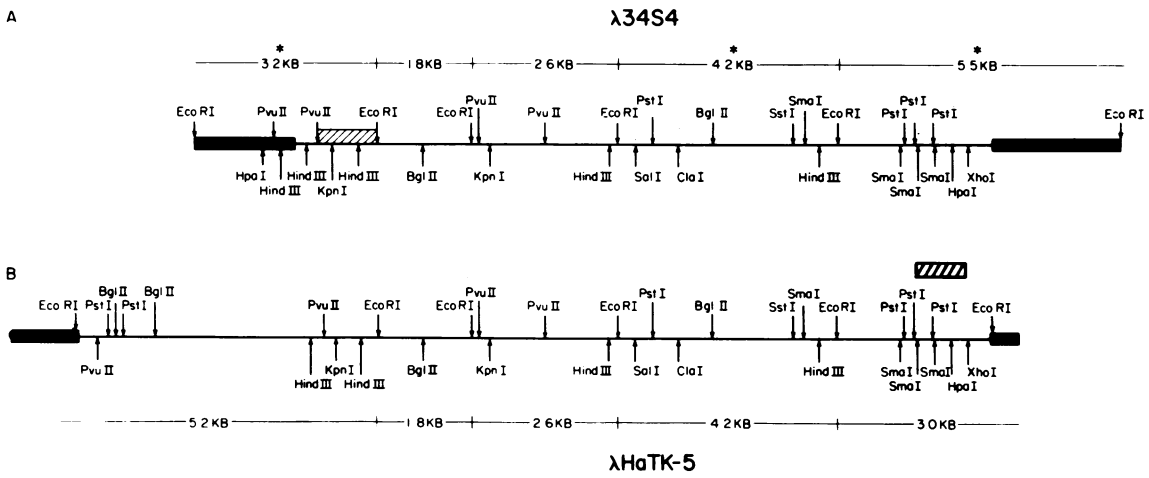


FIG. 1. Partial restriction maps of recombinant phage containing the Chinese hamster TK gene. (A) λ34S4 in an L47 recombinant containing the Chinese hamster TK gene within a 12.2-kb *Bcl*I insert fragment cloned from genomic DNA of a mouse Ltk$^-$ cell line transformed to Tk$^+$ with DNA from the Chinese hamster ovary cell line A-29 (47). The three *Eco*RI fragments which hybridize in Southern blot analysis with the chicken TK gene (43, 58) are indicated by asterisks. The 900-bp *Eco*RI-*Pvu*II fragment of the 3.2-kb *Eco*RI fragment used as a Benton and Davis hybridization probe to screen the Chinese hamster ovary A-29 λGT10 cDNA library is cross-hatched. (B) λHaTK.5 is an EMBL 4 recombinant phage containing the Chinese hamster TK gene within a 17-kb *Bcl*I fragment cloned from size-fractionated, *Bcl*I-digested DNA of the Chinese hamster ovary cell line A-29. The details of this cloning are described in Materials and Methods. Insert DNA sequences are drawn as thin lines; phage DNA sequences are indicated as heavy lines.

der reaction conditions specified by the vendor. cDNA-linker ligation products were digested extensively with *Eco*RI (New England BioLabs), phenol-chloroform (1:1) extracted, and concentrated by ethanol precipitation.

Recombinant phage construction and screening. λgt10 phage were propagated to high titer in NZCYM (52) broth cultures of *Escherichia coli* Bu881, concentrated by polyethylene glycol 6000 precipitation, and banded to equilibrium on CsCl gradients as described (81). λgt10 DNA was purified by sequential extractions with sodium dodecyl sulfate-proteinase K (Boehringer-Mannheim) at 100 μg/ml in 10 mM Tris hydrochloride (pH 7.4)–100 mM NaCl–5 mM EDTA–0.2% sodium dodecyl sulfate and phenol-choroform (1:1) and concentrated by ethanol precipitation. *Eco*RI-digested λgt10 DNA was ligated to double-stranded cDNA at a concentration of 250 μg/ml at 9°C for 24 h with T4 DNA ligase. The ligation reaction was heated to 70°C, cooled slowly to room temperature, and packaged in vitro using freeze-thaw and sonic extracts prepared by the methods of Enquist and Sternberg (17). The efficiency of recombinant phage construction, which ranged as high as 5% at a vector-to-cDNA mass ratio of 40:1, was determined by titration on *E. coli* C600. Recombinant phage were plaqued at a density of 3×10^3 per 90-mm dish on *E. coli* C600 Hfl-A to suppress the growth of nonrecombinant phage (33) and screened by the plaque hybridization procedure of Benton and Davis (4) as modified by S. Woo (80). A 900-bp *Eco*RI-*Pvu*II subclone of the 3.2-kilobase (kb) *Eco*RI fragment of λ34S4 was radiolabeled by nick-translation with ^{32}P as described by Rigby et al. (66). DNAs of recombinant λgt10 phage identified by hybridization were prepared from 50-ml liquid lysate NZCYM cultures (47), digested with *Eco*RI, and analyzed by Southern blot hybridization to estimate the size of the recombinant inserts and confirm the original positive hybridization reaction.

Nucleotide sequence analysis. Nucleotide sequence data were generated using the dideoxynucleotide chain-terminating method of Sanger et al. (67). Radiolabel was incorporated in the sequencing reactions through the use of deoxyadenosine 5′-(α-[^{35}S]thio)triphosphate (New England Nuclear Corp.) essentially as described by Biggin et al. (5). The Klenow fragment of *E. coli* polymerase I was obtained from Bethesda Research Laboratories. Restriction enzyme fragments of the cDNA insert of λcTK.90 and appropriate fragments of the recombinant phage λHaTK.5 were subcloned in the vectors M13mp10 and M13mp11, obtained as replicate-form DNA from New England BioLabs. Recombinant M13 phage were propagated on the host strain *E. coli* K-12 JM101 for the preparation of single-stranded sequencing template essentially as described (67). Four 15-mer oligonucleotides were prepared by an automated phosphite method using an Applied Biosystems 380A DNA synthesizer (1). The sequences of these oligonucleotides were based on regions of the cDNA which had previously been sequenced on both strands.

Construction of hamster TK minigenes. The Chinese hamster TK minigenes presented in Fig. 4 were constructed as described below. All restriction and modification enzymes were products of New England BioLabs or International Biotechnologies Inc. and were used under reaction conditions specified by the vendor. In general, molecular ligations combined insert and vector fragments isolated from agarose gels by electroelution and purified by benzoylated naphthoylated DEAE cellulose chromatography (Sigma). Recombinant plasmids were cloned in *E. coli* DH.1 essentially as described by Hanahan (29), using ampicillin (100 mg/ml) selection, and amplified as described (11).

The Chinese hamster TK minigene designated pHaTK.1 was constructed in the vector pBA (a derivative of pXf3 carrying a 102-bp polylinker including an *Xho*I site inserted between the *Eco*RI and *Bam*HI sites of pXf3) (52). A recombinant pBA vector carrying the 330-bp *Eco*RI cDNA fragment of λcTK.90 in the appropriate orientation was digested with *Xho*I and *Sma*I (New England BioLabs) and recombined with a 1.6-kb *Xho*I-partial *Sma*I digestion fragment isolated from the 3.0-kb *Eco*RI fragment of λHaTK.5

FIG. 2. Nucleotide sequence of the 1,219-bp insert of the hamster TK cDNA λcTK.90 and the amino acid sequence predicted for the 702-bp open reading frame. The numbered sequence excludes the *Eco*RI linkers at the 5′ and 3′ extremes of the λcTK.90 cDNA and the oligo(dA) tail of some 30 bp.

(see Fig. 1), generating the recombinant plasmid p9SXX containing a unique *Eco*RI site. The construction of pHaTK.1 was completed by cloning the 910-bp *Eco*RI cDNA fragment of λcTK.90 into p9SXX.

pHaTK.1 was modified to pHaTK.1a by (i) digesting pHaTK.1 with *Xho*I and *Hpa*I, (ii) rendering the *Xho*I site blunt ended by reaction with the Klenow fragment of *E. coli* polymerase I, and (iii) religating the Klenow reaction products with T4 DNA ligase. pHaTK.2 was constructed by combining the 5.2-kb *Eco*RI fragment of λHaTK.5 with the p9SXX vector described above. pHaTK.1b was constructed by recombining the *Hin*dIII-*Xho*I insert fragment of pHaTK.1 with the *Hin*dIII-*Xho*I vector fragment of pHaTK.2. pHaTK.2a was modified from pHaTK.2 as described for pHaTK.1 above. pHaTK.2b was constructed by cloning the *Stu*I-*Xho*I fragment of pHaTK.2 into the *Stu*I-*Xho*I vector fragment of pHaTK.1. pHaTK.3 was constructed by cloning an *Sst*I-*Xho*I fragment of λHaTK.5 into the *Sst*I-*Xho*I vector fragment of pHaTK.1. pHaTK.3a was constructed by cloning a *Cla*I-*Xho*I insert fragment of pHaTK.3 into the *Cla*I-*Xho*I vector fragment of pHaTK.1b. All recombinant plasmids were redigested with the appropriate enzymes to insure the integrity of the restriction enzyme sites used for construction.

