Volume 12 Number 9 1984 Nucleic Acids Research

The nucleotide sequence of the chicken thymidine kinase gene and the relationship of its predicted polypeptide to that of the vaccinia virus thymidine kinase

T.Jesse Kwoh and Jeffrey A.Engler[1]

Department of Veterinary Microbiology, University of Saskatchewan, Saskatoon, Saskatchewan S7N 0W0, Canada, and [1]Department of Biochemistry, University of Alabama in Birmingham, University Station, Birmingham, AL 35294, USA

Received 20 February 1984; Accepted 11 April 1984

ABSTRACT

The entire DNA nucleotide sequence of a 3.0 kilobase pair HindIII fragment containing the chicken cytoplasmic thymidine kinase gene was determined. Oligonucleotide linker insertion mutations distributed throughout this gene and having known effects upon gene activity (Kwoh, T.J., Zipser, D., and Wigler, M. 1983. J. Mol. Appl. Genet. 2, 191-200), were used to access regions of the HindIII fragment for sequencing reactions. The complete nucleotide sequence, together with the positions of the linker insertion mutations within the sequence, allows us to propose a structure for the chicken thymidine kinase gene. The protein coding sequence of the gene is divided into seven small segments (each less than 160 base pairs) by six small introns (each less than 230 base pairs). The proposed 244 amino acid polypeptide encoded by this gene bears strong homology to the vaccinia virus thymidine kinase. No homology with the thymidine kinases of the herpes simplex viruses was found.

INTRODUCTION

The thymidine kinase genes have proven to be extremely useful as selective markers for introducing desired genes into animal cells (1). Much of this work involves the herpes simplex virus thymidine kinase genes, which were used to demonstrate that individual genes could be inserted into cells by DNA-mediated gene transfer, more commonly referred to as transfection (2, 3). The ability to transfect genes and the powerful positive and negative selection methods for the thymidine kinase genes (4) has in turn allowed the isolation of both viral and cellular thymidine kinase genes. To date, the viral thymidine kinase genes have been isolated from both herpes simplex virus types 1 (5-8) and 2 (9) and from vaccinia virus (10, 11). The genes for cytoplasmic thymidine kinase have also been cloned from chicken (12), human (13, 14) and hamster (15) cells. The transfection and selection methods have also aided the studies on the regulation of expression of the herpes simplex (16, 17) and chicken (18) thymidine kinase genes.

We have been studying the chicken thymidine kinase gene. Cytoplasmic thymidine kinase genes are expressed in almost all actively growing cell types and their expression is cell cycle dependent (19-22). In synchronized cells, enzyme levels are low or absent during G1 phase, rise

© IRL Press Limited, Oxford, England.

during S phase, and then decline during G2. As a first step toward examining the control mechanisms involved in this regulation, we have characterized the structure of the cloned chicken gene using both genetic and physical approaches. Previously, we reported the construction of a library of 125 linker insertion mutants which have an eight base pair oligonucleotide inserted at random locations within the cloned 3.0 kilobase pair (kbp) HindIII fragment containing the active chicken thymidine kinase gene (18). These mutants were used to analyze the structure of the coding region for chicken thymidine kinase. By transfection of each mutant plasmid into mouse Ltk^- cells and selection for conversion to tk^+, it was found that some mutations within the putative gene have no effect on gene activity, suggesting that these frameshifting insertion mutations lie within introns. We were able to propose a structure for the chicken thymidine kinase gene based upon this mutational analysis and upon physical characterization of the mRNA (18). The transcriptional unit is 2.3 kbp long and results in a mRNA that is 2.0 kilobases (kb) long including the poly-A tail. The gene appears to contain at least four small introns which separate the protein coding sequences into at least five segments. The gene has a 900 base pair (bp) 3'-untranslated region.

Here we report the use of the linker mutations described previously to determine the DNA nucleotide sequence of the 3.0 kb HindIII fragment that contains the chicken thymidine kinase gene. The locations of many of the linker mutations have been precisely aligned with the DNA sequence. By correlating the positions of the linker mutations that abolish thymidine kinase activity with their position along the nucleotide sequence, we have tentatively identified the probable coding frame and splice junctions for the chicken thymidine kinase mRNA. When the amino acids encoded in this presumptive coding frame were compared with the polypeptide sequences of other thymidine kinase genes, the vaccinia virus thymidine kinase polypeptide was highly conserved with respect to chicken thymidine kinase polypeptide whereas herpes virus thymidine kinase polypeptide shares no apparent homology.

MATERIALS AND METHODS

(a) Preparation of plasmid DNA

The wild-type and linker insertion mutant plasmids containing the chicken thymidine kinase gene that were used for determination of the DNA nucleotide sequence have been described previously (18). The plasmid DNA was grown in *E. coli* strains 294 (18) and HB101 (23), and cleared lysates of the chloramphenicol treated bacteria were prepared by the methods of either Guerry *et al.* (24) or Birnbaum and Doly (25). The plasmid DNA was purified by banding in CsCl-ethidium bromide density equilibrium gradients.

