MOLECULAR AND CELLULAR BIOLOGY, Sept. 1984, p. 1769–1776
0270-7306/84/091769-08$02.00/0

Vol. 4, No. 9

Genetic and Physical Analysis of the Chicken *tk* Gene

GARY F. MERRILL,† RICHARD M. HARLAND, MARK GROUDINE, AND STEVEN L. McKNIGHT*
Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98104

Received 17 April 1984/Accepted 8 June 1984

Several aspects of the structure of the chicken thymidine kinase gene (*tk*) have been resolved as a result of genetic experiments and nucleotide sequencing. Deletion mapping established the locations of two functional boundaries in a region thought to correspond to the 5′ terminus of the gene. One such boundary coincides with a transcriptional promoter, and the other coincides with the translation start codon of the chicken tk polypeptide. Similar deletion mapping assays identified a functional boundary at the 3′ terminus of the gene. DNA sequence analysis confirms the prediction that this 3′ region encodes the carboxyl terminus of the tk polypeptide. A recombinant cDNA clone complementary to genomic *tk* sequences was isolated. A comparison between genomic and cDNA sequences reveals the locations of six intervening sequences and allows prediction of the complete amino acid sequence of the chicken tk polypeptide.

In animal cells thymidine kinase (tk) and many other enzymes involved in DNA precursor biosynthesis are preferentially expressed during the replicative phase of the cell cycle (3, 6, 11, 19, 24, 26, 27). The levels of these proteins also decline as cells withdraw from the cell cycle during terminal differentiation (20). We are interested in defining the molecular mechanisms that account for these regulatory phenomena.

Direct biochemical analysis of this class of gene products is hampered by the fact that they occur in small amounts. For example, Leys and colleagues have estimated that a proliferating mouse cell may transcribe the gene encoding dihydrofolate reductase (*dhfr*) as infrequently as once per hour (E. J. Leys, G. F. Kraus, and R. E. Kellems, submitted for publication). Our measurements of the cellular *tk* mRNA concentration in proliferating chicken embryo fibroblasts fall in the same range as measurements of the *dhfr* mRNA concentration in standard mouse cell lines.

Cell lines that express substantially elevated levels of *dhfr* mRNA can be established by selection in culture medium containing the folate analog methotrexate. The increased *dhfr* mRNA concentration in methotrexate-resistant animal cells results directly from the amplification of the chromosomal *dhfr* gene (1). Attempts to selectively amplify the cellular *tk* gene have failed. Thus, we have adopted an alternate, genetic approach to study the regulation of tk enzyme expression in animal cells.

Experiments initially reported by Schlosser et al. (23) show that the cellular *tk* gene retains its S-phase-specific expression pattern when transferred as a native chromosomal DNA fragment into an otherwise tk^- host cell. This observation has recently been confirmed and extended with recombinant copies of the cellular *tk* gene of humans (2) and hamsters (10). Also, as described in the accompanying report (17), tk activity is appropriately regulated in differentiating mouse muscle cells that have been transformed with a cloned copy of the chicken *tk* gene. These results indicate that information sufficient for appropriate regulation of tk activity is retained on the recombinant clones bearing the respective cellular *tk* genes. It should therefore be possible to use an in vitro genetic approach to identify components of the cellular *tk* gene responsible for appropriate regulation.

To carry out genetic experiments that focus on the regulation of tk enzyme expression we first needed to resolve the structure of the cellular *tk* gene. To this end we created two nested sets of deletion mutations of the chicken *tk* gene and used the hypoxanthine-aminopterin-thymidine (HAT) selection assay to establish functional boundaries at both ends of the gene. These functional assays, coupled with the resolution of the nucleotide sequence of both genomic and cDNA copies of the gene, identify the complete protein-coding region of the chicken *tk* gene and establish the location of six intervening sequences. This information facilitated experiments described in the accompanying report (17) which localize a component of the gene that is essential for its regulation during terminal differentiation.

MATERIALS AND METHODS

Cell culture and cell transformations. Mouse L cells were maintained at 37°C in a humidified atmosphere of 95% air–5% CO_2. Growth medium was Dulbecco modified minimal essential medium, supplemented with 10% calf serum and antibiotics. For selection and maintenance of transformed lines, growth medium was supplemented with 10^{-4} M hypoxanthine, 4×10^{-7} M aminopterin, and 10^{-5} M thymidine (HAT). Cell transformations were carried out by the calcium phosphate precipitation method (5), as modified by Corsaro and Pearson (4). Cultures were initiated at 5×10^5 cells per 10-cm dish. The next day they were exposed to 0.1 μg of *tk*-containing plasmid and 20 μg of calf thymus carrier DNA. After 24 h, the DNA-containing medium was replaced with fresh growth medium, and after an additional 24 h, HAT selection was begun. Cultures were fixed 12 days after exposure to DNA.

Northern transfer analysis of RNA. The 100 to 500 independent L cell transformants obtained from an individual transformation experiment were pooled and expanded. Cultures containing ca. 10^7 proliferating cells were harvested in 4 ml of SET buffer (10 mM Tris-hydrochloride [pH 7.5], 5 mM EDTA, 1% sodium dodecyl sulfate) containing 0.2 mg of proteinase K per ml. After digestion for 1 h at 50°C, lysates were twice extracted with phenol and chloroform, precipitated with ethanol, and suspended in water. The RNA content was estimated by subtracting the DNA concentration, determined fluorometrically by a dye binding assay (9), from the concentration of total nucleic acids, determined spectrophotometrically at an absorbance at 260 nm. Samples were digested with DNase I (Worthington Diagnostics) that had

* Corresponding author.
† Present address: Department of Biochemistry and Biophysics, Oregon State University, Corvallis, OR 97331.

been pretreated with iodoacetate. DNase I-digested samples were again extracted with phenol and chloroform, precipitated with ethanol, and suspended in water to a concentration of 2 mg of RNA/ml. Samples were denatured with glyoxal, electrophoresed, and transferred to nitrocellulose (25). Hybridizations were carried out in 5× SSC (0.15 M NaCl plus 0.015 M sodium citrate)–25 mM sodium phosphate (pH 6.5)–0.02% bovine serum albumin–0.02% Ficoll 400–0.02% polyvinylpyrrolidone–250 μg of salmon sperm DNA per ml–50% formamide. Radiolabeled probe was prepared by nick translation of a 1,400-nucleotide *Kpn*I-*Bgl*II restriction fragment of the chicken *tk* gene (see Fig. 1). Filters were washed twice with 2× SSC–0.5% sodium dodecyl sulfate, followed by three 15-min rinses in 0.1× SSC–0.5% sodium dodecyl sulfate at 55°C. Filters were exposed to Kodak X-ray film at −70°C with the aid of an intensifying screen.