Construction of exoIII-S1 deletion mutants. Approximately 10 μg of *Hpa*I (New England BioLabs)-digested pHaTK.2 was digested at 20°C for various times up to 15 min with 100 U of exonuclease III (exaIII; New England BioLabs) per μg in a buffer of 10 mM Tris hydrochloride (pH 7.4)–50 mM NaCl–5 mM $MgCl_2$. The reaction products were diluted 10-fold into S1 nuclease buffer (30 mM sodium acetate, pH 4.5, 10 mM $ZnSO_4$, 300 mM NaCl), digested at 20°C for 30 min with 1,000 U of S1 nuclease (Boehringer-Mannheim) per μg, phenol-chloroform (1:1) extracted, and concentrated by ethanol precipitation. The extent of exoIII-S1 digestion of pHaTK.2 was estimated by agarose gel electrophoresis after *Eco*RI or *Hin*dIII digestion. Plasmid samples with appropriately sized deletions were ligated at 4°C for 20 h with an equivalent mass of 8-mer *Xho*I linkers (CCTCGAGG; Collaborative Research) prepared for ligation by phosphorylation with polynucleotide kinase. The ligation products were digested extensively with *Xho*I, phenol-chloroform (1:1) extracted, and concentrated by ethanol precipitation. The *Xho*I digestion products were ligated at 4°C for 12 h at a concentration of 1 μg/ml with T4 DNA ligase and cloned by transformation into *E. coli* DH.1 using ampicillin (100 μg/ml) selection. Since the pHaTK.2 minigene contains an endogenous *Xho*I site 1.5 kb 5′ to the *Hpa*I site, this mutagenesis scheme left the deletion endpoint of each mutant juxtaposed to the same plasmid vector sequences.

Gene isolation. High-molecular-weight genomic DNA from the Chinese hamster ovary cell line A-29 was digested with *Bcl*I (New England BioLabs) and size fractionated on 10 to 40% sucrose gradients prepared in 10 mM Tris hydrochloride (pH 7.4)–100 mM NaCl–5 mM EDTA. *Bcl*I restriction fragments greater than 15 kb in size were ligated with T4 DNA ligase to sucrose gradient-purified, *Bam*HI-digested arms of the λ cloning vector EMBL 4 and packaged in vitro as described above. Recombinant phage were screened by the plaque hybridization procedure of Benton and Davis (4) using the 4.2- and 2.6-kb *Eco*RI fragments of the isolate λ34S4 as nick-translated probes.

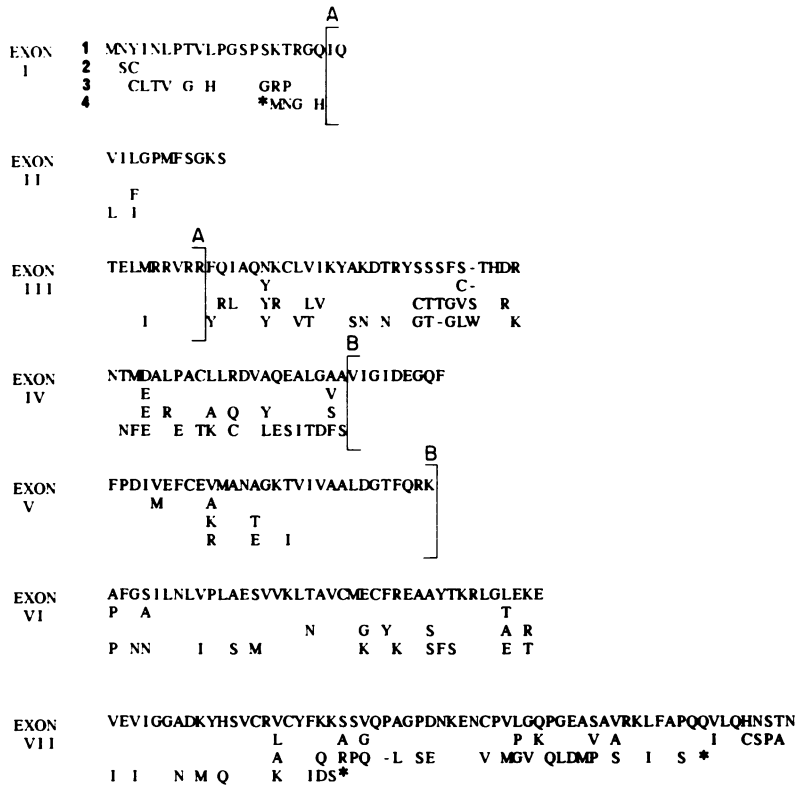


FIG. 3. Comparative amino acid analyses of the Chinese hamster (47), human (8), chicken (43, 58), and vaccinia virus (78) TK proteins. The amino acid sequence of the hamster TK protein is presented in line 1, encoded according to standard International Union of Pure and Applied Chemistry conventions and organized according to the exon-intron organization of the Chinese hamster TK gene. Amino acid sequence divergences of the human (line 2), chicken (line 3), and vaccinia virus (line 4) proteins from the Chinese hamster sequence are noted where appropriate. The highly conserved regions A and B described in the text are enclosed by brackets.

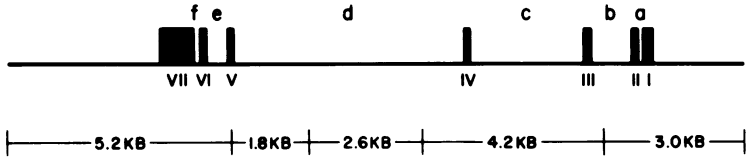
Mouse Ltk$^-$ cell transformation assay. The various TK minigene constructions described below and their exoIII-S1 deletion derivatives were transfected as $CaPO_4$ precipitates in the presence of mouse Ltk$^-$ carrier DNA (20 μg/ml) onto mouse Ltk$^-$ cells seeded 24 h earlier at a density of 10^6 cells per 100-mm dish, essentially as described by Wigler et al. (79). Tk$^+$ transformant colonies were fixed with methanol and identified by Giemsa staining after a 14-day period of hypoxanthine-aminopterin-thymidine selection. Plasmid DNAs used for Ltk$^+$ transfection analysis were purified by banding to equilibrium on ethidium bromide-CsCl gradients essentially as described (47). Plasmid DNA concentrations were read by UV spectroscopy and confirmed by estimations of relative ethidium bromide fluorescence after agarose gel electrophoresis of linearized DNA.

RESULTS AND DISCUSSION

Construction and identification of Chinese hamster TK cDNA clones. My colleagues and I have previously reported (47) (i) that the 3.2-, 4.2-, and 5.5-kb *Eco*RI fragments of the recombinant phage λ34S4 containing the Chinese hamster TK gene reacted in Southern blot hybridizations with a 3.0-kb *Hin*dIII fragment containing the cloned chicken TK gene, and, further, (ii) that each of these three *Eco*RI fragments, in Northern blot hybridizations, detected a 1,400-nucleotide RNA in cytoplasmic poly(A)$^+$ RNA from the Chinese hamster cell line A-29, but was unable to detect such an RNA in cytoplasmic poly(A)$^+$ RNA from a CHO Tk$^-$ cell line obtained from M. Harris (30). We concluded, on the basis of these data, that the Chinese hamster TK mRNA was a 1.4-kb poly(A)$^+$ species whose sequences were partly homologous to the chicken TK gene and were distributed through 11.2 kb of hamster genomic DNA.

To clone the hamster TK mRNA, my colleagues and I prepared double-stranded, *Eco*RI-methylated, and size-fractionated cDNA to cytoplasmic poly(A)$^+$ RNA from the Chinese hamster ovary cell line A-29 and cloned this cDNA, using *Eco*RI linkers, into the *Eco*RI-accepting phage vector λgt10 (33), generating a library of approximately 3×10^6 recombinant phage. Using a 900-bp *Eco*RI-*Pvu*II subclone of the 3.2-kb *Eco*RI fragment of λ34S4 (free of L47 phage sequences), we screened nearly 8×10^5 recombinants by the plaque hybridization method of Benton and Davis (4) and identified 20 candidate Chinese hamster TK cDNA recombinant clones. The DNAs of all 20 phage, prepared from 50-ml liquid lysate cultures, were digested with *Eco*RI and analyzed by Southern blotting using either the 900-bp *Eco*RI-*Pvu*II fragment or the 4.2- or the 5.2-kb *Eco*RI fragment of λ34S4 as a probe. We found the isolate designated λcTK.90 to carry the largest recombinant insert; its 1,250 bp were contained in two *Eco*RI fragments, one approximately 900 bp in size which hybridized with the *Eco*RI-*Pvu*II probe and the other 330 bp in size which hybridized with both the 4.2- and 5.2-kb *Eco*RI fragment probes. Since both *Eco*RI fragments of the λcTK.90 phage appeared TK specific by this Southern blot analysis, we concentrated our sequencing attention on the λcTK.90 isolate and variously subcloned this isolate into M13mp10 and M13mp11 vectors.

A.



B.