(b) Enzymes and buffers

Restriction endonucleases were purchased from New England Biolabs or Bethesda Research Laboratories and the manufacturer-recommended digestion conditions were used. Agarose gel electrophoresis buffer (E buffer)

consisted of 40 mM Tris base and 2 mM EDTA adjusted to pH 7.2 with glacial acetic acid. TBE buffer for polyacrylamide-urea sequencing gels was 135 mM Tris base, 45 mM boric acid, and 2 mM EDTA.

(c) Nucleotide sequence analysis

Ten to 20 pmole of plasmid DNA was digested to completion with a restriction enzyme. To sequence from a linker insertion mutation, the plasmid was digested with XhoI. BglII, HindIII, XbaI or EcoRI were used to sequence from naturally occurring restriction enzyme sites. The DNA was extracted with phenol and chloroform and precipitated with ethanol. The 3'-ends of the resulting linear molecules were filled-in with [α-^{32}P] dATP or dTTP (3000 Ci/mmol; Amersham or New England Nuclear), the three other deoxynucleotides, and the large fragment of DNA polymerase I (New England Nuclear). After labeling for 30 minutes at 12°C unincorporated nucleotides were removed by several ethanol precipitations and the labeled DNA was redigested with a second restriction endonuclease. The products of the digestion were separated by electrophoresis on an agarose gel in E buffer. After autoradiography, the labeled DNA was recovered from the excised gel slices either by electroelution or by using sodium iodide and glass powder, as described by Vogelstein and Gillespie (26). The nucleotide sequences were determined by the chemical degradation method of Maxam and Gilbert (27). The sequences obtained were recorded and analyzed using computer programs developed by Staden (28, 29), Gingeras et al. (30), and Blumenthal et al. (31). Hydrophobicity plots of the predicted amino acid sequences were calculated as described by Kyte and Doolittle (32), using computer programs written by Dr. J. Feiffer.

RESULTS AND DISCUSSION

The chicken thymidine kinase gene used in our studies is carried on a 3.0 kbp HindIII fragment. This fragment was subcloned into plasmid pBR322 from a Charon 4a recombinant phage, which had been isolated from a genomic library of chicken DNA (12). Fig. 1a shows the restriction endonuclease cleavage map for the 3.0 kb HindIII fragment that encodes the chicken thymidine kinase gene. The following restriction endonucleases cleave this fragment once: BglII, ClaI, EcoRI, KpnI and XbaI. SstI and PvuII each cut the fragment twice. Previously, we have characterized this gene by inserting an eight base XhoI linker with the sequence CCTCGAGG into the middle of random AluI, HaeIII and RsaI sites (18). Based upon transfection studies of these linker insertion mutations, we proposed that the protein coding sequence, which is divided into at least five small segments by at least four small introns, is located between the EcoRI and BglII sites. The 900 bp 3'-untranslated region of the mRNA is encoded by the segment between BglII and HindIII (18). Fig. 1b shows the sequencing strategy and the positions of XhoI linker mutations used for DNA sequencing. Greater than 95% of the sequences for both DNA strands has been determined.