Manipulation of recombinant *tk* clones. The parental chicken *tk* gene, generously supplied by M. Wigler, consists of a 3-kilobase *Hin*dIII fragment cloned in pBR-322 (21). Extragenic deletion mutations were prepared by enzymatic methods, using exonuclease III and S1 nuclease according to previously published procedures (22). The 5′ deletion was initiated either from the leftward *Hin*dIII restriction endonuclease recognition site (according to the diagram shown in Fig. 1) or from the unique *Eco*RI site. Deleted molecules were ligated to synthetic *Bam*HI linkers, doubly digested with *Bam*HI and *Kpn*I restriction endonucleases, and recloned in a vector that contained all chicken *tk* DNA sequences downstream of the *Kpn*I restriction site. Similar procedures were used to establish two sets of 3′ deletion mutations. Molecules truncated in a region surrounding the 3′ terminus of the chicken *tk* gene were prepared enzymatically after an initial restriction cut at the rightward *Hin*dIII site of the intact parental plasmid (see Fig. 2). Molecules truncated in a region just internal to the 5′ end of the mRNA-coding segment of the chicken *tk* gene were prepared, starting at the unique *Kpn*I restriction endonuclease recognition site. The *Kpn*I-digested substrate was treated with S1 nuclease before exonuclease III digestion to render the molecules flush ended. Endpoints of deletion mutations were judged preliminarily by electrophoretic sizing and then resolved to the nucleotide level by chemical DNA sequencing (13). The nomenclature of both 5′ deletion mutations (Δ5′) and 3′ deletion mutations (Δ3′) derives from the position of the deletion terminus relative to the rightward *Hin*dIII site on the parental chicken *tk* genomic clone. For example, Δ3′-49 maintains a 3′ deletion terminus located 49 nucleotides from the rightward *Hin*dIII site and is therefore missing those 49 residues; Δ5′-2302 maintains a 5′ deletion terminus located 2,302 nucleotides from the rightward *Hin*dIII site and is missing all genomic DNA sequences upstream of its terminus.

Promoter fragments of the chicken and herpesvirus *tk* genes were recombined with their opposing mRNA-coding segments at either the natural *Bgl*II restriction site of the herpesvirus *tk* gene (14) or the synthetic *Bam*HI restriction site that was used to terminate both 5′ and 3′ deletion mutants of the chicken *tk* gene. A terminal fragment of the herpesvirus *tk* gene containing its polyadenylation signals was prepared by ligating a *Bam*HI linker at the *Sma*I restriction site located 20 base pairs (bp) upstream from the viral *tk* translation stop codon (14). This procedure was carried out by using a recombinant herpesvirus *tk* gene that contains a synthetic *Hin*dIII linker 359 bp downstream from the translation stop codon (15). The 379-bp fragment containing the polyadenylation signal of the herpesvirus *tk* gene was prepared by double digestion with *Hin*dIII and *Bam*HI restriction enzymes. This fragment was ligated in vitro to the terminus of each 3′ deletion mutation of the chicken *tk* gene, and the hybrid molecules were cloned in a pBR-322 plasmid vector.

Identification of a chicken *tk* cDNA clone. Polyadenylated RNA was isolated from the testes of 4-month-old roosters. Double-stranded cDNA was synthesized by methods described by Fritsch (12), except that *Escherichia coli* DNA polymerase I was used for second strand synthesis. Synthetic *Eco*RI linkers were ligated to the double-stranded cDNA, and after *Eco*RI digestion, the cDNA was fractionated on a washed column (0.7 by 28 cm) of Sepharose 4B-CL in 0.3 M NaCl–10 mM Tris-hydrochloride (pH 7.8)–1 mM EDTA–0.05% sodium dodecyl sulfate. The trailing fractions containing small cDNA and linkers were discarded. Of the material subjected to S1 nuclease digestion after second strand cDNA synthesis, 51% was recovered from the leading peak of the Sepharose 4B-CL column. Approximately 100 ng of cDNA was ligated to *Eco*RI-digested λgt10 DNA (7). The ligated DNA was packaged in vitro (12) and used to infect *E. coli* C600 *hfl*. Approximately 2×10^6 recombinant bacteriophage were thus obtained. An amplified library was screened for sequences complementary to the chicken *tk* gene. Three positive bacteriophage were obtained from a screen of ca. 10^5 plaques.

Nucleotide sequence analysis. Both genomic and cDNA copies of the chicken *tk* gene were sequenced by the chemical method (13). Roughly 80% of the genomic sequence was resolved by radiolabeling and sequencing from the endpoints of deletion mutations (see Fig. 1 and 2). The remaining 20% of the genomic copy was sequenced by using naturally occurring restriction endonuclease recognition sites. An intact sequence of 2,411 nucleotides was established (see Fig. 4). This genomic sequence progresses from the rightward *Hin*dIII restriction endonuclease recognition site to a point 103 nucleotides to the left of the unique *Eco*RI restriction site (see Fig. 1 and 4). Roughly 50% of the sequence was resolved for both DNA strands. The sequence of the cDNA recombinant was obtained by radiolabeling and sequencing from naturally occurring restriction endonuclease recognition sites that were predicted from the genomic DNA sequence. An intact cDNA sequence of 716 nucleotides was established (see Fig. 4). This cDNA sequence progresses from a point 85 nucleotides to the right of the unique *Bgl*II restriction endonuclease recognition site to a point 20 nucleotides to the left of the unique *Kpn*I restriction endonuclease recognition site (see Fig. 1 and 4).