EXON I	CAGATCCAG.GTGCGGGGTCCGGGC.....TGACTTCTCTCCTAG.GTGATCCTC	EXON II
EXON II	AGGGAAAAG.GTAATGAATGGGCTT.....TGCTGACTCTTGCAG.CACCGAGCT	EXON III
EXON III	ACATGACCG.GTCAGTCCCACCACC.....CTTTCTTCTTCCCAG.GAACACCAT	EXON IV
EXON IV	GGACAGTTT.GTAAGTTGGCTTGAA.....CACCAACTTTCCCAG.TTCCCTGAC	EXON V
EXON V	CAGAGAAAG.GTAAGGTGTCTCACC.....CCCCTGTATATATAG.GCTTTCGGC	EXON VI
EXON VI	GAGAAGGAG.GTACCTTCTTGTGCT.....TGGCCCTGTCCACAG.GTGGAGGTG	EXON VII

FIG. 4. Exon-intron organization of the Chinese hamster TK gene. (A) The seven exons of the Chinese hamster TK gene, numbered I through VII, are drawn relatively to scale as blocked vertical projections from the line scaled according to the *Eco*RI fragments of the hamster TK gene recombinant clone λHaTK.5. Exons contain at least (exon I) 79, (exon II) 32, (exon III) 111, (exon IV) 94, (exon V) 90, (exon VI) 120, and (exon VII) 680 bp. The sizes of the six intervening sequences of the hamster TK gene, designated by lowercase letters, have been estimated by nucleotide sequencing or restriction enzyme mapping data to be (a) 90 bp, (b) 920 bp, (c) 3.1 kbp, (d) 5.4 kbp, (e) 650 bp, and (f) 100 bp. (B) Nucleotide sequences of the 12 exon-intron borders of the hamster TK gene. Exon sequences are separated from intron sequences by a single space and a dot. Nucleotides homologous to the 5′ donor consensus sequences AG/GTAAGT and the 3′ acceptor consensus sequences PyPyPyCAG (9) are indicated with a dot over the line. Nucleotides homologous to the corresponding chicken TK gene splice donor and acceptor sequences are underlined.

λcTK.90 sequence analysis. The nucleotide sequence of the 1,219-bp cDNA insert of λcTK.90 presented in Fig. 2 was established using the dideoxynucleotide chain-terminating method of Sanger et al. (67). The cDNA insert of λcTK.90 was sequenced on both strands with the exception of a 150-bp segment of 3′ noncoding DNA bordered by *Hin*dIII and *Eco*RI sites. All nucleotide sequence data were computer filed and analyzed using programs in general use at Cold Spring Harbor Laboratory (36).

The λcTK.90 cDNA contains six ATG-initiated open reading frames (ORFs) exceeding 90 bp, by far the largest of which is a 702-bp ORF which specifies a 25,625dalton protein which we identify as the hamster TK protein based on its close amino acid sequence homology to the human (8), chicken (43, 58), and vaccinia virus (78) TK proteins previously described. The ATG codon which initiates the TK ORF is embedded in the sequence CCGCCATGA, which is homologous at every nucleotide with the consensus sequence for translation initiator ATGs of higher eucaryotic mRNAs defined by Kozak (41). The TK ORF of λcTK.90 is followed by a 3′ noncoding region of 491 nucleotides, which contains, centered at nucleotide 1184, a 26-bp stretch of extreme A/T richness composed of five interdigitated sequences each of which approximates the consensus polyadenylation signal AATAAA identified by Fitzgerald and Shenk (19). The final such sequence, ATTAAA, is followed closely by (i) the sequence TTTCTACTTG, which resembles the consensus sequence TTTTCACTG found by Benoist et al. (3) to lie 3′ to the polyadenylation signal in a variety of higher eucaryotic mRNAs, and finally by (ii) an oligo(dA) tract of some 30 nucleotides. We are persuaded by the presence of this oligo(dA) tract, situated an appropriate distance from a cluster of putative polyadenylation signals, that the λcTK.90 cDNA contains the entire 3′ noncoding region of the hamster TK mRNA, a region some 171 nucleotides smaller than the 3′ noncoding region of human TK mRNA described by Bradshaw and Deininger (8).

The hamster TK mRNA sequence is 56% G/C rich (58% G/C coding; 53% G/C noncoding), identical in this respect to the base composition of the Chinese hamster metallothionein I and II mRNAs (26). It is, however, considerably more G/C rich than the Chinese hamster dihydrofolate reductase (DHFR) mRNA (43% G/C) (57) (44% coding; 43% noncoding) and the Chinese hamster hypoxanthine phosphoribosyltransferase (HPRT) mRNA (42% G/C) (40) (42% G/C coding; 42% G/C noncoding), the only other Chinese hamster mRNA sequences available to date (Gen Bank Release, vol. 29, 22 February 1985). Clearly the base composition of Chinese hamster mRNAs can vary widely, even among low-abundance mRNAs encoding enzymes important to nucleotide metabolism whose expression is cell cycle and growth phase dependent (i.e., TK, DHFR, and HPRT). Despite its G/C richness, the hamster TK mRNA sequence is relatively poor, like many other eucaryotic mRNAs, in occurrences of the dinucleotide CpG (6); this doublet occurs at only 35% of the expected random frequency for a 56% G/C-rich sequence.

The G/C richness of the hamster TK mRNA is in large measure a consequence of a nonrandom codon usage. In the TK reading frame of 702 nucleotides, 75% of 234 codons contain either C or G in the third base position. Of the 123 codons with a pyrimidine in the third base position, 94 use C in that position. C is preferred to U as a third-base-position pyrimidine by the codon family of every amino acid. Of the 111 codons with a purine in the third base position, 80 use G in that position. This apparent overall preference for G as a third-base-position purine is largely a reflection, however, of a striking preferential use of GUG over GUA (16:1) as a

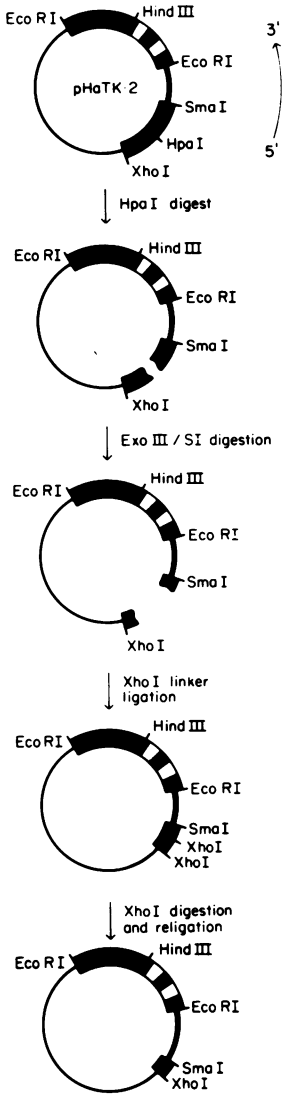


FIG. 5. Strategy for the construction of 5′ deletion mutants of the Chinese hamster TK minigene pHaTK.2. pHaTK.2 DNA was linearized by *Hpa*I digestion, resected with exoIII and S1 nucleases, and recircularized by ligation to synthetic 8-mer *Xho*I linkers (CCTCGAGG). Ligation products were digested with *Xho*I, religated, and cloned by transformation into *E. coli* DH.1 to yield pHaTK.2 deletion mutants, whose deletion endpoints were juxtaposed to the same pBA vector sequences. The size of various deletions was first estimated by *Eco*RI-*Xho*I digestion and agarose gel electrophoresis. The endpoints of appropriately sized deletions were defined by dideoxynucleotide sequencing of M13mp10 clones containing *Xho*I-*Eco*RI fragments of the various deletion mutants.

valine codon, of CUG over CUA (12:1) as a leucine codon, and of CAG over CAA (11:1) as a glutamine codon. It is interesting, in the context of mRNA translation strategies, that the nucleotide sequence motifs which flank the translation initiator AUG of the Chinese hamster TK, DHFR, and HPRT mRNAs contain G as the −3 position purine whereas 78% of the eucaryotic mRNAs surveyed by Kozak (41) contained A. Since G as a −3 purine has been shown to diminish by threefold the efficiency with which the preproinsulin mRNA is translated in vivo (relative to A as a −3 purine) (42), the use of G as a −3 purine by these three mRNAs may constitute an element in the regulation of the expression of these genes whose products are normally expressed at low level in a cell cycle-modulated fashion.

The Chinese hamster TK protein. The 702-bp ORF of the hamster TK cDNA λcTK.90 specifies a protein with a molecular weight of 25,625 whose 234 amino acids are 90% conserved with the human TK protein (8) (25,504 daltons, 234 amino acids), a value intermediate in the range of 88 to 94% homology reported for other Chinese hamster/human amino acid sequence comparisons (35, 56). The hamster TK protein is 80% conserved with the chicken TK protein (24,844 daltons, 225 amino acids) (43, 58) and 60% conserved with the vaccinia virus TK protein (20,102 daltons, 177 amino acids) (78) (Fig. 3 and 4). These interspecies protein sequence conservations are realized through nucleotide sequence conservations of 82, 75, and 60% between the coding sequences of hamster and human, chicken, and vaccinia virus genes, respectively. The hamster TK gene shows no extended homology at either the nucleotide or amino acid sequence level to the TK gene of herpes simplex virus type 1 (54).