Fig. 2 shows the DNA nucleotide sequence of the l-strand (of the same

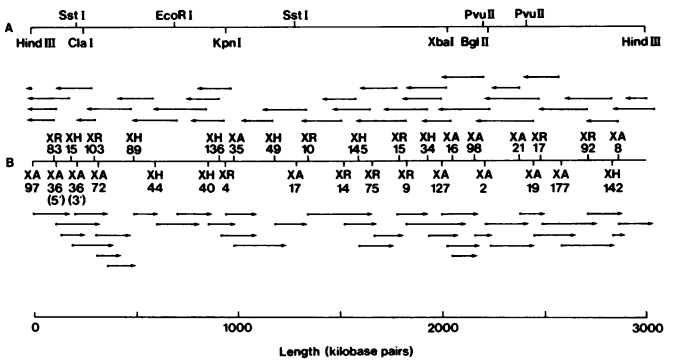
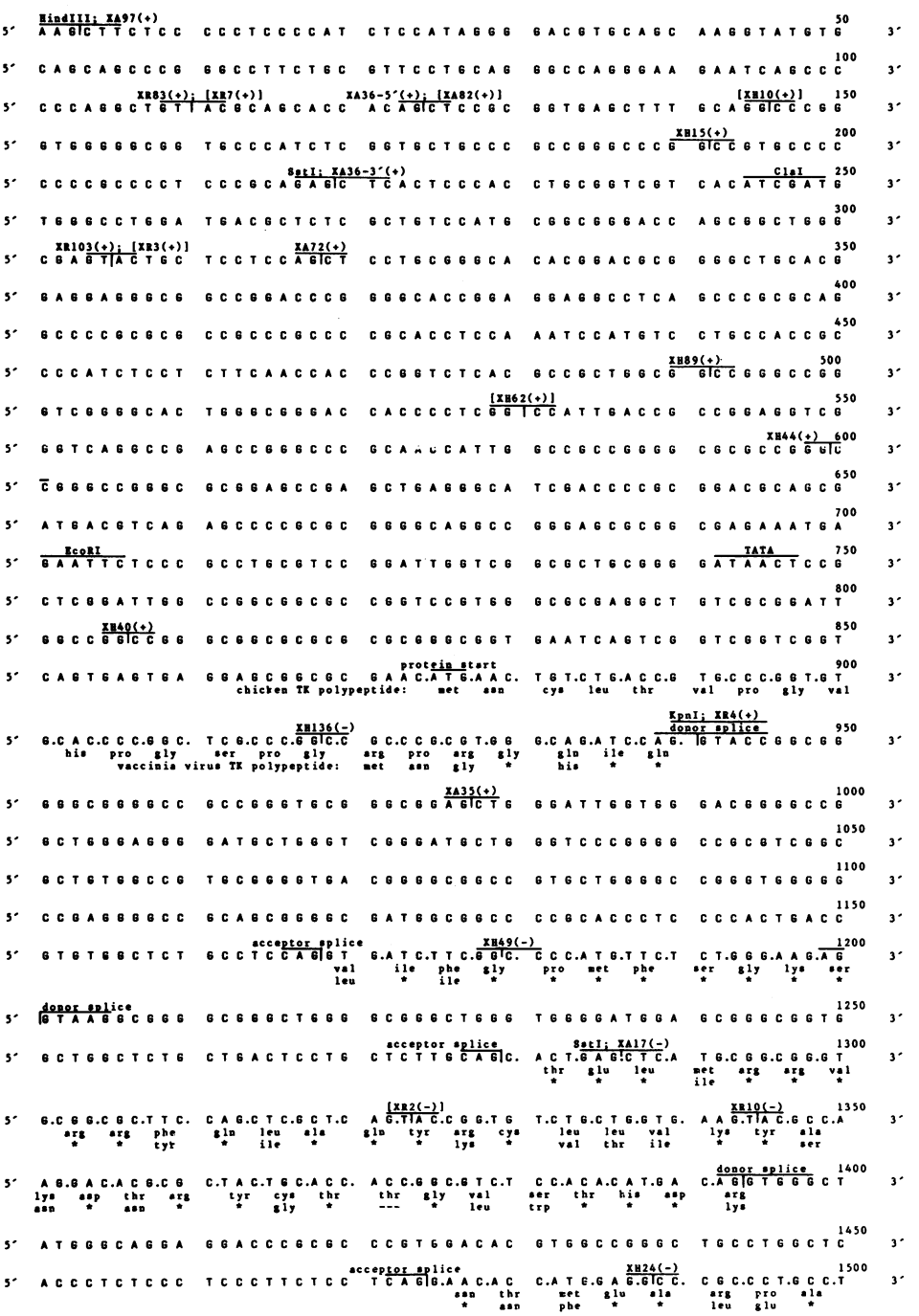


Fig. 1: Restriction endonuclease cleavage sites and sequencing strategy for the 3.0 kb HindIII fragment encoding the chicken thymidine kinase gene. (a) Cleavage sites within the fragment. (b) Sequencing strategy for the fragment. Sequences covered in individual reactions are shown as arrows pointing from 3' (the labeled end) to 5'. The names of the linker insertion mutations used for the sequencing are shown above and below the line.

polarity as the presumed chicken tk mRNA). The entire sequence is 3008 bp long. Since linker insertion mutations were used to access regions of the cloned fragment, the precise location of these mutations was determined in the course of this work. The locations of 34 mutations at 32 separate sites was determined experimentally. Of the 34 mutations, only one showed any sequence abnormality at the mutation site. Mutation XR103 at base 304 contains a nine base insert, CCTCGGAGG, on the 5' side of the linker (data not shown). The additional sequence resembles the linker sequence and may have resulted from a minor contaminant of the linker preparation used. In addition to the mutation positions determined experimentally, we were able to infer the location of 15 additional mutations at 10 positions based upon whether the mutation was derived from linker insertion into AluI, HaeIII or RsaI sites and the relative order of the mutations as determined by restriction enzyme mapping. Due to multiple isolates, another 51 mutations could be assigned to sites that were experimentally located.

