MOLECULAR AND CELLULAR BIOLOGY, Nov. 1984, p. 2316–2320
0270-7306/84/112316-05$02.00/0

Vol. 4, No. 11

Human Thymidine Kinase Gene: Molecular Cloning and Nucleotide Sequence of a cDNA Expressible in Mammalian Cells

HARVEY D. BRADSHAW, JR., AND PRESCOTT L. DEININGER*

Department of Biochemistry, Louisiana State University Medical Center, New Orleans, Louisiana 70112

Received 2 May 1984/Accepted 17 August 1984

A cDNA containing the entire coding region of the human thymidine kinase gene has been molecularly cloned. The cDNA is under the control of a simian virus 40 promoter and is expressible in mammalian cells. The complete nucleotide sequence of the human thymidine kinase cDNA has been determined. The cDNA is 1,421 base pairs in length and has a large open reading frame of 702 base pairs capable of specifying a protein with a molecular weight of 25,504. Genomic Southern blotting experiments show that sequences homologous to the human thymidine kinase cDNA are conserved among many vertebrates, including prosimians (lemur), tree shrews, rats, mice, and chickens. Direct comparison of the nucleotide sequences of the human thymidine kinase cDNA and the chicken thymidine kinase gene reveals ca. 70% overall homology. This homology is extended further at the amino acid sequence level, with greater than 74% amino acid residues matched between the human and chicken thymidine kinase proteins.

Thymidine kinase (ATP:thymidine 5′-phosphotransferase, EC 2.7.1.21) catalyzes the ATP-dependent phosphorylation of thymidine to thymidine 5′-monophosphate, which is subsequently converted to thymidine 5′-triphosphate and utilized for DNA synthesis. Thymidine kinase activity is intimately associated with DNA replication in many systems. In actively dividing cells, thymidine kinase activity rises dramatically with the onset of the S phase of the cell cycle, then declines rapidly in the G2 phase (1). In terminally differentiating cells, such as cardiac muscle cells, the cessation of DNA synthesis and withdrawal from the cell cycle are accompanied by the loss of thymidine kinase activity (5).

The fine structure and regulation of the herpes simplex virus type 1 thymidine kinase gene have been elucidated in considerable detail (19), largely due to the availability of recombinant DNA clones (9), complete nucleotide sequence information (29), and specific antisera raised against the protein (4). Unfortunately, the viral thymidine kinase activity is not regulated in the S phase-specific manner characteristic of the cellular enzyme. Developments in gene transfer and rescue, coupled with the powerful selections both for and against cells expressing thymidine kinase, have allowed the molecular cloning of chicken (24), hamster (16), and human (3, 15, 18) cellular thymidine kinase genes. The cloned human thymidine kinase gene is regulated appropriately when introduced into thymidine kinase-deficient mouse L cells (Ltk^-) (3), demonstrating that the control signals specifying the S phase increase in thymidine kinase activity are present on the clone and are recognized by mouse cells. Recently, the chicken thymidine kinase gene has been shown to be properly regulated when introduced into thymidine kinase-deficient mouse myoblasts, which can be induced to terminally differentiate (21). Further progress in identifying the elements regulating cellular thymidine kinase activity has been hampered by the lack of a detailed transcriptional map for the eucaryotic thymidine kinase genes. The extremely low abundance of the thymidine kinase mRNA makes such mapping very difficult. We have circumvented the problem of mRNA rarity by isolating molecularly cloned cDNA for human thymidine kinase. Here, we report the first complete nucleotide sequence for a eucaryotic thymidine kinase cDNA and the amino acid sequence deduced from it. Significant nucleic acid sequence homology between the human thymidine kinase cDNA and the DNA of other vertebrates has been detected by Southern blotting, and extensive homology with the chicken thymidine kinase gene has been identified precisely.

* Corresponding author.