The hamster, human, chicken, and vaccinia virus TK proteins are least conserved at the far carboxy terminus. The chicken TK and vaccinia virus TK proteins, for example, are abbreviated at the carboxy terminus by 10 and 42 amino acids, respectively, with respect to the hamster and human sequences, and the hamster-human and hamster-chicken amino acid sequence homologies are reduced to 75 and 50%, respectively, through amino acids 184 to 234 of the hamster TK sequence. In contrast to this pronounced interspecies C-terminal sequence divergence, the four TK proteins contain two regions of 22 and 39 amino acids, which we designate regions A and B (amino acid residues 21 through 42 and 93 through 131, respectively, of the Chinese hamster TK protein, shown bracketed in Fig. 4) and which show interspecies conservation of 95 and 96%, respectively (total conserved residues/total residues). My co-workers and I assume that regions A and B are indispensably important to TK catalysis as domains which perhaps constitute substrate binding sites or participate directly in the thymidine phosphorylation reaction. We analyzed the hydrophobic character of regions A and B, relying on the hydropathy profile established for the chicken TK gene by Kwoh and Engler (43), and found that both regions A and B are composed of subregions of significant hydrophobic character (region A, residues 21 through 30; region B, residues 98 through 110 and 118 through 140) flanked by amino acid stretches which are neither remarkably hydrophobic nor hydrophilic. It is tempting to speculate that these hydrophobic subregions of regions A and B are, in fact, the crucially important elements for TK catalysis. Neither region A nor region B is encoded completely by a single TK gene exon (region A is encoded by exons I, II, nd III, region B by exons IV and V), although the major hydrophobic subregion of region A appears to be encoded uniquely by exon II, an exon-intron arrangement, consistent with the suggestions of Gilbert (23) and others (24, 51, 71, 74) that functional domains of proteins are organized genetically as discrete exon elements. The hydrophobic peaks of region B do not, however, fit so discrete an exon-specific pattern.

My colleagues and I have used the Seq HP program (25) to search the Dayhoff data base for oligopeptides homologous to regions A and B which may be contained as functional elements in other procaryotic or eucaryotic proteins. Using a default score of −40 as a homology stringency criterion,

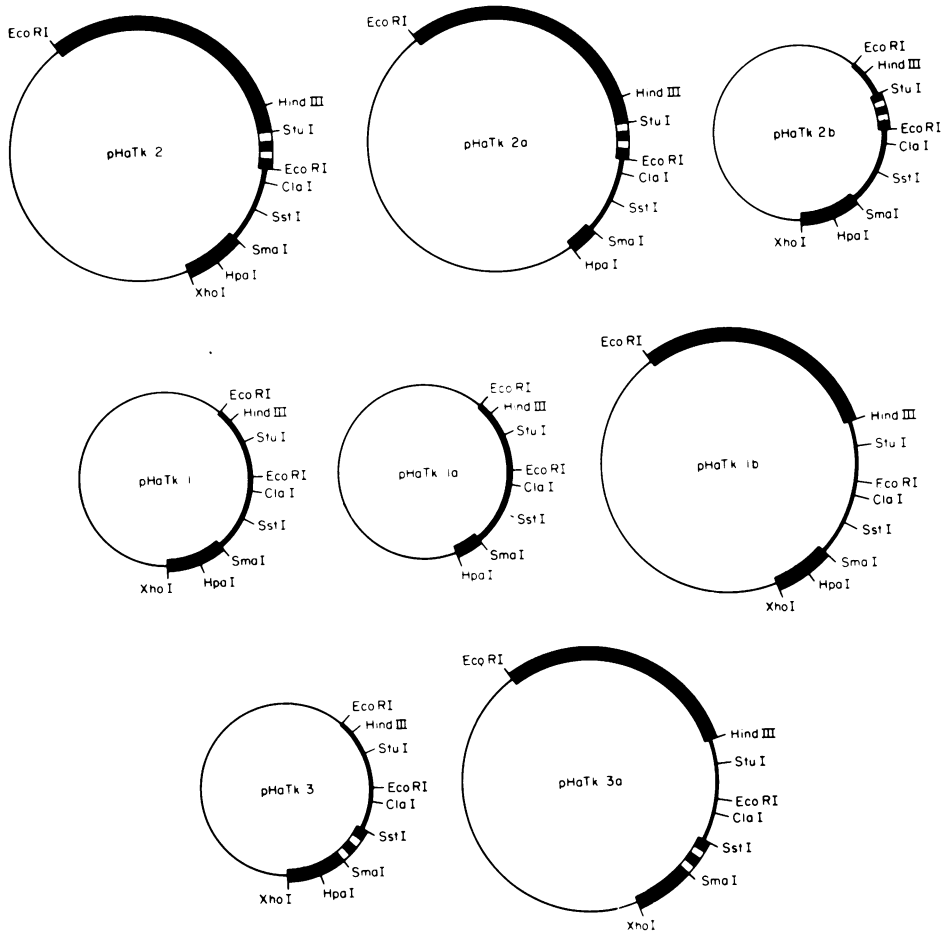


FIG. 6. Partial restriction maps of various Chinese hamster TK minigenes. The construction of these genes in the vector pBA is described fully in Materials and Methods. Medium-thick lines represent sequences derived from the Chinese hamster TK cDNA λcTK.90. Heavy lines represent sequences derived from the hamster TK gene clone λHaTK.5. (Light-thick represent sequences from the vector pBA.) Intervening sequences are indicated as open spaces within heavy line segments.

we have extracted some 50 oligopeptide sequences ranging in size from 13 to 27 amino acids, with homologies to region B which range from 25 to 37%. None of these oligopeptides, however, are significantly homologous to region B by the criteria of the Seq DP program (using up to 100 randomly generated sequences.) It remains a formal possibility that a core peptide of either region A or B (perhaps the hydrophobic subregions we have identified) is highly conserved among ATP or nucleoside binding proteins, though we are unable to identify such a peptide by using the computer parameters we have selected for reasons of practicality. Neither region A nor region B contains the conserved amino acid sequence motif Gly-X-Gly-X-X-Gly-$(X)_n$-Lys, described as a core element of various ATP binding proteins (14). Neither region A nor region B shows significant sustained amino acid sequence homology with the various ATP binding proteins surveyed by Walker et al. (77), though region B does contain the tripeptide glycine-lysine-threonine (residues 116, 117, and 118) which appears as a conserved tripeptide in six of the seven proteins in the Walker series.

Chinese hamster TK gene structure. My colleagues and I have defined the exon-intron organization of the hamster TK gene by comparing the nucleotide sequence of the λcTK.90 cDNA with nucleotide sequence data derived from various M13 subclones of the recombinant phage λHaTK.5 which hybridize in Southern blot analysis with the λcTK.90 cDNA. The λHaTK.5 isolate contains the Chinese hamster TK gene within a 17-kb *Bcl*I insert fragment cloned directly from DNA of the Chinese hamster ovary cell line A-29. The λHaTK.5 clone contains 3.5 kb of genomic DNA at the 3′ end of the hamster TK gene that is not contained in the λ34S4 TK gene isolate that we described previously and which was cloned from genomic DNA of an Ltk^- cell line transformed to Tk^+ with hamster A-29 DNA (45). The recombinant inserts of λHaTK.5 and the λ34S4 are otherwise structurally identical.

The exon-intron structure of the Chinese hamster TK gene is presented in Fig. 5. The sequences of the λcTK.90 cDNA are distributed through 11.2 kb of Chinese hamster genomic DNA in at least seven exons ranging in size from 32 bp (exon II) to 680 bp (exon VII). The seven exons are segregated by introns which vary over 50-fold in size, from the 90-bp intron *a* which segregates exons I and II to the 5.4-kb intron *d* segregating exons IV and V. Although the 5′ and 3′ coding nucleotides joined directly by splicing can be defined unambiguously for only one of the six intervening sequences, it is nonetheless possible to consider all six introns as sequences initiated by the dinucleotide GT and terminated by the