RESULTS AND DISCUSSION

Functional boundaries of the chicken *tk* gene. A recombinant plasmid containing a 3-kilobase *Hin*dIII insert of genomic chicken DNA efficiently transforms tk^- mouse L cells to the tk^+ phenotype (21). A restriction map of this *Hin*dIII fragment is shown in Fig. 1. Strand-selective hybridization experiments (8) have shown that, as the fragment is drawn, the polarity of transcription is from left to right.

To establish the 5′ boundary of the chicken *tk* gene, a series of deletion mutations missing progressively greater amounts of sequence at the lefthand end of the parental clone were constructed and tested for their ability to transform tk^- mouse L cells to the tk^+ phenotype. The deletion mutations were prepared in vitro by an enzymatic method (see above). The deletion terminus of each mutation was marked by a synthetic *Bam*HI linker. Two deletion muta-

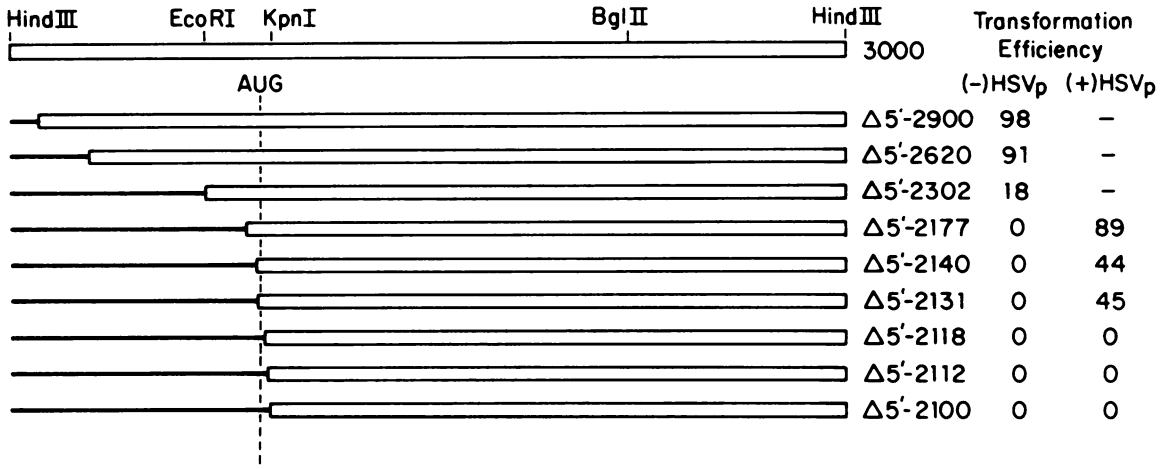

FIG. 1. Structure and relative transformation efficiency of 5′ deletion mutations of the chicken *tk* gene. A restriction map of the parental clone, pCH-tk5 (8), is shown at the top of the figure. A series of 5′ deletion mutations were made as described in the text. All mutations maintain a synthetic *Bam*HI linker at their 5′ terminus. The numerical designation of each deletion mutation corresponds to the distance in nucleotides between the deletion endpoint and the rightward *Hin*dIII site of the parental DNA fragment. Each deletion mutation was tested for its ability to transform tk^- mouse L cells to the tk^+ phenotype as described in the text. Mutations Δ5′-2177, Δ5′-2140, Δ5′-2131, Δ5′-2118, Δ5′-2112, and Δ5′-2100 were also assayed after ligation to a fragment of the HSV *tk* gene that harbors the viral promoter. The assessed transformation efficiency of each deletion mutation is shown to the right of the figure. Values for transformation efficiency represent the average of two determinations carried out on separate days.

tions that terminated to the left of the natural *Eco*RI site are wild-type in their ability to rescue tk^- mouse L cells, and six deletion mutations that terminated to the right of the *Eco*RI site are incapable of rescuing tk^- mouse L cells (Fig. 1). The phenotypes of mutants that maintained a deletion endpoint close to the natural *Eco*RI site are complex. For example, deletion mutant Δ5′-2302, which contains a *Bam*HI linker attached to the *Eco*RI site, was fivefold less efficient than the parental clone in the tk^- L cell rescue assay. In contrast, direct cloning of the *Eco*RI-*Hin*dIII fragment into a pBR-322 *Eco*RI-*Hin*dIII vector resulted in a plasmid which is indistinguishable from the full-length clone in its ability to rescue tk^- L cells (data not shown).

The observation that the nature of the plasmid DNA sequences flanking the *Eco*RI restriction site influences tk enzyme expression might reflect the existence of a nearby transcriptional control region. Although this interpretation is speculative, we have found that a DNA fragment that encompasses this region of the chicken *tk* gene functions as a transcriptional promoter in both cultured mouse cells (17) and frog oocytes (S. McKnight, unpublished data) when it is ligated to a promoterless herpes simplex virus (HSV) *tk* gene.

If deletion mutations terminating downstream from the *Eco*RI site of the chicken *tk* gene are unable to rescue tk^- L cells because they lack a transcriptional promoter, then provision of a promoter from a heterologous gene might restore *tk* function. To test this possibility, the HSV *tk* promoter was ligated to the deletion terminus of all six nonfunctional 5′ deletion mutations, and each recombinant was tested in the L cell rescue assay. The HSV *tk* promoter restored *tk* function when ligated to deletion mutations Δ5′-2177, Δ5′-2140, and Δ5′-2131 but not when ligated to deletion mutations Δ5′-2118, Δ5′-2112, and Δ5′-2100. The complementation observed for the first three deletion mutations (Δ5′-2177, Δ5′-2140, and Δ5′-2131) is consistent with promoter replacement. A possible explanation of the failure of the HSV *tk* promoter to complement the last three deletion mutations (Δ5′-2118, Δ5′-2112, and Δ5′-2100) is that the translation start codon for the chicken tk polypeptide resides between the deletion endpoints of Δ5′-2131 and Δ5′-2118.