MATERIALS AND METHODS

Cloning the human thymidine kinase cDNA. We had previously identified a 1.6-kilobase (kb) *Xho*I-*Eco*RI fragment of the human thymidine kinase genomic clone λTK46, which hybridizes to a 1.5-kb mRNA presumed to be the thymidine kinase mRNA (3), based on the sizing experiments of Lin et al. (17). The 1.6-kb *Xho*I-*Eco*RI fragment was subcloned into the plasmid vector pUC8 (28) as described previously (3). This plasmid is designated pXR1. To prepare probe for screening the cDNA library, pXR1 was cleaved with *Eco*RI and *Bam*HI (all restriction enzymes were from New England Biolabs or Boehringer-Mannheim Biochemicals and were used as recommended by the vendor), and the 1.6-kb fragment was eluted from low-melting-point agarose (Bethesda Research Laboratories) by published procedures (30). The fragment was further purified by electrophoresis through a 4% polyacrylamide gel followed by electroelution into a dialysis bag (11). The fragment was nick translated (25) to a specific activity of 2×10^8 cpm/μg with [α-^{32}P]dATP (all radiolabeled compound were from New England Nuclear Corp. or ICN Biomedicals, Inc.). The Okayama-Berg cDNA library (23) fractions containing inserts of 1.0 to 2.0 kb were used to transform competent (6) *E. coli* SF8 to ampicillin resistance. The library (300,000 clones) was screened by the high density colony screening method of Hanahan and Meselson (12). Nitrocellulose filters (Schleicher and Schuell) were prehybridized in a buffer containing NaPIPES [sodium piperazine-N,N′-bis(2-ethanesulfonic acid)], 50 mM; NaCl, 1 M; bovine serum albumin, 0.1%; polyvinylpyrrolidone, 0.1%; Ficoll, 0.1%; yeast RNA, 100 μg/ml; EDTA, 1 mM; and formamide, 50% (vol/vol). The hybridization was carried

out in the same buffer with the addition of 2×10^6 cpm/ml of the ^{32}P-labeled 1.6-kb fragment from pXR1. The filters remained in the hybridization solution for 48 h at 37°C. The hybridized filters were washed twice for 30 min in 2× SSC–0.5% sodium dodecyl sulfate (SDS) (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) at room temperature and once for 60 min in 0.2× SSC–0.5% SDS at 65°C. The filters were air dried and then autoradiographed for 16 h with Kodak XRP X-ray film and a Du Pont Cronex Lightning-Plus intensifying screen. Hybridizing clones were picked, replated at low density, and rescreened.

DNA sequence analysis. DNA sequence was performed by the chain termination method of Sanger, Nicklen, and Coulson (26). The human thymidine kinase cDNA clone pTK11 (see below) was chosen for sequence analysis. Plasmid DNA was prepared from a 1,000-ml culture by the alkaline lysis method (2). Three shotgun libraries (for a recent review, see reference 8) were prepared in the bacteriophage M13mp8 (22). A 1.5-kb *Bam*HI fragment of pTK11 containing most of the cDNA insert was eluted from low-melting-temperature agarose, self-ligated, and either cleaved with *Alu*I or sonicated, repaired with T4 DNA polymerase, and size fractionated by the method of Deininger (7). The resulting blunt-ended fragments were cloned into *Sma*I-cleaved and phosphatase-treated M13mp8. Another library was made by sonicating, repairing, size-fractionating, and subcloning the entire plasmid pTK11. The resulting M13 subclones were screened by hybridization to the 1.5-kb *Bam*HI fragment of pTK11. Sequence data were compiled and analyzed with the IntelliGenetics BION computer system.

Southern blot analysis. For Southern blot (27) analysis of thymidine kinase homology among vertebrates, DNA from each organism (5 μg) was digested with *Hin*dIII, fractionated by electrophoresis through a vertical 1% agarose gel, and transferred to nitrocellulose. The 1.25-kb *Sma*I-*Bam*HI fragment of pTK11 was labeled by nick translation and used as probe. The blot was hybridized and washed as described for the cDNA library screen above, except that the final wash was carried out at 37°C.