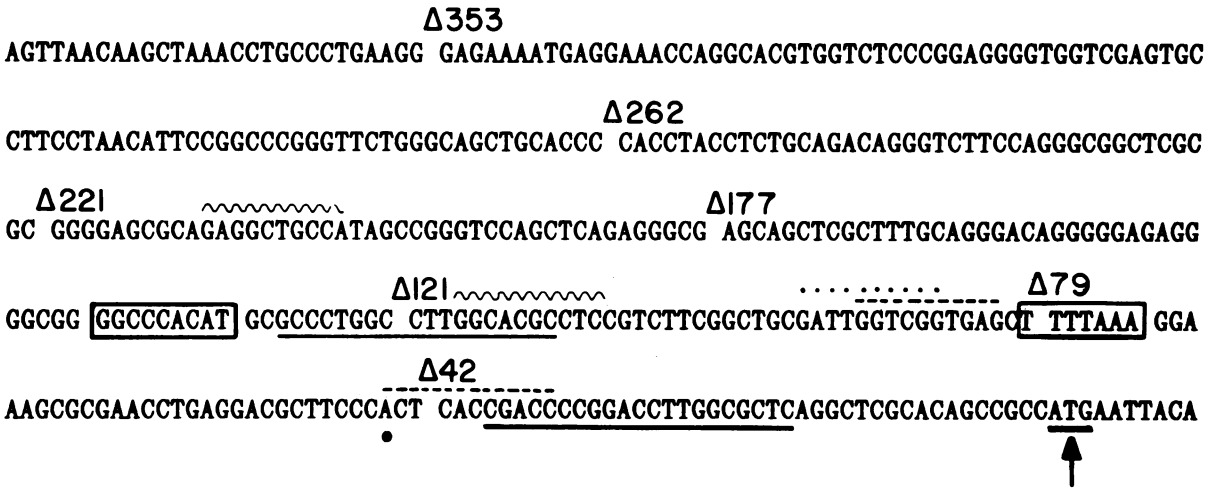


FIG. 7. Nucleotide sequence of 380 bp of 5′ flanking DNA. The deletion endpoints of the exoIII-S1 derivatives of pHaTK.2 are indicated above the sequence, numbered to reflect the nucleotides remaining 5′ to the translation initiator AUG. The translation initiator ATG is indicated with an arrow. The T/A-rich sequence and the CAT box homologous sequences are enclosed in boxes. The 19-bp imperfect direct duplications are underscored; the inverted repeat sequences related to this duplication are overscored as wavy or dashed lines. The 10-bp sequence which appears in the 5′ flanking region of the chicken TK gene as a sequence triplication is indicated by superscript dots.

dinucleotide AG, in accord with the GT/AC rule summarized by Breathnach and Chambon (9). The six 3′ splice acceptor sequences (considered as the 15 most 3′ nucleotides of the intervening sequence) conform in all cases with the PyAG consensus expectations (61), are pyrimidine rich (ranging from 9/15 pyrimidines to 13/15 pyrimidines), and contain a total of only four dipurines, none of which is AG. The 5′ and 3′ splice donor and acceptor sequences of the hamster TK gene are unremarkable at this level of nucleotide sequence analysis (9, 69).

The exon-intron organization of the hamster TK gene, as I have described it here, is similar to that of the chicken TK gene as detailed by Kwoh and Engler (43) and Merrill et al. (58). The coding sequences of both genes are interrupted by intervening sequences which separate coding nucleotides 66/67, 98/99, 209/210, 303/304, 393/394, and 513/514 (if the appropriate corrections are made for the 3-bp insertion in the chicken coding sequence relative to the hamster sequence at position 198). Since cDNA clones encompassing the complete 3′ noncoding region of the chicken TK gene have not yet been described, it remains a formal possibility that additional intervening sequences may interrupt this region, which has been estimated by S1 mapping and agarose gel electrophoresis to be as large as 1,400 nucleotides, nearly 1,000 nucleotides larger than the 3′ noncoding region of the hamster TK gene (63). Despite these overall organizational similarities, the hamster and chicken TK genes differ significantly in size. The introns of the hamster TK gene, which range from 90 bases to 5.4 kb in size, total nearly 10 kb, whereas by contrast the introns of the chicken TK gene, not one of which exceeds 230 bases, total only 1 kb (43, 58).

The Chinese hamster TK gene promoter. The 5′ extreme of the hamster TK gene has not yet been characterized in complete detail. My colleagues and I have, however, defined a 5′ boundary for the Chinese hamster TK gene promoter by constructing the TK minigene pHaTK.2 (Fig. 6) and a series of exoIII-S1 deletion mutants of pHaTK.2 which were tested for their Tk^+ transforming activity in mouse Ltk^- cells (Table 1). The parental Chinese hamster TK minigene pHaTK.2 is composed of (i) a 1.6-kb *Xho*I-partial *Sma*I fragment from λHaTK.5 which contains approximately 1.5 kb of genomic DNA 5′ to the translation initiator AUG codon plus 52 bp of exon I; (ii) a 250-bp *Sma*I-*Eco*RI insert fragment of λcTK.90 which contains the TK coding sequences of exon I from the *Sma*I site to the *Eco*RI site in exon V; and (iii) a 5.2-kb *Eco*RI fragment derived from λHaTK.5 which contains the remainder of exon V, exons VI and VII, and 3.5 kb of 3′-flanking DNA. This pHaTK.2 minigene and the pHaTK.2a derivative of pHaTK.2, which carries a 1-kb deletion between the *Hpa*I and *Xho*I sites (Fig. 6), transform Ltk^- cells to Tk^+ with an efficiency of three to five colonies per ng. We therefore concluded that the Chinese hamster TK gene promoter is contained 3′ to the *Hpa*I site, within approximately 400 bp 5′ to the translation initiator AUG. To define the 5′ boundary of the TK promoter with greater precision, we prepared a nested set of exoIII-S1 deletion mutants of pHaTK.2 as described in the legend to Fig. 7, according to a strategy that left the deletion endpoint of each mutant juxtaposed to the same plasmid sequence. Seven mutant derivatives of pHaTK.2, with deletions 3′ to

TABLE 1. Ltk^+ transformation efficiency of various Chinese hamster TK minigenes[a]

Minigene	Total colonies	Relative efficiency
pHaTK.1	33	1.0
pHaTK.1a	28	0.8
pHaTK.1b	106	3.2
pHaTK.2	775	23.4
pHaTK.2a	706	21.3
pHaTK.2b	421	12.7
pHaTK.3	257	7.7
pHaTK.3a	465	14.0

[a] TK minigenes (10 ng) were transfected as $CaPO_4$ precipitates in the presence of 20 μg of carrier DNA onto mouse Ltk^- cells as described by Wigler et al.(79). Tk^+ colonies were scored after 17 days of hypoxanthine-aminopterin-thymidine selection. Data reported here are averages of two transfection experiments. Total colonies represent the data from two experiments. The transformation efficiency of pHaTK.1 is set as relative efficiency of 1.0.

the *Hpa*I site ranging from 30 to 350 bp, were transfected onto mouse Ltk$^-$ cells and scored for Tk$^+$ transformation efficiency. The results of these assays (data not presented) show clearly that pHaTK.2 derivatives which retain at least 121 bp 5′ to the translation initiator AUG are just as efficient at Tk$^+$ transformation as the parental minigene pHaTK.2. The deletion mutant pHaTK.2-79, however, which retains only 79 bp 5′ to the initiator AUG, transform Ltk$^-$ cells with a nearly 100-fold reduction in efficiency. This reduction is unlikely to be an effect of proximal pBA vector sequences influencing TK gene expression since a derivative of pHaTK.2-79, from which 400 bp of pBA DNA 5′ to the *Xho*I site has been deleted, is just as inefficient as pHaTK.2-79 in the TK$^+$ transformation assay. Furthermore, the reduction is clearly not the result of a coding sense mutation occurring during exoIII-S1 mutagenesis since pHaTK.2-79 transforms with the efficiency of pHaTK.2 when combined with a 900-bp *Bam*HI-*Xho*I fragment containing the herpes simplex virus type 1 TK gene promoter (43). We have concluded that the transcription of the hamster TK gene at levels sufficient to transform Ltk$^-$ cells to Tk$^+$ depends on a genetic element(s) contained within 121 bp 5′ to the translation initiator AUG.

Nucleotide sequence organization of the Chinese hamster TK gene promoter. The nucleotide sequence of the 400 bp 5′ to the translation initiator AUG is presented in Fig. 7. Within the region my colleagues and I define as essential to TK gene transcription is found heptanucleotide TTTTAAA at −79/−72, which resembles the TATAA box element described by Breathnach and Chambon (9). This hamster T/A-rich sequence lies 60 bp 3′ to the sequence GGC CCACAT at −139/−131, which is striking in its homology to the canonical CAT box sequence which Corden et al. (12) have shown to lie 50 to 60 bp 5′ to the conventional TATAA element. (The hamster CAT box homolog, it should be noted, lies 5′ to the boundary we have proposed for a minimally functional TK gene promoter.) If the hamster T/A-rich sequence is in fact essential to the positioning of TK gene transcription initiation, I suggest that the transcription of the hamster TK gene is initiated at the A residue, at −44 which lies (i) in the pyrimidine-rich sequence CTTCCC CACTC and (ii) within 33 bp of the 5′-most T of the T/A-rich sequence, and thereby fulfills consensus sequence expectations for transcription initiation sites defined by Corden et al. (12) and Breathnach and Chambon (9).