Similar methods were used to study the 3′ end of the chicken *tk* gene. We prepared a nested series of deletion mutations missing progressively greater amounts of the right-hand portion of the genomic *tk* fragment. These 3′ deletion mutations are shown diagramatically in Fig. 2, along with their assessed effectiveness in rescuing tk^- mouse L cells. Deletion mutation Δ3′-734, which removes 734 bp of DNA from the right-hand *Hin*dIII site, is the most-truncated deletion mutation capable of efficiently rescuing tk^- L cells. Deletion of an additional 38 bp produces a gene (Δ3′-772) which is an order of magnitude less efficient in the L cell rescue assay. All mutations that maintain endpoints internal to Δ3′-772 are completely incapable of L cell rescue.

The functional boundary that occurs in the vicinity of deletion mutation Δ3′-772 could reflect the location of an element important for transcription termination or polyadenylation. However, this area might encode the carboxyl terminus of the tk polypeptide. To distinguish between these alternatives, we fused to each 3′ deletion terminus a 379-bp fragment of the herpesvirus *tk* gene that contains the viral tk polyadenylation signals (14). Each of these hybrid genes was then tested in the L cell transformation assay. Figure 2 shows the results of these assays. Two observations are noteworthy. First, provision of the HSV *tk* polyadenylation signal did not restore function to deletion mutations shorter than Δ3′-734. This result suggests that the loss of tk function which occurs when the 38 nucleotides between Δ3′-734 and Δ3′-772 are removed is not due to removal of a polyadenylation signal. We are therefore led to believe that this region encodes the carboxy terminus of the chicken tk polypeptide. Second, ligation of the HSV *tk* polyadenylation signal to deletion mutations terminating downstream from Δ3′-772 resulted in a substantial increase in transformation efficiency. In fact, deletion mutations provided with the viral tk polyadenylation signal were even more efficient in rescuing tk^- L cells than the full-length chicken *tk* genomic clone. This set of observations raises the possibility that the

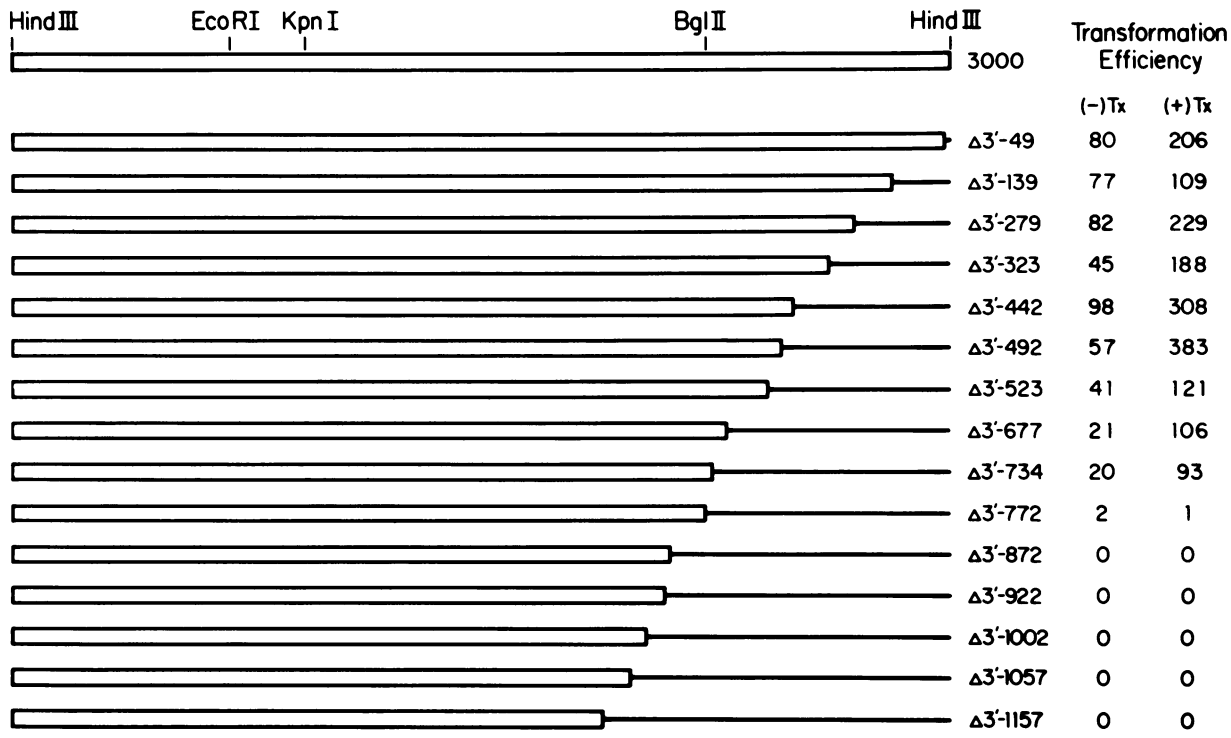

FIG. 2. Structure and relative transformation efficiency of 3′ deletion mutations of chicken *tk* gene. A series of 3′ deletion mutations of the chicken *tk* gene were constructed as described in the text. The numerical designation of each deletion mutation corresponds to the distance in nucleotides between the deletion endpoint and the rightward *Hin*dIII restriction site of the parental DNA fragment. A fragment of the HSV *tk* gene that harbors the viral polyadenylation signal was ligated to the terminus of each 3′ deletion mutation. Mutations containing [(+)Tx] or lacking [(−)Tx] the viral polyadenylation signal were assayed for their ability to transform tk^- mouse L cells to the tk^+ phenotype as described in the text. The results of these assays are presented to the right of the figure. Values for transformation efficiency represent the average of two determinations carried out on separate days.

parental 3-kilobase *Hin*dIII fragment lacks a polyadenylation signal.

