Mutational Analysis of the Cloned Chicken Thymidine Kinase Gene

T. Jesse Kwoh, David Zipser, and Michael Wigler

Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, U.S.A.

Summary: We have analyzed the transcription and coding unit of the chicken thymidine kinase (tk) gene. We have constructed a library of mutant chicken tk genes by the *in vitro* linker insertion method of Heffron et al. A total of 125 mutations within a 3.0 kbp *Hind*III fragment containing the gene were isolated and mapped. The effect of each mutation upon the thymidine kinase gene was determined by measurement of the transfection efficiency in mouse Ltk⁻ cells. The chicken tk mRNA is about 2 kb and polyadenylated. The direction of transcription was also determined. From these results, we propose a structure for the gene in which at least three small introns separate the amino acid coding region into at least four segments. Key Words: Thymidine kinase—*In vitro* mutagenesis—Gene structure—DNA-mediated gene transfer.

We have begun a study of the chicken gene encoding cytoplasmic thymidine kinase (tk). Thymidine kinase is a pyrimidine salvage pathway enzyme which catalyzes the conversion of thymidine to thymidine monophosphate (1). This enzyme is present in almost all actively growing cell types (2). The synthesis of cytoplasmic thymidine kinase is cell cycle dependent (3–6). Because the enzyme is not tissue specific and synthesis seems to be coordinated with DNA replication, we believe the regulation of tk genes may have interesting and novel features. The chicken tk gene was isolated by plasmid rescue, a method which isolates genes based on their ability to cause a phenotypic change in cells (7). The gene, which is contained on a 2.3 kb *Eco*RI–*Hind*III fragment, induces the synthesis of a cytoplasmic tk in mouse Ltk⁻ cells with the same isoelectric point as that of chicken but different from that of mouse or herpes simplex virus-1 (7). As a first step toward studying this gene, we sought to determine the chicken tk gene structure and sequence. We report here the construction *in vitro* of a collection of mutations in the chicken tk gene.

METHODS

Cells

Mouse Ltk⁻ apr⁻ cells were maintained in Dulbecco’s modified Eagle’s media (Flow Labs) containing antibiotics (GIBCO) and 10%
calf serum (Flow Labs). Chicken 249 cells, which were derived from an MC29 induced hepatoma, were obtained from G. Shultz, Heidelberg, and were grown in the above medium plus 2% chicken serum (GIBCO).

**Bacteria and Plasmids**

Plasmids pBR322, pCHTK2, pCHTK3, and pCHTK5 were grown in *Escherichia coli* 294, F<sup>−</sup>, λ<sup>+</sup>, r<sup>x</sup>, m<sup>L</sup>, *endA*, * thi* (obtained from Dr. M. Ptashne). Bacteria were grown in LB media (8) supplemented with 100 μg ampicillin or 12.5 μg tetracycline per ml. The restriction maps of the plasmids are shown in Fig. 1. The isolation of plasmid pCHTK2 has been described previously (7). Plasmid pCHTK5 was derived by cloning the 3.0 kbp HindIII fragment of ACHTK-1, a charon 4A phage isolate containing chicken tk (7), into the HindIII site of pBR322. Plasmid pCHTK3 was obtained by cleaving pCHTK2 with *PvuII*, adding *BamHI* linkers to the *PvuII* created ends, cleaving with *BamHI*, and recyclizing at low DNA concentrations with T4 DNA ligase.

**Nucleic Acid Extraction**

Mini plasmid DNA preparations were made by the phenol lysis procedure of Klein et al. (9). Batch preparations of plasmid DNA from chloramphenicol-amplified cultures (10) were made by the procedure of Guerry et al. (11). High molecular weight cellular DNA was prepared as previously described by Wigler et al. (12). Cytoplasmic RNA was prepared by the method of Sharp et al. (13). Poly-A<sup>+</sup> RNA was purified with oligo-dT cellulose by the method of Aviv and Leder (14).