Previously, we had proposed that transcription begins in the segment between EcoRI at base 701 and the KpnI site at base 940 and proceeds rightward to the HindIII site at base 3003 (18). However, since the gene contains at least four small introns which split the protein coding sequence into at least five small segments, and because the mRNA occurs in very low abundance in both chicken cells and in mouse cell transfectants, the precise 5'-end of the mRNA and the location of the splice junctions are unknown (18). In Xenopus oocytes injected with the chicken thymidine kinase gene, transcription begins at bases 766,767 and 769 (33). If transcription in chicken and mouse cells intitiates at the same bases as Xenopus transcription, then regulatory sequences should be located upstream

CHICKEN THYMIDINE KINASE DNA SEQUENCE:



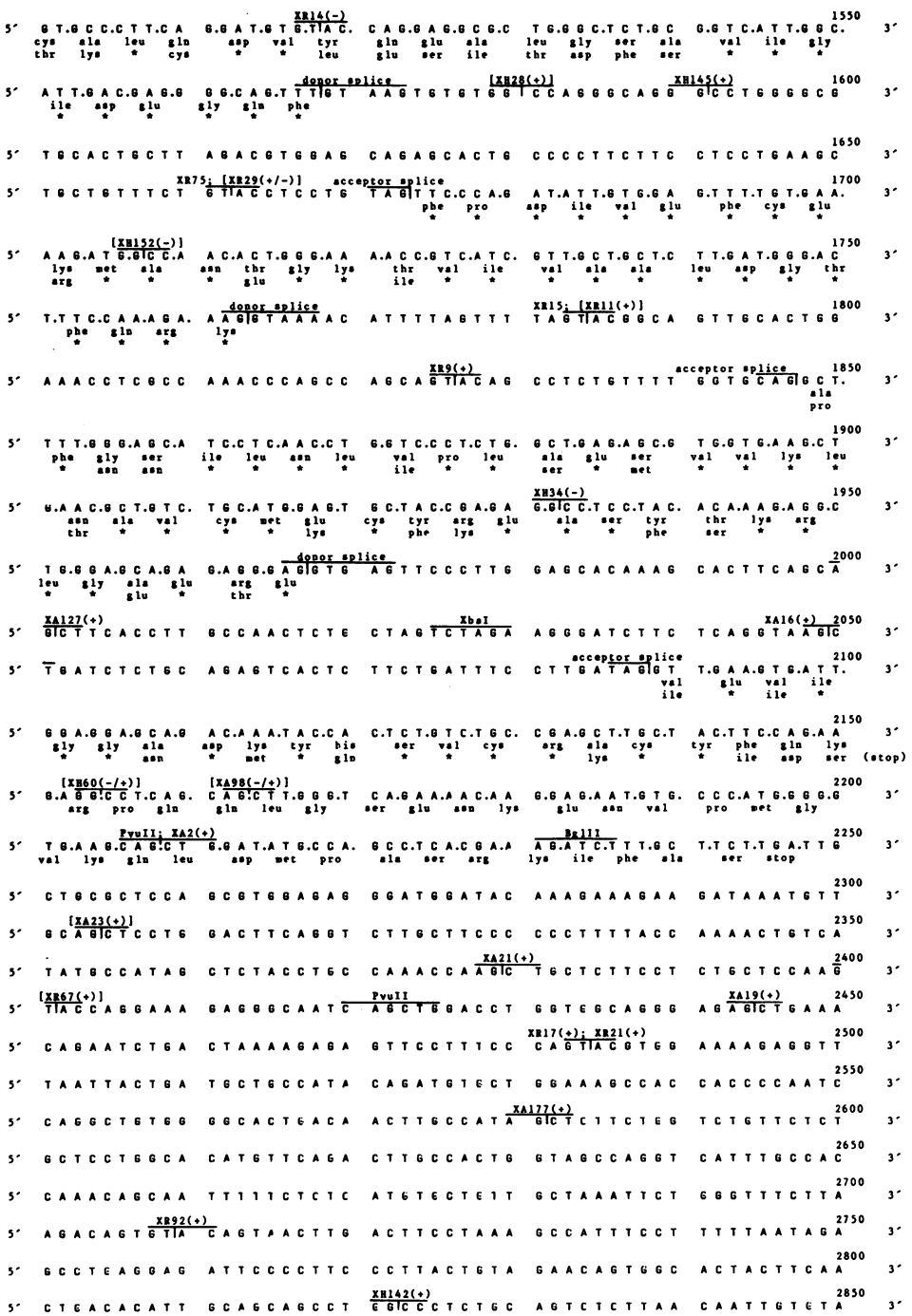


Fig. 2: The nucleotide sequence of the DNA encoding the chicken thymidine kinase gene. The sequence shown is of the same polarity as that of the chicken thymidine kinase mRNA. The positions of restriction endonuclease cleavage sites and linker insertion mutations are indicated above the sequence. The positions of mutations shown in brackets were assigned based upon restriction enzyme mapping. The effect of each linker mutation on tk activity is indicated by (+) if the mutated gene were still able to transfect mouse Ltk$^-$ to tk$^+$ and (-) if the mutated gene could not. Those mutations with a reduced ability to donate thymidine kinase in transfections were shown as (+/-) and (-/+) based upon their relative abilities (18). The amino acid sequence of the predicted chicken thymidine kinase polypeptide and the positions of donor and acceptor splice points are indicated. In addition, the amino acid sequence of the vaccinia virus thymidine kinase polypeptide (10, 11) is written below the predicted chicken sequence for comparison. Where the two sequences agree, an * is shown; in addition a one amino acid deletion between the two thymidine kinase sequences is indicated by --- (nucleotides 1371-1373).

of base 766. No obvious CAAT- or TATA-like sequences are apparent. Such sequences have been found to be involved in transcription initiation of other animal genes, and are usually located 80 and 30 bases upstream of the penultimate mRNA base, respectively (34). In the chicken thymidine kinase gene, the sequence ATAACT at base 742 is 25 bases upstream of the putative mRNA start site used in Xenopus and may be a variant of the TATA sequence.