Despite that fact that 734 bp of DNA can be removed from the 3′ end of the parental clone without impairing *tk* function, the bulk of this region is transcribed and present on *tk* mRNA produced by transformed L cells. Northern transfer experiments were performed by using total RNA isolated from polyclonal populations of L cells transformed with the parental chicken *tk* clone, 3′ deletion mutations of the parental clone, or 3′ deletion mutations ligated to the HSV tk polyadenylation signal. Cells transformed with the parental *tk* clone (lane 1) gave a very faint band of hybridization at a position corresponding to a transcript length of 2,000 ribonucleotides (Fig. 3). Cells transformed with 3′ deletion mutations of the parental clone (lanes 2, 4, and 6) failed to show a detectable *tk* mRNA signal, even though the level of tk enzyme activity in these transformants is comparable to the level in cells transformed with the parental clone (Fig. 3). In contrast, cells transformed with 3′ deletion mutations to which the HSV *tk* polyadenylation signal is ligated (lanes 3, 5, and 7) produced discrete transcripts, the lengths of which are truncated by an amount equivalent to the extent of 3′ deletion (Fig. 3).

These results suggest that sequences very close to the right-hand *Hin*dIII site of the parental clone might be capable of serving as a polyadenylation signal. Deletion of this signal apparently results in the production of *tk* messages of heterogeneous length that are not detected as a discrete band by the Northern transfer assay. Provision of the HSV tk polyadenylation signal to 3′ deletion mutations establishes a uniform mRNA terminus. Thus, each of the deleted genes that contain the viral tk polyadenylation signal produces a discrete transcript.

Nucleotide sequence of genomic and cDNA copies of the chicken *tk* gene. The deletion mutations described in the previous section facilitated experiments which allow the tentative identification of four functional boundaries of the chicken *tk* gene. The 5′ deletion mutations reveal regions that might correspond to the transcriptional promoter and translation start site of the gene. The 3′ deletion mutations reveal regions that might correspond to the polyadenylation signal and translation stop codon of the gene. To examine these tentative assignments in more detail we have resolved the nucleotide sequence of a genomic copy of the chicken *tk* gene, determined the endpoint of each 5′ and 3′ deletion mutation, and resolved the sequence of a cDNA copy of the chicken *tk* gene. The results of our sequencing experiments are presented in Fig. 4.

(i) Protein coding domain of the gene. Consistent with genetic assignment of the translation start site, an ATG triplet is located in the 13-nucleotide region between the termini of deletion mutations Δ5′-2131 and Δ5′-2118. Recall from Fig. 1 that deletion mutation Δ5′-2131 is restored to a functional state when it is provided with the HSV *tk* promoter, whereas Δ5′-2118 is not. For reasons that will be outlined subsequently we strongly believe that the ATG triplet located within this 13-nucleotide region is the translation start codon for the chicken tk polypeptide.

Our transformation assays of 3′ deletion mutations identified a region close to the naturally occurring *Bgl*II restriction

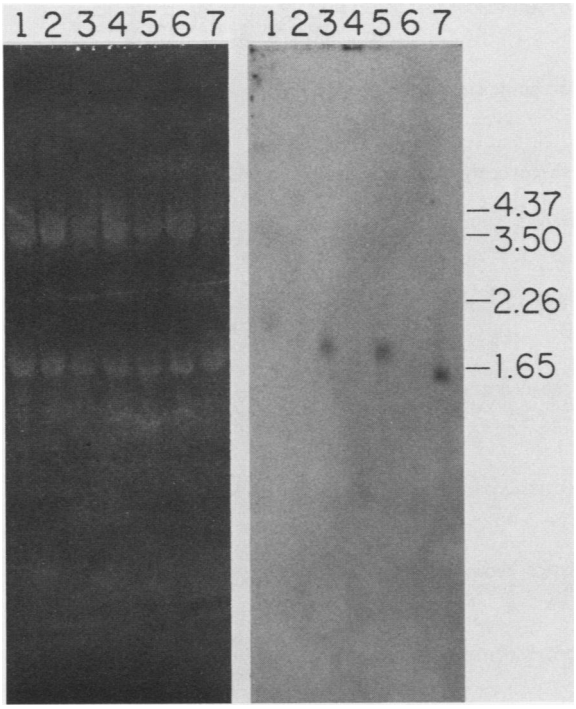

FIG. 3. Northern transfer analysis of *tk* mRNA from mouse L cells transformed with 3′ deletion mutations of chicken *tk* gene lacking or containing an HSV *tk* 3′ terminus. Total RNA was prepared from a pooled population of mouse L cell transformants and sized by gel electrophoresis after glyoxal denaturation. Cells had been transformed with a recombinant plasmid containing either the full-length chicken *tk* gene (lane 1); deletion mutations Δ3′-279, Δ3′-323, and Δ3′-523 (lanes 2, 4, and 6, respectively); or the same three deletion mutations ligated to a DNA fragment containing the polyadenylation signal of the herpesvirus *tk* gene (lanes 3, 5, and 7, respectively). The panel on the left shows the ethidium bromide staining pattern of the gel which confirms the concentration and integrity of each RNA sample. The panel on the right shows an autoradiographic exposure of a Northern transfer of the same RNA samples, probed with radiolabeled chicken *tk* DNA (see the text). Molecular weight markers are glyoxalated *Hin*dIII fragments of bacteriophage lambda DNA and 28S and 18S ribosomal RNA.

enzyme site that might encode the carboxyl terminus of the tk polypeptide. DNA sequence analysis reveals the presence of a TGA triplet between the termini of Δ3′-772 and Δ3′-734. That this triplet indeed represents the translation stop codon is strongly supported by sequence information obtained from a cDNA copy of the chicken *tk* gene.