**In Vitro Mutagenesis**

Plasmids were mutated by a modification of the method of Heffron et al. (15). The restriction enzymes *AluI* (BRL), *HaeIII* (BRL), or *RsaI* (N. E. Biolab), which leave blunt-ended molecules, were used to digest 100 μg pCHTK5 plasmid DNA partially in a 50-μl reaction volume such that 10–30% of the supercoiled plasmids were converted to full-length linear DNA. Digestion was terminated by one phenol extraction and then two chloroform extractions. After ethanol precipitation, the DNA was resuspended in 100 μl 10 mM Tris, pH 8.0, 1 mM EDTA, and stored at −20°C. Eight base synthetic *XhoI* linkers were purchased from Collaborative Research, kinased with ATP by the method of Maxam and Gilbert (16), and reannealed by the method of Bahl et al. (17). Once reannealed, linkers

---

were stored at -20°C and were never heated to more than 4°C until after completion of the ligation reactions. Linkers were mixed with an aliquot containing 2 μg of linear DNA, at a 30:1 molar ratio. Ligation of the linkers to the linear molecules was carried out in a 30-μl reaction volume at 4°C with T4 DNA ligase (BRL) for 24 h. The DNA was then phenol extracted once. CHCl₃ extracted twice, ethanol precipitated, resuspended in 100 μl buffer, and digested with XhoI (BRL). Full-length linear DNA was separated from circular DNA by agarose gel electrophoresis, and the DNA was isolated by electroelution and purified by chromatography on Whatman DE-52. The recovered DNA was again digested with XhoI and then cyclized by ligation at a DNA concentration of 1 μg/ml. The cyclized DNA was used to transform E. coli by the method of Cohen et al. (18). Our procedure varied from that of Heffron et al. (15) in one essential way: restriction enzymes were used to generate flush-end linear plasmid. This cleavage method yields a collection of insertion mutations that is only pseudorandom. In contrast, Heffron et al. (15) constructed a random mutation library by cleaving plasmids with DNase I. However, the staggered DNA ends resulting from DNase I cleavage must be repaired with DNA polymerase prior to linker addition. This repair step results in short sequence duplications surrounding the inserted linker (19) and an undetermined frameshift. Our use of restriction enzymes that leave flush-ends and the use of eight bp synthetic linkers creates a frameshift as well as insertion mutation.

Transfection Assay

Nucleic acid from mini plasmid DNA preparations was digested with Sall and RNase A. After phenol extraction, two chloroform extractions, ethanol precipitation, and resuspension in water, aliquots were subject to agarose gel electrophoresis and stained with ethidium bromide. The plasmid DNA concentrations in the stained gels were estimated after short wavelength UV illumination. Transfection of mouse Ltk⁻ cells was performed as previously described by Wigler et al. (20). Each linker mutant plasmid was used at either 5–10 or 250 ng/10 cm plate. After 2 weeks selection in media containing hypoxantheine, aminopterin, and thymidine (HAT), colonies were fixed with formaldehyde, stained with Giemsa, and counted. In each set of transfections, pCHTK5 and pBR322 were included as positive and negative controls, respectively. The transfection efficiency of each mutant plasmid was determined in at least two different assays.

Other Procedures

Restriction enzymes and E. coli exonuclease III were purchased from Bethesda Research Laboratories and New England Biolab. The manufacturer-recommended digestion conditions were used. Blot hybridizations of DNA and RNA were performed by the methods of Southern (21) and Thomas (22), respectively. Nick translation of DNA was performed as described by Peruchó et al. (23). Labeling of the 5’-end of RNA and DNA and the 3’-end of DNA were performed as described by Maxam and Gilbert (16) and by Challberg and Englund (24), respectively. DNA sequence determinations were performed by the chemical degradation procedure of Maxam and Gilbert (16). The method of mapping mRNA using nuclease S1 was performed as described by Berk and Sharp (25), except nuclease-resistant heteroduplexes were detected by the blot-hybridization procedure of Southern (21).