The exact 3'-end of the mRNA is unknown since a cDNA clone of the chicken thymidine kinase mRNA has not yet been isolated. Hybridization-nuclease S1 protection experiments suggest the 3'-end is very close to the HindIII site (18). The canonical polyadenylation signal sequence, AATAAA (35), was not found in the sequence. However the sequence AATTAAA was found at base 2981, in the region where the 3'-terminus of the mRNA is approximately locate. This variant sequence has been shown to function as a polyadenylation signal in a variety of other eukaryotic mRNAs, including early region 3 of adenovirus-2 (36), chicken lysozyme (37), anglerfish somatostatin (38), mouse pancreatic α-amylase (39), rat skeletal muscle actin (40), and some interferons (41). By analogy to these other examples, this variant sequence may also be used as the polyadenylation signal sequence for chicken thymidine kinase mRNA.

In order to determine the protein coding sequence of the chicken thymidine kinase gene, we compared the potential reading frames and the potential RNA splice junctions with the location and effect of our linker insertion mutations. Fig. 3 shows the protein-coding capacity of the HindIII fragment in all three potential reading frames. Previous work has suggested that the polypeptide coding region for the thymidine kinase gene

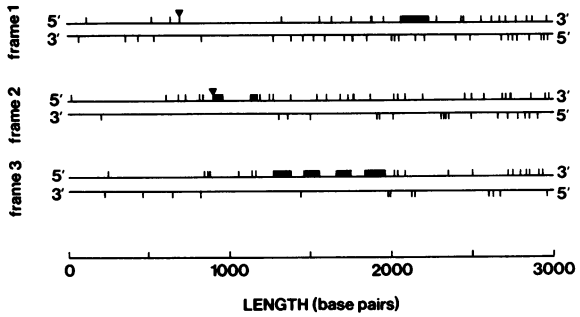


Fig. 3: The protein-coding potential of the cloned 3.0 HindIII fragment. The three potential reading frames of the fragment are shown. Frames 1, 2 and 3 begin at nucleotides 1, 2 and 3, respectively. Vertical lines show the positions of stop codons (UAG, UGA, UAA) in each reading frame. Some closely spaced stop codons may appear as a single vertical line due to the scale of this figure. Vertical lines with filled triangles denote the positions of AUG codons which are in the area of the thymidine kinase amino terminus and which may serve as sites for initiation of translation. The seven protein coding regions predicted from the nucleotide sequence are shown as seven filled rectangles.

lies between the EcoRI cleavage site (nucleotide 701) and the BglII site (nucleotide 2231). Only one ATG codon (at base 875) was found between the EcoRI and KpnI cleavage sites. This codon obeys the rules described by Kozak for functional translation initiators (42); there are purines at positions -3 and +4 with respect to the ATG. Our linker insertion mutations are consistent with the ATG at base 875 being the translation initiator. Mutation XH40 at base 805 does not inactivate the thymidine kinase gene whereas mutation XH136 at base 917 abolishes gene activity. Deletion mutagenesis of this region and replacement of sequences in this region with the herpes simplex thymidine kinase transcription initiators are also consistant with translation beginning at base 875 (33; Kwoh, unpublished observation). The nearest ATG to the EcoRI-KpnI fragment is located at base 697. Although this codon also follows Kozak's rules, and does start the major open reading frame in the region, this codon is unlikely to be used since the mutant XH40 does not inactivate gene activity *in vivo*.

The chicken thymidine kinase gene sequence downstream of the EcoRI cleavage site was also inspected for the presence of sequences that conformed to the AG-GT rule for donor and acceptor splice junctions (43). Six potential donor and acceptor junction sequences could be found (see Fig. 2); each donor can be paired with an acceptor lying directly downstream. Potential donors are found at nucleotides 941, 1201, 1393, 1569, 1764, and 1968; acceptor sequences can be found at nucleotides 1168, 1279, 1474, 1673, 1847 and 2088. Other examples of each of these putative donor and acceptor splice sequences have been shown to be bona fide

splicing sites *in vivo* (43). With the exception of the linker insertion mutations in the last proposed coding segment, the position of these splice junction sequences is consistent with the positions of linker mutants that abolish thymidine kinase expression *in vivo*: insertions that disrupt thymidine kinase activity are located between acceptor and donor splice sites, suggesting that a portion of the chicken thymidine kinase gene is encoded in these regions. Similarly, linker insertions that do not eliminate thymidine kinase activity *in vivo* are located between donor and acceptor splice sites in regions that should correspond to intron sequence to be removed by RNA splicing. The mutations XH60 and XA98 (bases 2152 and 2162 respectively) do not completely abolish gene activity and mutation XA2 (base 2207) has normal gene activity despite these three mutations occurring in the last protein coding region. This implies that the carboxyl terminus of the chicken thymidine kinase is dispensible to some degree.