RESULTS

Isolation of a functional human thymidine kinase cDNA. We have recovered a cDNA clone for human thymidine kinase from the Okayama-Berg expression library (23). The library was constructed in a vector containing the simian virus 40 early promoter, modified late splice junctions, and polyadenylation signal, permitting expression of inserted cDNAs in mammalian cells. This unique vector has proved instrumental in verifying the identity of the cDNA from human hypoxanthine-guanine phosphoribosyltransferase (13) and the human thymidine kinase cDNA (in our studies). In the case of thymidine kinase, we had previously isolated a genomic clone which is capable of transforming Ltk$^-$ cells to tk$^+$ (3). The region coding for thymidine kinase mRNA within the genomic clone could not easily be identified due to the very low abundance of thymidine kinase mRNA even in logarithmically growing cells. A 1.6-kb *Xho*I-*Eco*RI fragment of the genomic clone was found to hybridize to an mRNA of 1.5 kb, the expected size of the thymidine kinase mRNA (17). We reasoned that if this were the authentic thymidine kinase mRNA, an expressible cDNA would be capable of transforming Ltk$^-$ cells to tk$^+$. When we used the 1.6-kb *Xho*I-*Eco*RI fragment of λTK46 as a hybridization probe to screen the 1.0-to-2.0-kb-insert fractions of the Okayama-Berg library, 22 hybridizing clones were present in the 300,000 ampicillin-resistant colonies examined. Four of the cDNA clones were purified, and their plasmid DNA was extracted. DNA from two of the clones was capable of transforming Ltk$^-$ cells to tk$^+$ with an efficiency greater than 10^3 per ng of DNA. We did not perform these experiments with higher dilutions of the plasmid DNA, so we do not know the upper limit of this transfection frequency. Both of these clones had cDNA inserts of ca. 1.5 kb, whereas the two nonfunctional clones had inserts of only 1.25 kb. Based on the 1.5-kb size for the thymidine kinase mRNA seen on Northern blots (3), our conclusion was that the two functional clones contain essentially full-length cDNA for human thymidine kinase. One of these clones, pTK11, was chosen for further analysis.

Nucleotide sequence analysis of the human thymidine kinase cDNA. The sequencing strategy and partial restriction map of the cDNA insert in pTK11 are shown in Fig. 1. The sequence was determined on both strands for 95% of the length of the cDNA, including all but three nucleotides of the protein-coding region. The complete nucleotide sequence, the deduced amino acid sequence, and the composition of the human thymidine kinase cDNA are presented in Fig. 2. The insert is 1,421 nucleotides in length, exclusive of the polydeoxyadenylate tail at the 3′ end and the polydeoxyguanylate tail at the 5′ end. The first methionine (ATG) codon, presumed to be the translation initiation site, is located 58 base pairs (bp) from the 5′ end of the cDNA. Because the reading frame is entirely open before the ATG at position 58, we cannot absolutely rule out an alternative translation start site. However, because the cDNA is 1,421 bp long, the addition of the normal 100 to 200 adenylate residues at the 3′ end of the mRNA would account for the 1.5-kb mRNA seen on Northern blots. We believe that the