RESULTS

To begin our analysis of the chicken tk gene, we constructed a library of "pseudorandom" linker insertion mutations. This method entails partial digestion of the chicken tk plasmid, pCHTK5, with frequently cutting restriction endonucleases, preparation of full unit-length linear molecules, addition of synthetic XhoI "linker" molecules to the blunt-ended linear, cyclization after XhoI cleavage, and transformation of E. coli with the resulting constructs.
XhoI linkers were chosen, since pCHTK5 contains no XhoI restriction endonuclease sites. A total of 495 transformants were screened, of which 206 were derived from the AluI digestion, 180 were obtained from the HaeIII digestion, and 109 were made from the Rsal digestion. Table 1 summarizes the analysis of these transformants. The plasmids were designated XA... XH... XR... depending upon whether they were derived after AluI, HaeIII, or Rsal digestion, respectively. Digestion of mini plasmid DNA preparations with XhoI showed that 168/495 of the transformants contained a new XhoI site. The low yield of mutant plasmids probably results from incomplete separation of the linear form III plasmid DNA and nicked circular form II plasmids.

Double digestion with XhoI and HindIII was used to determine whether the new XhoI site was located within the 3.0 kbp tk fragment or within the pBR322 sequence. A total of 125 plasmids with XhoI linkers in chicken sequences were found.

Digestion of the chicken sequence mutants with XhoI and SalI allowed us to form an approximate order of the mutations along the 3.0 kbp chicken DNA fragment. The precise position of the linker insertion mutation was determined using side-by-side comparisons of the mutant plasmids in double and triple restriction enzyme digests. Insertion locations differing by 20–50 bp could be detected this way. This analysis also revealed that not all mutations were simple insertions. Several plasmids contained small sequence deletions at the XhoI linker addition site, and several were duplications separated by the XhoI linker. These structures were confirmed by sequencing in several instances. They presumably arose during the preparation of full unit-length linear plasmids after partial digestion. Such a preparation would inevitably be contaminated by monomeric plasmid molecules cleaved twice at nearby restriction sites (yielding deletion/insertions) and by dimeric plasmid molecules cleaved at restriction sites on each DNA unit (yielding duplication/insertions). Figure 1 shows the location and type of mutations obtained.

The combined mutation library has 41 unique simple linker insertion mutations (Table 1). The greatest distance between mutations is the 200 bp gap between the mutation sites of XH10 and XR3. The average distance between mutations is 74 bases and the smallest distance is less than 10 bases. It is likely that our library of mutant plasmids contain insertions at all Rsal sites since we have multiple independent insertion mutations at most of the Rsal restriction sites (Fig. 2). Not all possible insertions of AluI or HaeIII sites have been obtained. A number of mutations were found to inactivate known restriction sites. The sequence recognized by AluI is internal to the HindIII, SstI, and PvuII recognition sequences. Linker insertion into the AluI site alters the hexameric sites for the above enzymes. Plasmids XA97 and XA2 in-

<table>
<thead>
<tr>
<th>TABLE 1. Mutant plasmids derived from pCHTK5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restriction enzyme*</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Transformants</td>
</tr>
<tr>
<td>AluI</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Mutants</td>
</tr>
<tr>
<td>206</td>
</tr>
<tr>
<td>pBR322 mutants</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>Chicken tk mutants</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>Insertion mutants</td>
</tr>
<tr>
<td>20 (14)</td>
</tr>
<tr>
<td>Deletion mutants</td>
</tr>
<tr>
<td>9 (8)</td>
</tr>
<tr>
<td>Duplication mutants</td>
</tr>
<tr>
<td>1 (1)</td>
</tr>
</tbody>
</table>

* Restriction enzyme used to generate linear DNA (see Methods).
* Number of mutant plasmids isolated with the number of mutation locations in parentheses.

activate a HindIII site and a PvuII site, respectively, by linker insertion. SstI and PvuII sites are destroyed in XA36, XA26, and XA33 because each of the plasmids has a deletion endpoint within the enzyme recognition sequences. The mutation in plasmid XR4 inactivates the KpnI site by linker insertion into the Rsal site within the KpnI hexameric sequence.