A cDNA library prepared from chicken testes mRNA was screened for bacteriophage plaques containing sequences homologous to the genomic chicken *tk* gene. Three *tk*-positive cDNA clones were obtained. The recombinant containing the longest cDNA insert was sequenced. The 5′ terminus of the testes *tk* cDNA clone extends to a position 45 bp downstream of the ATG triplet tentatively identified by genetic assays as the translation start codon (Fig. 4). Since there are no potential intron acceptor sites in the genomic sequence for a distance of at least 68 nucleotides upstream from the 5′ terminus of the *tk* cDNA, the 45 nucleotides between the putative translation start site and the 5′ terminus of the *tk* cDNA are likely to be present in the mature *tk* mRNA. The reading frame that is initiated at the putative ATG translation start site is open for 224 amino acids and terminates at the TGA triplet located between the deletion termini of Δ3′-772 and Δ3′-734. Recall that this TGA triplet corresponds to the region identified by genetic experiments to encode the carboxyl terminus of the protein. Finally, the presence of an in-frame TGA stop codon 21 nucleotides upstream from the putative translation start site suggests that this ATG triplet is not an internal methionine codon.

Comparison of the nucleotide sequence of the genomic and cDNA clones reveals the presence of six intervening sequences. The approximate locations of four of these introns were predicted correctly by Kwoh et al. (8) on the basis of insertional mutagenesis experiments. Each of the six introns is flanked by splice donor and acceptor sequences that closely match consensus sequences compiled from the analysis of other eucaryotic structural genes (18). The largest of the six introns segments the first two exons. It is 228 nucleotides in length and notably guanine · cytosine rich (81% guanine-plus-cytosine content). The sizes of the remaining five introns are closely matched, ranging from 79 to 121 nucleotides in length. The largest exon encodes the carboxyl-terminal domain of the polypeptide (52 amino acid residues in length), as well as untranslated sequences of the mRNA extending 3′ from the translation termination codon. The remaining six exons range in length from 66 to 120 nucleotides.

Bradshaw and colleagues (submitted for publication) have isolated and sequenced a cDNA copy of the human *tk* gene. Whereas the human tk polypeptide is predicted to be slightly longer than the length we predict for the chicken tk protein (233 versus 224 amino acids), the two proteins are quite homologous. Discounting the carboxyl terminus of the human tk polypeptide, which extends 9 amino acid residues beyond that of the chicken protein, the two proteins match in amino acid sequence at 196 of 224 residues (82% homology).

(ii) Transcriptional control signals upstream and downstream of the protein-coding domain. Although the boundaries of the protein-coding region of the chicken *tk* gene appear to be soundly resolved, the location of the transcription initiation site and polyadenylation signal are still unknown. Genetic experiments indicate that sequences upstream from the translation start codon on the parental chicken *tk* clone are capable of serving as a transcriptional promoter. Deletion of this region results in loss of tk function, and replacement of this region with the transcriptional promoter of the HSV *tk* gene restores the chicken *tk* gene to a functional state (Fig. 1). Furthermore, when chicken sequences upstream from the translation start site are ligated to a HSV *tk* gene that lacks its own promoter, the resulting recombinant efficiently rescues both tk^- mouse L cells (data not shown) and tk^- mouse muscle cells (17). Furthermore, when this chicken *tk* promoter-HSV *tk* structural gene recombinant is microinjected into frog oocytes, it directs the synthesis of between 20- and 40-fold more tk enzyme than does a promoterless HSV *tk* gene (S. McKnight, unpublished data).

There are several DNA sequence similarities between the putative transcriptional promoter of the chicken *tk* gene and the promoter of the HSV *tk* gene. Both genes contain an *Eco*RI restriction site upstream of their respective protein-coding domains. Both genes contain an adenosine-thymine-rich hexanucleotide 40 to 50 nucleotides downstream from their unique *Eco*RI site (5′-TATTAA-3′ is located ~50 nucleotides from the HSV *tk Eco*RI site; 5′GATAAC-3′ is located ~40 nucleotides from the chicken *tk Eco*RI site). When the HSV *tk* gene is microinjected into frog oocytes, it directs the synthesis of a transcript bearing a 5′ terminus that maps 26 nucleotides downstream from the 5′-TATTAA-3′ hexanucleotide (16). When the chicken *tk* gene is microin-