The proposed number of intervening sequence based upon inspection of the nucleotide sequence is two more than that identified by the analysis of the linker insertion mutations. The additional introns encompasses the regions from base 1201 to 1279 and from base 1394 to 1474. These regions lack *Alu*I and *Rsa*I cleavage sites and would have escaped detection of even a complete linker insertion library constructed using these enzymes. While the base 1201 to 1279 intron is also devoid of *Hae*III sites, the base 1394 to 1474 intron contains a single *Hae*III site and could have been detected by a complete library. Two linker insertion mutations examined are interesting with respect to RNA splicing. The insertion site of mutant XR4 (base 942) changes a donor splice junction from AG/GTACC to AG/GTCCTCGAGGACC without affecting gene activity, or, presumably, splicing. Mutation XR75 and XR29 are located at base 1662, eleven bases upstream of an acceptor splice junction. Linker insertion at this site reduces gene activity, despite being in an intron presumably by interfering with splicing. This mutation suggests that acceptor splice junction have sequence upstream of the junction that are important.

Using the ATG initiation codon at nucleotide 875 and the potential splice junction sequences described above, a possible amino acid sequence for the chicken thymidine kinase gene can be derived from the nucleotide sequence (see Fig. 2). The putative polypeptide contains 224 amino acids and has a calculated molecular weight of 24,844. This predicted sequence was compared with the sequences of other known thymidine kinase genes from Herpes simplex virus-1 (376 amino acids; 44, 45) Herpes simplex virus-2 (376 amino acids; 46) and vaccinia virus (176 amino acids; 10, 11) As shown in Fig. 2, vaccinia virus thymidine kinase shares 66% homology with the predicted chicken thymidine kinase polypeptide, while the Herpes virus thymidine kinase shares no obvious homology (data not shown). To maximize the homology between chicken and vaccinia thymidine kinase genes, the sequence of the vaccinia thymdine kinase has to be compared starting at amino acid 17 of chicken thymidine kinase.

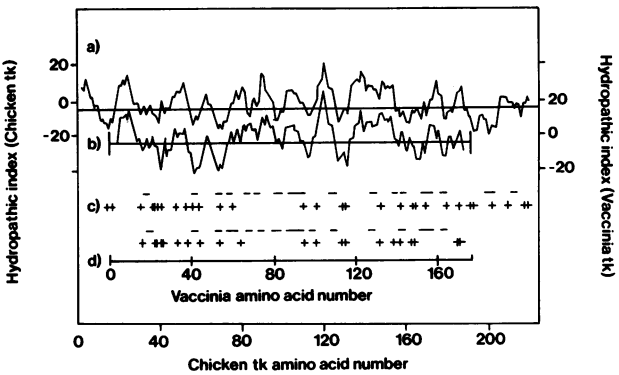


Fig. 4: Computer analysis of the predicted amino acid sequences of chicken and vaccinia virus thymidine kinase genes. a) and b) show the computer predicted hydrophobic regions of the chicken (a) and vaccinia virus (b) thymidine kinase polypeptides. The hydropathic index was calculated using values assigned by Kyte and Doolittle (32) by summing the hydropathic values for each set of seven continguous amino acids. The values have been plotted according to position of the fourth of the seven amino acids considered in each calculation. The baseline represents the average hydropathic value for all twenty-one amino acids. The plot for vaccinia virus thymidine kinase has been offset from that of chicken in order to maximize correspondence of the peaks; the position of the first amino acid for vaccinia corresponds to amino acid 17 of chicken thymidine kinase (see Fig. 2). c) and d) show the positions of glutamic and aspartic acid residues (-) or lysine and arginine residues (+) in chicken (c) and vaccinia virus (d) thymidine kinase genes. The positions of these amino acids in vaccinia virus thymidine kinase are offset from those of chicken as described above in order to maximize homology.

In order to further study this homology, computer programs for predicting hydrophobic regions from primary amino acid sequence data were used on the predicted chicken and vaccinia virus tk polypeptides. Fig. 4 shows the hydrophobic regions predicted according to the method described by Kyte and Doolittle (32). The hydropathic index measures the probability that a given amino acid will be found in a hydrophobic region of the polypeptide. Both the chicken and vaccinia tk proteins correspond well in this plot. Also shown are the positions of glutamic or aspartic acid (-) or arginine and lysine (+) amino acids along the predicted polypeptide sequences of the chicken and vaccinia virus genes. Many of the positions of charged amino acids correspond precisely in the two polypeptides. The strong homology of the chicken and vaccinia polypeptides supports the proposed organization of the chicken gene. The same computer programs were also used to study the herpes virus-1 and -2 thymidine kinase polypeptides. No correspondence between the herpes and the vaccinia or chicken thymidine kinase proteins could be identified either by hydrophobicity prediction or by position of charged amino acids (data not shown).