Each mutant plasmid was used in the mouse Ltk" transfection assay system to determine whether the mutation altered the chicken tk gene activity. Figure 3 shows the transfection efficiency of linker insertion mutants compared to wild-type plasmid. Mutations in the 0.00–0.80 kbp and 2.22–3.00 kbp portions of the cloned chicken DNA did not disturb gene activity. The region from 0.80 to 2.22 kbp is a mosaic of sites sensitive to or resistant to linker insertion. Four "islands of resistance" are observed in this region. Since the linker used is octameric, it should cause frameshift mutations when inserted into the coding region. The most likely explanation of these "islands of resistance" is, therefore, that they correspond to intron sequences. Since independent plasmid isolates with the same linker insertion site gave similar transfection results, we can rule out retention of activity in a mutant plasmid due to the insertion of three XhoI linkers. In untranslated sequences, linker insertion mutations would alter gene activity only if the insertion site was within a regulatory sequence. However, this latter type of mutant should be rare, given the small target size of regulatory sequences. The transfection results suggest that the coding region for tk stretches from coordinate 0.90 kbp (mutant XH136) to coordinate 2.19 kbp (mutant XA98). In general, the trans-
fection results for the deletion and duplication mutations agreed with those of the insertion mutations (results not shown). However, one deletion mutant did not conform to this rule. The mutant plasmid, XH154, which contains a deletion that spans two resistant mutant plasmids, XH40 (at 0.80 kbp) and XH44 (at 0.61 kbp), had a very low (although detectable) gene activity (data not shown). Since the original plasmid res阮ant, pCHTK2, lacks the sequences from 0.00 to 0.72 kbp, the loss of activity in XH154 indicates some sequence important for tk expression lies between the EcoRI site (at 0.72 kbp) and the linker insertion site of XH40 (at 0.80 kbp).

In addition to the analysis of mutant plasmids, we have determined the size of the tk transcript and the direction of transcription. Poly-A selected cytoplasmic RNA from chicken tissue culture cells (249 cells) was fractionated in agarose gels. Transfer to nitrocellulose and subsequent hybridization with 32P-labeled pCHTK2 revealed a single band (Fig. 4). We estimate the size of tk mRNA to be ~2.0 kbp. A mRNA of the same size is found in chicken embryos and in mouse cells transfected with either λCHTK-1 or pCHTK2 (results not shown). Mouse Ltk− poly-A mRNA alone does not hybridize to pCHTK2.

To determine the direction of transcription,
we prepared probes labeled only on one strand (Fig. 5). The plasmid pCHTK3 was cleaved either by EcoRI or BamHI. EcoRI treatment followed by DNA synthesis with \(^{32}P\)-dXTP and the E. coli polymerase I large fragment resulted in the tk region being labeled only on one coding strand. We then used these asymmetrically labeled probes in Northern blots to poly-A mRNA from a transfected mouse cell which contained 20-fold more chicken tk activity than 249 cells. The results obtained indicate that transcription begins near the EcoRI site and proceeds toward the BamHI site (Fig. 5).

An attempt was made at mapping the tk mRNA by the S1 nuclease method of Berk and Sharp (25). The 3.0 kbp HindIII fragment containing the tk gene was purified from pCHTK5 and hybridized to poly-A\(^+\) cytoplasmic RNA from chicken 249 tissue culture cells. Digestion of the resultant hybrids with nuclease S1 leaves a DNA fragment that is \(\sim 900\) bases long (Fig. 6). Even with longer autoradiographic exposure times, we could not detect any other fragments.

**DISCUSSION**

We have constructed a library of linker insertion mutations for the cloned chicken tk gene. This library is useful for a) mapping of the tk coding region; b) determination of the chicken tk DNA sequence; c) accessing regions of the tk gene for subsequent in vitro mutagenesis; and d) studying DNA recombination in animal cells. The use of the mutant plasmids to determine the DNA sequence of the cloned chicken tk gene will be presented in a subsequent publication. In this paper, we present the genetic mapping of the tk coding region. A simple assay system, transfection of tk\(^+\) mouse cells, makes this genetic approach feasible.