The hamster T/A-rich sequence may of course be functionally irrelevant to TK gene transcription since it now appears that TATA box elements are dispensable in the canonical form and location for the transcription of various eucaryotic genes, notably the chicken TK gene (43, 58), the mouse DHFR (13) and HPRT (57) genes, the hamster hydroxymethylguanine coenzyme A reductase gene (65), and the human adenosine deaminase gene (76), all represented in the cytoplasm by low-abundance mRNAs. At present, the most conservative interpretation of hamster TK gene transcription is that it might be initiated at any nucleotide 3′ to −121. This region contains the sequence GGTCG GT at −90/−84 which might be considered a candidate donor sequence that splices to the CCTTGGCGCTCAG sequence at −28/−16. My colleagues and I have seriously considered such a splicing possibility since, as we would expect by this argument, the final 9 bp of the λcTK.90 cDNA diverges from the genomic sequence at the AG dinucleotide of the −28/−16 sequence. We have sequenced multiple M13 clones of both the λcTK.90 cDNA and λHaTK.5 isolate through this region and consistently find this discrepancy. The final nine nucleotides of the λcTK.90 cDNA which diverge from the genomic sequence do not, however, lie 5′ to the putative splice donor sequence at −90/−84, a result we would expect if the splicing alternative we propose is correct. It remains a formal possibility, of course, that the transcription of the hamster TK gene is initiated at various promoters, under various conditions of cell growth, and that the λcTK.90 cDNA was prepared to a hamster TK mRNA whose transcription was initiated at a promoter 5′ to the promoter we have characterized in the pHaTK.2 minigene.

My colleagues and I have considered several other explanations for this cDNA/genomic sequence divergence, namely, (i) that the TK genes in the Chinese hamster A-29 cell line are polymorphic, the λcTK.90 cDNA being made to a TK mRNA transcribed from an allele of the TK gene cloned in λHaTK.5, and (ii) that the divergence is a simple matter of artifactual reverse transcription at the 5′ extreme of the hamster TK mRNA. To consider the alternative of polymorphism, we are currently sequencing the appropriate subclones of other TK gene isolates from the A-29 cell line. It is important to point out in the context of allelism that this 9-bp divergence is the only such example we have encountered in comparing the 1,219 bp of λcTK.90 with the λHaTK.5 isolate.

The 5′ region flanking the translation initiator AUG contains the perfect 10-bp inverted repeat sequences at −43/−34 and −90/−81 and the nearly perfect 10-bp inverted repeat sequences at −118/−109 and −210/−201. It is tempting to speculate that these sequences contribute to the modulation of hamster TK gene transcription through cruciform transitions which would isolate the T/A-rich and CAT homologous hamster sequences at −79/−72 and −121/−110 in single-stranded loops of 30 and 80 bp, respectively. We are ignorant of the organization of this 5′ flanking region into higher order hamster chromatin, however, and therefore are unable to argue the favorability of such transitions when they might be influenced by DNA binding proteins or RNA polymerase II-associated transcription factors. The 400 bp immediately 5′ to the translation initiator AUG of the chicken TK gene contains 12 inverted repeat sequence pairs of 7 bp or larger, though none of these invert repeats is obviously homologous, on the basis of nucleotide sequence, spacing, or position relative to the initiator AUG, to the two inverted repeats of the hamster 5′ flanking region.

It is important, in considering a functional significance for these hamster inverted repeat sequences, to place them in the context of the direct sequence repetitions which also occur in the 5′ region flanking the hamster TK gene promoter. This region contains, at −128/−110, the 19-bp sequence GCCCTGGCCTTGGCACGCC which is clearly a degenerate repeat of the sequence ACCCCGGACCTTG GCGCTC at −36/−18. These sequences are physically related to the hamster inverted repeat sequences since the final six nucleotides of the direct repeat at −128/−110 are the first seven nucleotides of the inverted repeat element at −118/−108 and the first five nucleotides of the −36/−18 inverted repeat element are the final five nucleotides of the inverted repeat element at −43/−34. A similar situation prevails in the 5′ flanking region of the chicken TK gene, which contains three direct repetitions of the consensus sequence PyPyGPyPyGGATTGGTCG, elements of which occur in two other locations within the 5′ flanking region as elements of inverted repeat sequences (43, 58). The core sequence of this triplicated chicken sequence, GATTG GTCG, occurs once in an undegenerate way in the 5′ flanking region of the hamster TK gene at −94/−85, and

elements of this sequence occur as part of the indirect repeat at −90/−81 in the hamster 5′ flanking DNA. It seems clear that the immediate 5′ flanking regions of the Chinese hamster and chicken TK genes have been modeled to a considerable extent by direct and indirect repetitions of a core nucleotide sequence. It is uncertain whether this modeling has a functional significance in TK gene transcriptional control or whether it reflects merely the inherent instability or susceptibility to rearrangement of certain DNA sequences which lie outside protein coding sequences.

It is important to point out that direct repeat sequences of 10 bp and longer have been reported to occur as well in 5′ flanking DNA of other higher eucaryotic genes (13, 57) whose low-level expression is growth phase dependent. My co-workers and I are intrigued to find that the 10-bp sequence CGGACCTTGG at −32/−23 in the 5′ flanking region of the hamster TK gene, and its degenerate homolog TG GATTGG at −125/−117, are closely related to the 10-bp, directly repeated sequence CGGAGCCTGG reported to lie within 100 bp of the translation initiator AUG of the mouse HPRT gene (57) and are related as well to the sequence CGGGGCCTTGG, which occurs as a core element of each of the four 40-bp direct repeats within the 5′ promoter-proximal DNA of the mouse DHFR gene (13, 18). We would argue that the direct repetition of a conserved 10-bp sequence within the 5′ flanking DNA of at least three genes whose low-level expression is growth phase dependent suggests a functionally important role for this sequence.

The 380 bp of DNA which flanks the translation initiator AUG of the hamster TK gene is 65% G/C rich and contains 27 occurrences of the dinucleotide CpG, nearly 75% the expected frequency for a sequence of this composition. By contrast, the TK mRNA sequence is 56% G/C rich and contains only 35% the expected CpG frequency. Similar patterns of relative CpG enrichment in 5′ flanking DNA have been reported as characteristics of other eucaryotic genes by McClelland and Ivarie (53), a finding that attracts attention since methylations at cytosines which occur in the sequence CpG have been proposed to mediate transcriptional repression in a number of genetic systems (15). It is conceivable, by analogy, that hamster TK gene expression in certain developmental contexts (e.g., the decline of TK activity during the terminal differentiation of various cell lineages) is accomplished by cytosine methylation within the CpG-enriched 5′ flanking DNA. Such a model is inconsistent, however, with recent data described by Merrill et al. (58) which argue for a posttranscriptional mechanism regulating the extinction of TK activity in mouse myoblast cultures induced to terminal differentiation. The significance of the CpG enrichment of the hamster TK gene therefore remains to be explained. The 5′ flanking region contains two occurrences of the nucleotide sequence GGGCGG (−233/−228, −145/−140), which appears altogether six times within the triplicated 21-bp region of the simian virus 40 early promoter, where it functions by binding the SP1 transcription factor (22). The significance of these GGGCGG sequences to the transcription of the Chinese hamster TK gene is unclear.

The DNA strands of this 5′ flanking region are marked by four stretches of extreme base composition asymmetry which occur (i) between −357 and −333, where 22 of 25 coding strand nucleotides are purines; (ii) between −320 and −306, where 12 of 15 nucleotides are purines; (iii) between −220 and −207, where 12 of 14 nucleotides are purines; and (iv) most remarkably between −161 and −138, where 22 of 24 nucleotides are purines. Although it is clear that strand asymmetries of this nature contribute to significant local variations in DNA helix stability through resonance stacking effects (75), my colleagues and I are uncertain whether the frequency and magnitude of the asymmetries described here are sufficient to modulate the rate of transcription complex formation at the hamster TK gene promoter. In striking contrast to these examples of strand asymmetry, this 5′ flanking region includes only a single nucleotide sequence as long as 9 bp of perfect purine-pyrimidine alternation, a sequence character shown by Nordheim and Rich (62) to be capable of Z DNA transitions under the proper conditions of ionic strength and superhelical density. This 9-bp sequence involves the final five nucleotides of the sequence GGC-CCACAT (shown boxed in Fig. 5), which as noted earlier conforms to the consensus sequence derived for upstream promoter elements.

Variables affecting TK minigene expression. My colleagues and I have constructed, in addition to pHaTK.2, four other hamster TK minigenes (Fig. 6). These genes were developed to analyze the structural determinants of the unexpected finding that the minigene pHaTK.1 (which we constructed to assess the importance of primary RNA structure to growth phase-dependent TK gene expression) transformed mouse Ltk$^-$ cells nearly 25-fold less efficiently than did the minigene pHaTK.2. Since pHaTK.1 is composed of the coding and 3′ noncoding sequences of the TK mRNA fused to the Chinese hamster TK gene promoter, it differs substantially in structure from pHaTK.2, which contains the cDNA sequences of pHaTK.1 as well as introns *e* and *f* and 3.5 kb of genomic DNA derived from the 3′ end of the hamster TK gene (isolate λHaTK.5). To consider the contribution of intervening sequences *e* and *f* alone to the transformation efficiency of pHaTK.2, we introduced these sequences into pHaTK.1 to create pHaTK.2b (for construction details see Materials and Methods) and found that this pHaTK.2b gene transformed Ltk$^-$ cells more than 10-fold more efficiently than pHaTK.1 (see Fig. 6). This transformation enhancement effect of introns *e* and *f* is not, however, sequence specific. We introduced introns *a* and *b* into pHaTK.1 to create pHaTK.3 and found that this gene also transformed Ltk$^-$ cells almost eightfold more efficiently than pHaTK.1. We do not yet understand how the inclusion of intervening sequences in our minigene constructions influences Tk$^+$ transformation efficiency, though we can imagine that these sequences either (i) facilitate the integration of the transfected DNAs into host genomic DNA, (ii) enhance TK gene transcription, or (iii) increase the efficiency with which primary transcripts of the hamster TK gene mature to cytoplasmic mRNA. It is interesting that Gasser et al. (21) have previously reported that the inclusion of DHFR intervening sequences in DHFR minigenes improved DHFR$^+$ transformation efficiencies as much as 30-fold. Both of these observations are consistent with earlier reports of Gruss et al. (27) and Hamer and Leder (28), which described the importance of intervening sequences to various aspects of simian virus 40 gene expression.