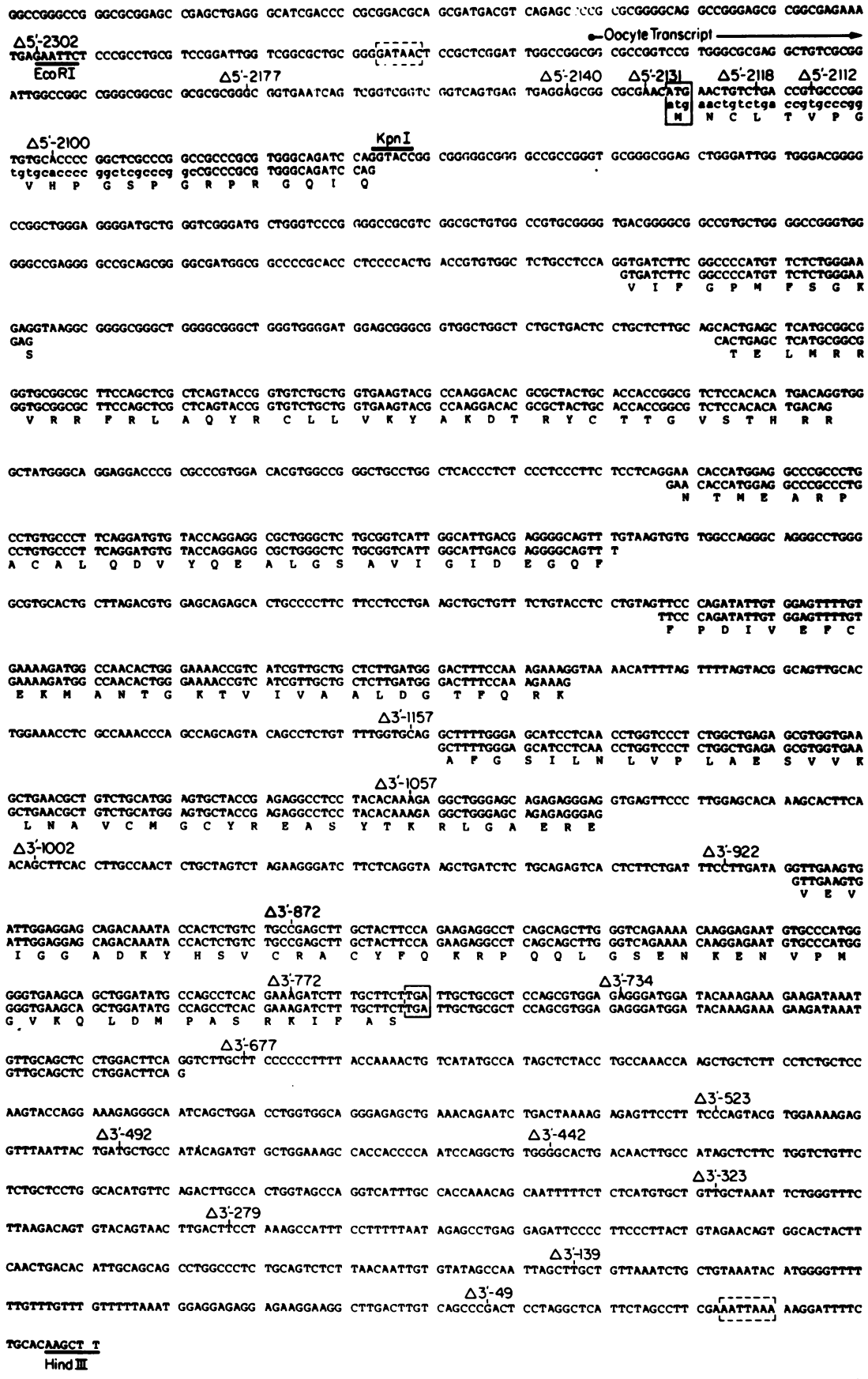

FIG. 4. Nucleotide sequence of the protein-coding region of the chicken *tk* gene. Nucleotide sequence was determined chemically (13), starting from naturally occurring restriction enzyme sites or from the synthetic *Bam*HI site present at the terminus of each 5′ and 3′ deletion mutation (Fig. 1 and 2). The genomic DNA sequence is continuous for 2,411 nucleotides and is presented as the top line of nucleotides. Numerical designations for deletion endpoints represent the distance in nucleotides from the initial A residue of the rightward *Hin*dIII restriction enzyme site. The bracketed hexanucleotide beginning at position 2,262 identifies a potential promoter signal that is located ca. 25 nucleotides upstream from the 5′ terminus of a chicken *tk* transcript synthesized in frog oocytes after microinjection of a plasmid bearing the

jected into frog oocytes, it directs the synthesis of a transcript whose origin maps 25 nucleotides downsteam from the 5'-GATAAC-3' hexanucleotide (McKnight, unpublished data). These results suggest that the sequences surrounding the *Eco*RI restriction site of the chicken *tk* gene may harbor its transcriptional promoter. However, it remains possible that microinjected frog oocytes and transformed mouse cells recognize and utilize these sequences in an adventitious manner and that the authentic chicken *tk* promoter is located somewhere upstream of the unique *Eco*RI site. Unequivocal resolution of this question must await experiments that map the location of the mRNA cap site on authentic chicken *tk* mRNA.

Similar uncertainty concerns the location of the chicken *tk* polyadenylation site. The genomic sequence contains the heptanucleotide sequence 5'-AATTAAA-3' 22 nucleotides to the left of the natural *Hin*dIII restriction site (Fig. 4). Deletion of this region does appear to destroy the ability of the gene to produce a discrete transcript in transformed mouse L cells (Fig. 3). However, deletion of this region does not reduce transformation efficiency, and provision of the HSV tk polyadenylation signal onto any 3' deletion mutations in this vicinity increases transformation efficiency to a level higher than that of the intact 3-kilobase genomic chicken *tk* clone. The testes cDNA clone that we have analyzed does not contain a tract of polyadenylate residues at the 3' boundary of the cDNA insert. We have recently isolated two chicken *tk* cDNA clones from a bursal lymphoma cell line. Unfortunately, neither of the lymphoma-derived cDNA clones contains a polyadenylate tract. However, preliminary sequence analysis of the lymphoma clones indicates that they contain contiguous DNA sequences downstream of the rightward *Hin*dIII site of the genomic clone (S. McKnight, unpublished data). Thus, our tentative identification of a polyadenylation signal close to the rightward *Hin*dIII site may well be incorrect.

Conclusions. The experiments included in this report identify the protein-coding domain of the chicken *tk* gene, predict the amino acid sequence of the chicken tk polypeptide, and locate six intervening sequences. The 3,000-nucleotide genomic DNA fragment that we have used in our studies may, or may not, contain the entire *tk* transcription unit. However, as shown in the accompanying report, this DNA fragment does contain information sufficient to specify appropriate regulation of tk enzyme expression during terminal differentiation of muscle cells.

ACKNOWLEDGMENTS

We thank M. Wigler for provision of a recombinant plasmid carrying the chicken *tk* gene. We also thank R. Kellems, T. Bradshaw, S. Caradonna, P. Deininger, and J. Kwoh for communicating unpublished data; R. Kingsbury, M. Peretz, and K. Hudson for technical assistance; and S. Bizak for clerical assistance.