Taken together, these results suggest that the vaccinia and chicken

thymidine kinase genes share a common precursor sequence that differs significantly from the herpes virus thymidine kinase genes. The vaccinia and chicken thymidine kinase polypeptides are much smaller (20 and 34 kilodaltons) than the herpes virus thymidine kinase (41 kilodaltons). That the chicken thymidine kinase protein is different from the herpes simplex protein is not surprising in that the two enzyme have markedly different properties. The herpes simplex enzyme has a broad substrate range and is actually a deoxypyrimidine kinase (47, 48). The enzyme is able to phosphorylate deoxycytidine, thymidine, dTMP and a wide range of thymidine analogues (47). Also, the herpes enzyme can use all four ribonucleotide triphosphates as the phosphate donor (48). In contrast, cellular cytoplasmic thymidine kinases can only phosphorylate thymidine and some halogenated analogues. Currently, there is insufficient evidence to allow comparison of the enzymatic activities of the vaccinia and chicken enzymes.

Two other cellular cytoplasmic thymidine kinase genes have been isolated (13-15). However, the nucleotide sequences of the hamster and human thymidine kinase genes or mRNA have not yet been reported. Both genes are much larger than the chicken thymidine kinase gene although the mRNA encoded by the two genes are approximately the same size as that of chicken. Lewis et al. found the hamster thymidine kinase gene has a region of strong homology to the chicken gene as determined by hybridization (15). Merrill et al. report that 18 of the last 19 amino acids of the human thymidine kinase gene are identical to those of the chicken enzyme (33). These results suggest that the primary difference between the chicken, hamster and human thymidine kinase genes may be in the size of the introns and these three genes may encode similar polypeptides.

The close similarity of the chicken and vaccinia thymidine kinase polypeptides and the lack of intervening sequences within the vaccinia thymidine kinase gene (10, 11) raise questions as to the origin of the viral gene. Could the virus have acquired its thymidine kinase gene by forming a DNA copy of a cellular mRNA? Proposals have been made for a mRNA intermediate in the formation of chromosomal pseudogenes (49, 50). Similar mechanisms may have been involved in the formation of the vaccinia thymidine kinase gene. Although vaccinia virus can be propagated on chicken embryos, chicken cells are not the natural host for vaccinia virus; it will be interesting to see if mammalian thymidine kinase genes have a stronger homology to the vaccinia enzyme than that reported here. Further analysis of the thymidine kinase genes from other poxviruses will also be needed before the relationship of the vaccinia gene and cellular genes can be resolved.

ACKNOWLEDGEMENTS

We wish to thank Lynn Smith for technical assistance and Jayne Dollar for preparation of the manuscript. We also thank Dr. John Feiffer for use of his computer programs for calculating hydrophobicity. This work was

aided by grants from National Cancer Institute of Canada (T.J.K.) and by grants IN-66U from the American Cancer Society and CA-1314812 from the U.S. National Cancer Institute. J.A.E. is a special fellow of the Leukemia Society of America.