Both the pHaTK.2b and pHaTK.3 minigenes, however, are measurably less efficient at Tk$^+$ transformation than pHaTK.2, a difference my colleagues and I considered might relate to the 3.5 kb of genomic DNA from the 3′ end of the hamster TK gene contained in pHaTK.2. To test the effects of this DNA segment on the transformation efficiency of other TK gene constructions, we introduced this DNA segment into pHaTK.1 to create the minigene pHaTK.1c and into pHaTK.3 to create pHaTK.3a. These minigenes proved to be approximately twofold more efficient at Tk$^+$ transformation than their respective parental TK minigenes.

We have not yet characterized those elements within the 3.5 kb of 3′ TK genomic DNA which contribute to the Tk^+ transformation efficiencies of the pHaTK.1b, pHaTK.2, and pHaTK.3a minigenes, though we suspect they are related to the process of RNA polyadenylation. The pHaTK.1 minigene, though it contains multiple putative polyadenylation signals, lacks genomic sequences 3′ to the site of polyadenylation which may contribute to the efficiency of site recognition, RNA cleavage, or RNA transcript polyadenylation. We sequenced 100 bp of hamster genomic DNA beyond the site of poly(A) addition and failed to find the sequence YGTGTTYT described by McLauchlan et al. (55) as a highly conserved sequence lying 24 to 38 bp 3′ to the polyadenylation site in numerous eucaryotic mRNAs. This nucleotide 24 to 39 region in the hamster genome is distinctly pyrimidine rich, however, a characteristic of those mRNAs described by McLaughlan et al. (55) which lack this motif. It is conceivable that this pyrimidine richness in the primary RNA transcript of the hamster TK RNA contributes in part to a recognition sequence to which enzymes involved in polyadenylation are attracted. The incorporation of these sequences into TK minigenes may therefore improve the efficiency of minigene Tk^+ transformation by affecting steady-state levels of cytoplasmic TK mRNA.

ACKNOWLEDGMENTS

This research was supported by a National Science Foundation grant to J.A.L. (PCM 8309360).

I thank Diana Matkovich for superb technical assistance, Mike Wigler and David Kurtz for support and helpful suggestions, Greg Freyer for advice with cDNA synthesis, Steve Hinton for advice with exoIII-S1 mutagenesis, Mark Zoller for oligonucleotide synthesis, Chris Keller and Keith Deen for help with computer analyses, Harvey Bradshaw and Prescott Deininger for sharing sequence information before publication, and Lynne Bonde for her patient and always careful manuscript preparation.

LITERATURE CITED

1. **Adams, S. P., K. S. Kavka, E. J. Wykes, S. B. Holder, and G. R. Gallupi.** 1983. Hindered diaklylamino nucleoside phosphite reagents in the synthesis of two DNA 51-mers. J. Am. Chem. Soc. **105**:661–663.
2. **Bello, L. J.** 1974. Regulation of thymidine kinase synthesis in human cells. Exp. Cell Res. **89**:263–274.
3. **Benoist, C., K. O'Hare, R. Breathnach, and P. Chambon.** 1980. The ovalbumin gene sequence of putative control regions. Nucleic Acids Res. **8**:127–142.
4. **Benton, W. D., and R. W. Davis.** 1977. Screening λgt recombinant clones by hybridization to single plaques *in situ*. Science **196**:180–182.
5. **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.
6. **Bird, A. P.** 1980. DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res. **8**:1499–1504.
7. **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.
8. **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.
9. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
10. **Brent, T. P., J. A. V. Butler, and A. R. Crathorn.** 1965. Variations in phosphokinase activities during the cell cycle in synchronous populations of HeLa cells. Nature (London) **207**: 176–177.
11. **Clewell, D. B., and D. R. Helinski.** 1972. Effect of growth conditions on the formation of the relaxation complex of supercoiled ColE1 deoxyribonucleic acid and protein in *Escherichia coli*. J. Bacteriol. **110**:1135–1146.
12. **Corden, J., B. Wasylyk, A. Buchwalder, P. Sassone-Corsi, C. Kedinger, and P. Chambon.** 1980. Expression of cloned genes in new environment: promoter sequences of eucaryotic protein-coding genes. Science **209**:1406–1414.
13. **Crouse, G. F., C. C. Simonsen, R. N. McEwan, and R. T. Schimke.** 1982. Structure of amplified normal and variant dihydrofolate reductase genes in mouse sarcoma S180 cells. J. Biol. Chem. **257**:7887–7897.
14. **Debouck, C., A. Riccio, D. Schumperli, K. McKenney, J. Jeffers, C. Hughes, and M. Rosenberg.** 1985. Structure of the galactokinase gene of *Escherichia coli*, the last gene of the gal operon. Nucleic Acids Res. **13**:1841–1853.
15. **Doerfler, W.** 1983. DNA methylation and gene activity. Annu. Rev. Biochem. **52**:93–124.
16. **Eker, P.** 1965. Activities of thymidine kinase and thymine deoxyribonucleotide phosphatase during growth of cells in tissue culture. J. Biol. Chem. **240**:2607–2611.
17. **Enquist, L., and N. Sternberg.** 1979. *In vitro* packaging of λ and cosmid DNA. Methods Enzymol. **68**:299–309.
18. **Farnham, P. J., and R. T. Schimke.** 1985. Transcriptional regulation of mouse dihydrofolate reductase in the cell cycle. J. Biol. Chem. **260**:7675–7680.
19. **Fitzgerald, M., and T. Shenk.** 1981. The sequence 5′ AAUAAA 3′ forms part of the recognition sequence for polyadenylation in late SV40 mRNAs. Cell **24**:251–260.
20. **Flintoff, W. F., S. V. Davidson, and L. Siminovitch.** 1976. Isolation and partial characterization of three methotrexate-resistant phenotypes from Chinese hamster ovary cells. Somatic Cell Genet. **2**:246–261.
21. **Gasser, S., C. C. Simonsen, J. W. Schilling, and R. T. Schimke.** 1982. Expression of abbreviated mouse dihydrofolate reductase genes in cultured hamster cells. Proc. Natl. Acad. Sci. USA **79**:6522–6526.
22. **Gidoni, D., W. S. Dynan, and R. Tjian.** 1984. Multiple specific contacts between a mammalian transcription factor and its cognate promoters. Nature (London) **312**:409–413.
23. **Gilbert, W.** 1978. Why genes in pieces? Nature (London) **271**:501.
24. **Go, M.** 1983. Modular structural units, exons, and function in chicken lysozyme. Proc. Natl. Acad. Sci. USA **80**:1964–1968.
25. **Goad, W., and M. I. Kanehisa.** 1982. Pattern recognition in nucleic acid sequences. I. A general method for finding local homologies and symmetries. Nucleic Acids Res. **10**:247–263.
26. **Griffith, B. B., R. A. Walters, M. D. Enger, C. E. Hildebrand, and J. K. Griffith.** 1983. cDNA cloning and nucleotide sequence comparison of Chinese hamster metallothionein I and II mRNAs. Nucleic Acids Res. **3**:901–910.
27. **Gruss, P., C. J. Lai, R. Dhar, and G. Khoury.** 1979. Splicing as a requirement for biogenesis of functional 16S mRNA of simian virus 40. Proc. Natl. Acad. Sci. USA **76**:4317–4321.
28. **Hamer, D., and P. Leder.** 1979. Splicing and the formation of stable RNA. Cell **18**:1299–1302.
29. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**:557–580.
30. **Harris, M.** 1982. Induction of thymidine kinase in enzyme deficient Chinese hamster cells. Cell **29**:483–492.
31. **Helfman, D. M., J. R. Feramisco, J. C. Fiddes, G. P. Thomas, and S. H. Hughes.** 1983. Identification of clones that encode chicken tropomyosin by direct immunological screening of a cDNA expression library. Proc. Natl. Acad. Sci. USA **80**:31–35.
32. **Holmes, D. S., and J. Bonner.** 1973. Preparation, molecular weight, base composition and secondary structure of giant nuclear ribonucleic acid. Biochemistry **12**:2330–2338.
33. **Huynh, T., R. Young, and R. Davis.** 1985. Constructing and screening cDNA libraries in lambda gt10 and lambda gt11, p. 49–78. *In* D. Glover (ed.), DNA cloning: a practical approach, vol. 1. IRL Press, Arlington, Va.
34. **Johnson, L. F., L. G. Rao, and A. Muench.** 1982. Regulation of