This work was supported by Public Health Service grant AG 03180 from the National Institutes of Health to G.M. and by a grant from the March of Dimes Foundation to S.M.

LITERATURE CITED

1. **Alt, F. W., R. E. Kellems, J. R. Bertino, and R. T. Schimke.** 1978. Selective multiplication of dihydrofolate reductase genes in methotrexate-resistant variants of cultured murine cells. J. Biol. Chem. **253**:1357–1370.
2. **Bradshaw, H. D.** 1983. Molecular cloning and cell cycle-specific regulation of a functional human thymidine kinase gene. Proc. Natl. Acad. Sci. U.S.A. **80**:5588–5591.
3. **Brent, T.** 1971. Periodicity of DNA synthetic enzymes during the HeLa cell cycle. Cell Tissue Kinet. **4**:297–305.
4. **Corsaro, C. M., and M. L. Pearson.** 1981. Enhancing the efficiency of DNA mediated gene transfer in mammalian cells. Somatic Cell Genet. **7**:603–616.
5. **Graham, F., and A. van der Eb.** 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology **52**:456–467.
6. **Howard, D., J. Hay, W. Melvin, and J. Durham.** 1974. Changes in DNA and RNA synthesis and associated enzyme activities after the stimulation of serum-depleted BHK 21/c13 cells by the addition of serum. Exp. Cell Res. **86**:31–42.
7. **Huynh, T., R. Young, and R. Davis.** 1984. Constructing and screening cDNA libraries in lambda gt10 and lambda gt11. *In* D. Glover (ed.), DNA cloning: a practical approach. IRL Press, Arlington, Va. (In press.)
8. **Kwoh, J., D. Zipser, and M. Wigler.** 1983. Mutational analysis of the cloned chicken thymidine kinase gene. J. Mol. Appl. Genet. **2**:191–200.
9. **Labarca, C., and K. Paigen.** 1980. A simple, rapid and sensitive DNA assay procedure. Anal. Biochem. **102**:344–351.
10. **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.
11. **Littlefield, J.** 1966. The periodic synthesis of thymidine kinase in mouse fibroblasts. Biochim. Biophys. Acta **114**:398–403.
12. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual, Cold Spring Harbor Press, Cold Spring Harbor, N.Y.
13. **Maxam, A., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base specific chemical cleavages. Methods Enzymol. **65**:499–560.
14. **McKnight, S. L.** 1980. The nucleotide sequence and transcript map of the herpes simplex virus thymidine kinase gene. Nucleic Acids Res. **8**:5949–5964.
15. **McKnight, S. L., and E. R. Gavis.** 1980. Expression of herpes thymidine kinase gene in *Xenopus laevis* oocytes: an assay for the study of deletion mutants constructed *in vitro*. Nucleic Acids Res. **8**:5931–5947.
16. **McKnight, S. L., E. R. Gavis, R. Kingsbury, and R. Axel.** 1981. Analysis of transcriptional regulatory signals of the HSV thymidine kinase gene: identification of an upstream control region. Cell **25**:385–398.
17. **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.
18. **Mount, S. M.** 1982. A catalogue of splice junction sequences. Nucleic Acids Res. **10**:459–471.
19. **Navalgund, L. G., C. Rossana, A. J. Meunch, and L. F. Johnson.** 1980. Cell cycle regulation of thymidylate synthetase gene expression in cultured mouse fibroblasts. J. Biol. Chem. **255**:7386–7395.

chicken *tk* gene (McKnight, unpublished data). The arrow proceeding from the putative transcription start site designates transcriptional polarity. Boxed triplets correspond to the putative translation initiation and termination codons for the chicken tk polypeptide. The bracketed heptanucleotide located near the *Hin*dIII site may serve as a polyadenylation signal in L cells transformed with the full-length chicken *tk* clone. The line directly below the genomic sequence shows an exonic sequence corresponding to mature *tk* mRNA. This sequence is derived from a chicken *tk* cDNA clone prepared from testes mRNA (see text). The *tk* cDNA clone extends toward the 5' terminus of *tk* mRNA to a point 45 nucleotides downstream of the putative translation start codon. Nucleotides not actually present on the *tk* cDNA but deduced from the genomic sequence are represented by lower-case letters. The line directly below the cDNA sequence presents the amino acid sequence of the chicken tk polypeptide that is predicted from the genomic and cDNA nucleotide sequences.

20. **O'Neil, M., and R. C. Strohman.** 1969. Changes in DNA polymerase activity associated with cell fusion in cultures of embryonic muscle. J. Cell. Physiol. **73**:61–68.
21. **Perucho, M., D. Hanahan, L. Lipsich, and M. Wigler.** 1980. Isolation of the chicken thymidine kinase gene by plasmid rescue. Nature (London) **285**:207–211.
22. **Sakonju, S., D. F. Bogenhagen, and D. D. Brown.** 1980. A control region in the center of the 5S RNA gene directs specific initiation of transcription. I. The 5′ border of the region. Cell **19**:13–25.
23. **Schlosser, C., C. Steglich, J. de Wet, and I. 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. U.S.A. **78**:1119–1123.
24. **Stubblefield, E., and S. Murphree.** 1968. Synchronized mammalian cell cultures. II. Thymidine kinase activity in colcemid synchronized fibroblasts. Exp. Cell Res. **48**:652–660.
25. **Thomas, P. S.** 1983. Hybridization of denatured RNA transferred or dotted to nitrocellulose paper. Methods Enzymol. **100**:255–265.
26. **Turner, M., R. Abrams, and I. Lieberman.** 1968. Levels of ribonucleotide reductase activity during the division cycle of the L cell. J. Biol. Chem. **243**:3725–3728.
27. **Weidemann, L. M., and L. F. Johnson.** 1979. Regulation of dihydrofolate reductase synthesis in an overproducing 3T6 cell line during transition from resting to growing state. Proc. Natl. Acad. Sci. U.S.A. **76**:2818–2822.