REFERENCES

1. Wigler, M., Sweet, R., Sim, G.K., Wold, B., Pellicer, A., Lacy, E., Maniatis, T., Silverstein, S., and Axel, R. (1979) Cell 16, 777-785.
2. Wigler, M., Silverstein, S., Lee, L.-S., Pellicer, A., Cheng, Y.-C. and Axel, R. (1977) Cell 11, 223-232.
3. Maitland, N.J. and McDougall, J.K. (1977) Cell 11, 233-241.
4. Littlefield, J. (1964) Science 145, 709-710.
5. McKnight, S.L. and Gavis, E.R. (1980) Nucleic Acid Res. 8, 5931-5948.
6. Colbere-Garapin, F., Chousterman, S., Horodniceann, F., Kourilsky, P. and Garapin, A. (1979) Proc. Natl. Acad. Sci. USA 76, 3754-3759.
7. Enquist, L.W., Van de Woude, G.F., Wagner, M., Smiley, J. and Summers, W.C. (1980) Gene 7, 335-342.
8. Wilkie, N.M., Clements, J.B., Boll, W., Mantei, N., Lonsdale, D. and Weissman, C. (1979) Nucleic Acid Res. 7, 859-877.
9. McDougall, J.K., Masse, T.H. and Galloway, D.A. (1980) J. Virol. 33, 1221-1224.
10. Weir, J.P. and Moss, B. (1983). J. Virol. 46: 530-537.
11. Hruby, D.D., Maki, R.A., Miller, D.B. and Ball, L.A. (1983) Proc. Natl. Acad. Sci. USA 80, 3411-3415.
12. Perucho, M., Hanahan, D., Lipsich, L., and Wigler, M. (1980) Nature 285, 207-210.
13. Bradshaw, H.D., Jr. (1983) Proc. Natl. Acad. Sci. USA 80, 5588-5591.
14. Lin, P.F., Zhao, S.Y. and Ruddle, F.H. (1983) Proc. Natl. Acad. Sci. USA 80, 6528-6532.
15. Lewis, J.A., Shimizu, K., and Zipser, D. (1983) Mol. Cell. Biol. 3, 1815-1823.
16. McKnight, S.L. (1982) Cell 31, 355-365.
17. Zipser, D., Lipsich, L. and Kwoh, J. (1981) Proc. Natl. Acad. Sci. USA 78, 6276-6280.
18. Kwoh, T.J., Zipser, D. and Wigler, M. (1983) J. Mol. Appl. Genetics. 2, 191-200.
19. Stubblefield, E. and Murphree, S. (1967) Expl. Cell Res. 48, 652-656.
20. Klevecz, R.R. (1969) Science 166, 1536-1538.
21. Adams, R.L.P. (1969) Expl. Cell Res. 56, 49-54.
22. Brent, R.P., Butler, J.A.V. and Crathorn, A.R. (1965) Nature 207, 176-177.
23. Boyer, H.W. and Roulland-Dussoix, D. (1969) J. Mol. Biol. 41, 459-472.
24. Guerry, P., LeBlanc, D.J., Falkow, W.S. (1973) J. Bacteriol. 116, 1064-1066.
25. Birnbaum, H.C. and Doly, J.C. (1979) Nucleic Acids Res. 7, 1513-1523.
26. Vogelstein, B. and Gillespie, D. (1979) Proc. Natl. Acad. Sci. USA 76, 615-619.
27. Maxam, A.M. and Gilbert, W. (1980) in Methods in Enzymology, Grossman, L. and Moldave, K., eds., Vol 65, pp. 499-560, Academic Press, New York.
28. Staden, R. (1977) Nucleic Acids Res. 4, 4037-4051.
29. Staden, R. (1978) Nucleic Acids Res. 5, 1010-1015.
30. Gingeras, T.R., Milazzo, J.P., Sciaky, D. and Roberts, R.J. (1979) Nucleic Acids Res. 7, 529-545.

31. Blumenthal, R.M., Rice, P.I. and Roberts, R.J. (1982) Nucleic Acids Res. 10, 91-102.
32. Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105-132.
33. Merrill, G.F., Hauschka, S.D. and McKnight, S.L. (1984) Cell, in press.
34. Breathnach, R. and Chambon, P. (1981) Ann. Rev. Biochem. 50, 349-383.
35. Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211-214.
36. Ahmed, C.M.I., Chanda, R.S., Stow, N.D. and Zain, B.S. (1982) Gene 20, 339-346.
37. Jung, A., Sippel, A.E., Grez, M., and Schutz, G. (1980) Proc. Natl. Acad Sci. USA 77, 5759-5763.
38. Hobart, P., Crawford, R., Shen, L., Pictet, R. and Rutter, W.J. (1980) Nature 288, 137-141.
39. Hagenbuchle, O., Bovey, R. and Young, R.A. (1980) Cell 21, 179-187.
40. Shani, M., Nudel, U., Zevin-Sonkin, D., Zakut, R., Givol, D., Katcoff, D., Carmon, Y., Reiter, J., Frischauf, A.M. and Yaffe, D. (1981) Nucleic Acids Res. 9, 589-589.
41. Goeddel, D.V., Leung, D.W., Dull, T.J., Gross, M., Lawn, R.M., McCandliss M., Seeburg, P.H., Ullrich, A., Yelverton, E. and Gray, P.W. (1981) Nature 290, 20-26.
42. Kozak, M. (1981) Nucleic Acids Res. 9, 5233-5251.
43. Mount, S.M. (1982) Nucleic Acids Res. 10, 459-472.
44. McKnight, S.L. (1980) Nucleic Acids Res. 8, 5949-5964.
45. Wagner, M.J., Sharp, J.A. and Summers, W.C. (1981) Proc. Natl. Acad. Sci. USA 78, 1441-1445.
46. Swain, M.A. and Galloway, D.A. (1983) J. Virol. 46, 1045-1050.
47. Chen, M.S. and Prusoff, W. H. (1978) J. Biol. Chem. 253, 1325-1327.
48. Cheng, Y.-C. and Ostrander, M. (1976) J. Biol. Chem. 251, 2605-2610.
49. Lemischka, I. and Sharp, P.A. (1982) Nature 300, 330-335.
50. Hollis, G.F., Hieter, P.A., McBride, D.W., Swain, D. and Leder, P. (1982) Nature 296, 321-325.