FIG. 1. Sequencing strategy and restriction endonuclease cleavage map for the human thymidine kinase cDNA insert in pTK11. *Alu*I restriction fragments and random shear products generated by sonication of the 1.5-kb *Bam*HI fragment of pTK11 were cloned into the *Sma*I site of M13mp8, and their nucleotide sequences were determined as described in the text. Most of the unique restriction endonuclease cleavage sites are shown. The boxed-in region encloses the first large open reading frame, which specifies thymidine kinase.

```
           1                                  30                                  60                                  90
GGG GGG GGG ACT TAC TGC GGG ACG GCC TTG GAG AGT ACT CGG GTT CGT GAA CTT CCC GGA GGC GCA ATG AGC TGC ATT AAC CTG CCC ACT GTG CTG CCC GGC TCC CCC
                                                                                    Met Ser Cys Ile Asn Leu Pro Thr Val Leu Pro Gly Ser Pro

                      120                                 150                                 180
AGC AAG ACC CGG GGG CAG ATC CAG GTG ATT CTC GGG CCG ATG TTC TCA GGA AAA AGC ACA GAG TTG ATG AGA CGC GTC CGT CGC TTC CAG ATT GCT CAG TAC AAG TGC
Ser Lys Thr Arg Gly Gln Ile Gln Val Ile Leu Gly Pro Met Phe Ser Gly Lys Ser Thr Glu Leu Met Arg Arg Val Arg Arg Phe Gln Ile Ala Gln Tyr Lys Cys

210                   240                                 270                                 300
CTG GTG ATC AAG TAT GCC AAA GAC ACT CGC TAC AGC AGC AGC TTC TGC ACA CAT GAC CGG AAC ACC ATG GAG GCG CTG CCC GCC TGC CTG CTC CGA GAC GTG GCC CAG
Leu Val Ile Lys Tyr Ala Lys Asp Thr Arg Tyr Ser Ser Ser Phe Cys Thr His Asp Arg Asn Thr Met Glu Ala Leu Pro Ala Cys Leu Leu Arg Asp Val Ala Gln

                330                                 360                                 390                                 420
GAG GCC CTG GGC GTG GCT GTC ATA GGC ATC GAC GAG GGG CAG TTT TTC CCT GAC ATC ATG GAG TTC TGC GAG GCC ATG GCC AAC GCC GGG AAG ACC GTA ATT GTG GCT
Glu Ala Leu Gly Val Ala Val Ile Gly Ile Asp Glu Gly Gln Phe Phe Pro Asp Ile Met Glu Phe Cys Glu Ala Met Ala Asn Ala Gly Lys Thr Val Ile Val Ala

                          450                                 480                                 510
GCA CTG GAT GGG ACC TTC CAG AGG AAG CCA TTT GGG GCC ATC CTG AAC CTG GTG CCG CTG GCC GAG AGC GTG GTG AAG CTG ACG GCG GTG TGC ATG GAG TGC TTC CGG
Ala Leu Asp Gly Thr Phe Gln Arg Lys Pro Phe Gly Ala Ile Leu Asn Leu Val Pro Leu Ala Glu Ser Val Val Lys Leu Thr Ala Val Cys Met Glu Cys Phe Arg

        540                                 570                                 600                                 630
GAA GCC GCC TAT ACC AAG AGG CTC GGC ACA GAG AAG GAG GTC GAG GTG ATT GGG GGA GCA GAC AAG TAC CAC TCC GTG TGT CGG CTC TGC TAC TTC AAG AAG GCC TCA
Glu Ala Ala Tyr Thr Lys Arg Leu Gly Thr Glu Lys Glu Val Glu Val Ile Gly Gly Ala Asp Lys Tyr His Ser Val Cys Arg Leu Cys Tyr Phe Lys Lys Ala Ser

                      660                                 690                                 720
GGC CAG CCT GCC GGG CCG GAC AAC AAA GAG AAC TGC CCA GTG CCA GGA AAG CCA GGG GAA GCC GTG GCT GCC AGG AAG CTC TTT GCC CCA CAG CAG ATT CTG CAA TGC
Gly Gln Pro Ala Gly Pro Asp Asn Lys Glu Asn Cys Pro Val Pro Gly Lys Pro Gly Glu Ala Val Ala Ala Arg Lys Leu Phe Ala Pro Gln Gln Ile Leu Gln Cys

750                   780                                 810                                 840
AGC CCT GCC AAC TGA GGG ACC TGC AAG GGC CGC CCG CTC CCT TCC TGC CAC TGC CGC CTA CTG GAC GCT GCC CTG CAT GCT GCC CAG CCA CTC CAG GAG GAA GTC GGG
Ser Pro Ala Asn  .

                870                                 900                                 930                                 960
AGG CGT GGA GGG TGA CCA CAC CTT GGC CTT CTG GGA ACT CTC CTT TGT GTG GCT GCC CCA CCT GCC GCA TGC TCC CTC CTC TCC TAC CCA CTG GTC TGC TTA AAG CTT

                          990                                                                 1050
CCC TCT CAG CTG CTG GGA CGA TCG CCC AGG CTG GAG CTG GCC CCG CTT GGT GGC CTG GGA TCT GGC ACA CTC CCT CTC CTT GGG GTG AGG GAC AGA GCC CCA CGC TGT

      1080                                1110                                1140                                1170
TGA CAT CAG CCT GCT TCT TCC CCT CTG CGG CTT TCA CTG CTG AGT TTC TGT TCT CCC TGG GAA GCC TGT GCC AGC ACC TTT GAG CCT TGG CCC ACA CTG AGG CTT AGG

                    1200                                1230                                1260
CCT CTC TGC CTG GGA TGG GCT CCC ACC CTC CCC TGA GGA TGG CCT GGA TTC ACG CCC TCT TGT TTC CTT TTG GGC TCA AAG CCC TTC CTA CCT CTG GTG ATG GTT TCC

                                1320                                1350                                1380
ACA GGA ACA ACA GCA TCT TTC ACC AAG ATG GGT GGC ACC AAC CTT GCT GGG ACT TGG ATC CCA GGG GCT TAT CTC TTC AAG TGT GGA GAG GGC AGG GTC CAC GCC TCT

              1410                                1440
GCT GTA GCT TAT GAA ATT AAC TAA TTA AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA
```