The analysis of the mutations, coupled with the determination of the mRNA size, transcription direction, and the S1 mapping data, allow us to propose a tentative structure of the chicken tk mRNA. Transcription begins near the EcoRI site at coordinate 0.72 kbp and proceeds toward the HindIII site at position 3.00 kbp. Since the poly-A\(^+\) mRNA-containing chicken tk sequence is about 2.0 kb, the 3\(^{\prime}\)-terminus of the mature mRNA is possibly quite close to the HindIII site at 3.00 kbp. Indeed, DNA sequence data reveal the presence of the heptanucleotide AATTTAA, which can serve as a poly-A addition signal (26,27), about 20 bases upstream from the HindIII site (unpublished results). No other AATTTAA or AATTAAA sequences were found in the cloned HindIII fragment. Since the region between 2.22 kbp and 3.00 kbp is resistant to mutations, we infer this region constitutes the 3\(^{\prime}\) untranslated region, whereas the functional region of tk is between coordinates 0.80 and 2.22 kbp. This region contains at least three, possibly four, small introns dividing the coding region into at least four, possibly five, small segments. The existence of introns is inferred from the retention of gene activity by mutants which map within the coding region of the gene. The uncertainty in the number of introns and exons arises from not knowing whether the linker insertion at 0.92 kbp disrupts regulatory or coding
sequences. Additional very small introns could have been missed by our mutant scan. Other introns may exist downstream from the coding region (in the 2.22–3.00 kbp region) and would not be detected by our method of assay.

In order to map the location of introns further, we performed a nuclease S1 analysis of the tk gene. A 0.9 kb DNA fragment of the cloned HindIII was protected from nuclease digestion by heteroduplex formation with chicken poly-A+ cytoplasmic RNA. The only region of the tk gene large enough to accommodate the 0.9 kb fragment is the stretch that includes the last coding segment (at 2.19 kbp) and the 3'-untranslated region (2.22 to 2.98 kbp). The region from 0.72 to 2.08 kbp contains too many introns to accommodate the 0.9 kb fragment. Therefore, the results of the S1 analysis indicate that the 3'-untranslated region is free of introns. The absence of any other segments in the S1 digest besides the 0.9 kb fragment suggests that more introns may exist in the region from 0.72 to 2.04 kbp. RNA–DNA heteroduplexes such as those that would form for the exons at 1.74 and 1.95 kbp would likely escape detection by S1 analysis under the conditions used. However, a 0.39 kbp exon (coordinate 1.18–1.56 kbp) should be detectable, and the absence of such a fragment in our S1 digest suggests a small undetected intron(s) may exist in this region. Since the scarcity of mRNA (which we estimate at 0.001–0.005% of total mRNA) and the small size of the exons have made S1 analysis very
GENETIC ANALYSIS OF THYMIDINE KINASE

FIG. 6. Nuclease S1 mapping of tk mRNA. The 3.0 kbp HindIII chicken tk fragment was purified from pCHTK5 and hybridized to poly-A+ cytoplasmic mRNA from chicken 249 tissue culture cells. Nonhomologous regions were removed by nuclease S1 digestion, and the resultant digest was fractionated by alkaline agarose gel electrophoresis. The biotemplating procedure of Southern (21) was used to detect the fractionated DNA. Nick-translated pCHTK2 was the hybridization probe. Lane A, HindI digest SV40 DNA kinased with [γ-32P] ATP; lane B, no S1 treatment; lane C, S1 digest.

difficult, a more definitive description of the chicken tk transcriptional unit must await the cloning of the cDNA.

Acknowledgments: We wish to thank Celia Fraser and Mitch Goldfarb for technical assistance. This work was supported by grants from the National Institutes of Health and the Robertson Research Fund.

REFERENCES

3. Stubblefield E. Murphree S. Synchronized mamma-

20. Wigler M, Pellicer A, Silverstein S, Axel R. Biochemical transfer of single copy eucaryotic genes