thymidine kinase enzyme level in serum stimulated mouse 3T6 fibroblasts. Exp. Cell Res. **138**:79–85.
35. **Karin, M., and R. I. Richards.** 1982. Human metallothionein genes: molecular cloning and sequence analysis of the mRNA. Nucleic Acids Res. **10**:3165–3173.
36. **Keller, C., M. Corcoran, and R. J. Roberts.** 1984. Computer programs for handling nucleic acid sequences. Nucleic Acids Res. **2**:379–384.
37. **Kit, S., D. R. Dubbs, and P. M. Frearson.** 1965. Decline of thymidine kinase activity in stationary phase mouse fibroblast cells. J. Biol. Chem. **240**:2565–2573.
38. **Kit, S., and G. Jorgensen.** 1972. Formation of thymidine kinase and deoxycytidylate deaminase in synchronized cultures of Chinese hamster cells ts for DNA synthesis. J. Cell Physiol. **88**:57–63.
39. **Klevecz, R. P.** 1964. Temporal coordination of DNA replication with enzyme synthesis in diploid and heteroploid cells. Science **166**:1536–1538.
40. **Konecki, D. S., J. Brennand, J. C. Fuscoe, C. T. Caskey, and A. C. Chinault.** 1982. Hypoxanthine-guanine phosphoribosyltransferase genes of mouse and Chinese hamster: construction and sequence analysis of cDNA recombinants. Nucleic Acids Res. **10**:6763–6775.
41. **Kozak, M.** 1984. Comparison and analysis of sequences upstream from the translational start site in eucaryotic mRNAs. Nucleic Acids Res. **12**:857–872.
42. **Kozak, M.** 1984. Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preproinsulin *in vivo*. Nature (London) **308**:241–246.
43. **Kwoh, T. J., and J. A. Engler.** 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.
44. **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.
45. **Lewis, J. A., D. T. Kurtz, and P. W. Melera.** 1981. Molecular cloning of a Chinese hamster dihydrofolate reductase specific cDNA and the identification of multiple dihydrofolate reductase mRNAs in antifolate resistant Chinese hamster lung fibroblasts. Nucleic Acids Res. **9**:1311–1322.
46. **Lewis, J. A., and D. A. Matkovich.** 1986. Genetic determinants of growth phase-dependent and adenovirus 5-responsive expression of the Chinese hamster thymidine kinase gene are contained within thymidine kinase mRNA sequences. Mol. Cell. Biol. **6**:2262–2266.
47. **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.
48. **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.
49. **Littlefield, J. W.** 1965. Studies on thymidine kinase in cultured mouse fibroblasts. Biochim. Biophys. Acta **95**:14–22.
50. **Littlefield, J. W.** 1966. The periodic synthesis of thymidine kinase in mouse fibroblasts. Biochim. Biophys. Acta **114**:398–403.
51. **Lonberg, N., and W. Gilbert.** 1985. Intron-exon structure of the chicken pyruvate kinase gene. Cell **40**:81–90.
52. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
53. **McClelland, M., and R. Ivarie.** 1982. Asymmetrical distribution of CpG, an average mammalian gene. Nucleic Acids Res. **10**:7865–7877.
54. **McKnight, S. L.** 1980. The nucleotide sequence and transcript map of the herpes simplex virus thymidine kinase gene. Nucleic Acids Res. **8**:5949–5962.
55. **McLauchlan, J., D. Gaffney, J. L. Whitton, and J. B. Clements.** 1985. The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3′ termini. Nucleic Acids Res. **13**:1347–1365.
56. **Melera, P. W., J. P. Davide, C. A. Hession, and K. W. Scotto.** 1984. Phenotypic expression in *Escherichia coli* and nucleotide sequence of two Chinese hamster lung cell cDNAs encoding different dihydrofolate reductases. Mol. Cell. Biol. **4**:38–48.
57. **Melton, D. W., D. S. Konecki, J. Brennand, and T. Caskey.** 1984. Structure, expression, and mutation of the hypoxanthine phosphoribosyltransferase gene. Proc. Natl. Acad. Sci. USA **81**:2147–2151.
58. **Merrill, G. F., R. M. Harland, M. Groudine, and S. L. McKnight.** 1984. Genetic and physical analysis of the chicken *tk* gene. Mol. Cell. Biol. **4**:1769–1776.
59. **Mitchison, J. M.** 1971. The biology of the cell cycle. Cambridge University Press, London.
60. **Mittermayer, C. R., R. Bosselman, and V. Bremerskov.** 1968. Initiation of DNA synthesis in a system of synchronized L cells: rhymicity of thymidine kinase activity. Eur. J. Biochem. **4**:487–489.
61. **Mount, S. M.** 1982. A catalogue of splice junction sequences. Nucleic Acids Res. **10**:459–472.
62. **Nordheim, A., and A. Rich.** 1983. Negatively supercoiled simian virus 40 DNA contains Z-DNA segments within transcriptional enhancer sequences. Nature (London) **303**:674–676.
63. **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.
64. **Postel, E., and A. Levine.** 1975. Studies on the regulation of deoxypyrimidine kinases in normal, SV40-transformed, and SV40 and adeno-infected mouse cells in culture. Virology **63**:404–420.
65. **Reynolds, G. A., S. K. Basu, T. F. Osborne, D. J. Chin, G. Gil, M. S. Brown, J. L. Goldstein, and K. L. Luskey.** 1984. HMG CoA reductase. A negatively regulated gene with unusual promoter and 5′ untranslated regions. Cell **38**:275–285.
66. **Rigby, P., M. O. Rhodes, and P. Berg.** 1977. Labelling deoxyribonucleic acid to high specific activity in vitro by nick-translation with DNA polymerase I. J. Mol. Biol. **113**:237–251.
67. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
68. **Schlosser, C. A., C. Steglich, J. R. DeWet, and I. E. Scheffler.** 1981. Cell cycle dependent regulation of thymidine kinase activity introduced into mouse Lmtk$^-$ cells by DNA and chromatin mediated gene transfer. Proc. Natl. Acad. Sci. USA **78**:1119–1123.
69. **Seif, I., G. Khoury, and R. Dhar.** 1979. BKV splice sequences based on analysis of preferred donor and acceptor sites. Nucleic Acids Res. **6**:3387–3398.
70. **Stallings, R. L., G. M. Adair, J. Siciliano, J. Greenspan, and M. J. Siciliano.** 1983. Genetic effects of chromosomal rearrangements in Chinese hamster ovary cells: expression and chromosomal assignment of *TK*, *GALK*, *ACP1*, *ADA*, and *ITPA* loci. Mol. Cell. Biol. **3**:1967–1974.
71. **Stein, J. P., J. F. Catterall, P. Kristo, A. R. Means, and B. W. O'Malley.** 1980. Ovomucoid intervening sequences specify functional domains and generate protein polymorphism. Cell **21**:681–687.
72. **Stubblefield, E., and G. C. Mueller.** 1965. Thymidine kinase activity in synchronized HeLa cell cultures. Biochem. Biophys. Res. Commun. **20**:535–538.
73. **Stubblefield, E., and S. Murphree.** 1967. Thymidine kinase activity in colcemid synchronized fibroblasts. Exp. Cell Res. **48**:652–656.
74. **Sudhof, T. C., J. L. Goldstein, M. S. Brown, and D. W. Russell.** 1985. The LDL receptor gene: a mosaic of exons shared with different proteins. Science **228**:815–822.
75. **Tinoco, I., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla.** 1983. Improved estimation of secondary structure in ribonucleic acids. Nature (London) New Biol. **246**:40–43.
76. **Valerio, D., M. G. C. Duyvesteyn, B. M. M. Dekker, G. Weeda, T. M. Berkvens, L. van der Voorn, H. van Ormondt, and A. J. van der Eb.** 1985. Adenosine deaminase: characterization and expression of a gene with a remarkable promoter. EMBO J. **4**:437–443.

77. **Walker, J. E., M. Saraste, M. J. Runswick, and N. J. Gay.** 1982. Distantly related sequences in the α and β-subunits of ATP synthase, myosin, kinases, and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. **1**:945–951.
78. **Weir, J. P., and B. Moss.** 1983. Nucleotide sequence of the vaccinia virus thymidine kinase gene and the nature of spontaneous frameshift mutations. J. Virol. **46**:530–537.
79. **Wigler, M., S. Silverstein, L. S. Lee, A. Pellicer, Y. C. Cheng, and R. Axel.** 1977. Transfer of purified herpes virus thymidine kinase gene to cultured mouse cells. Cell **11**:223–232.
80. **Woo, S. L.** 1979. A sensitive and rapid method for recombinant phage screening. Methods Enzymol. **68**:389–395.
81. **Yamamoto, K. R., B. M. Alberts, R. Benzinger, L. Lawhorne, and G. Treiber.** 1970. Rapid bacteriophage sedimentation in the presence of polyethylene glycol and its application to large-scale virus purification. Virology **40**:734–744.