FIG. 2. Nucleotide sequence and deduced amino acid sequence for the human thymidine kinase cDNA insert in the Okayama-Berg expression vector pcD. All but the first three nucleotides of the protein coding sequence were determined on both strands.

cDNA must be very nearly full length and that our assignment of the initial methionine is, therefore, correct. This is further confirmed by comparison with other thymidine kinase genes, as described below. The reading frame remains open for 702 bp, closing with an opal (TGA) terminator. This reading frame specifies a protein with a molecular weight of 25,504, in contrast to the molecular weight of 44,000 estimated for purified human placental thymidine kinase by SDS-polyacrylamide gel electrophoresis (10). The reason for this discrepancy is unknown, but the nucleotide sequence data given here probably represent the most accurate measurement and are in close agreement with the size of related thymidine kinase genes (14, 20). The 3′ nontranslated region of the cDNA is 662 nucleotides long with a putative polyadenylation signal (AATTAAA) just upstream from the polyadenylate addition site. Curiously, there is a large open reading frame in addition to the one we have assigned to thymidine kinase. This second frame is open from position 369 through position 1151, for a total length of 783 bp. Thus, this open reading frame is actually longer than the one encoding thymidine kinase, and it overlaps the thymidine kinase reading frame for 390 bp. The second open reading frame has no potential translation start site until position 744, making it an unlikely candidate for translation from the mRNA whose cDNA we have cloned. A similar open reading frame is not predicted for chicken gene and cDNA sequences (14, 20). We have noticed that a 1-bp insertion into the thymidine kinase reading frame that we have identified would result in a frameshift into the second large open reading frame and would specify a much larger protein than we have predicted for human thymidine kinase. To minimize the potential for errors, we sequenced both strands of the cDNA for all but the first three bases of the coding sequence and resequenced the most critical regions with independently derived M13 clones.

Comparison of vertebrate thymidine kinase genes. The 1.25-kb *Sma*I-*Bam*HI fragment of pTK11 that contained most of the protein-coding region of the human thymidine kinase cDNA was used as a hybridization probe to look for nucleotide sequence homology among other organisms. Hybridizing bands were found in *Hin*dIII digests of human, lemur, tree shrew, rat, mouse, and chicken DNAs but not alligator DNA (Fig. 3). The human, lemur, and tree shrew DNAs display a similar simple 3-banded pattern. The rodent DNAs examined, however, show a multitude of hybridizing bands. It is not yet clear what these bands represent, and further blotting and cloning experiments will be required to sort out the possibilities. Chicken DNA contains a single hybridizing 3.0-kb *Hin*dIII band, the size reported for the fragment containing the functional chicken thymidine kinase gene (24). Whereas the chicken thymidine kinase gene probe has been reported to show no cross-hybridization to rodent DNA (24) we found easily detectable homology between the human thymidine kinase cDNA and rodent and chicken genomic DNA. We sought to determine whether the 3.0-kb *Hin*dIII fragment in chicken DNA recognized by the human

thymidine kinase cDNA probe was, in fact, the chicken thymidine kinase gene. The complete nucleotide sequence of a chicken thymidine kinase cDNA and a portion of the gene is available (14, 20). A comparison of the human and chicken nucleotide sequences revealed ca. 70% homology over the full coding region and greater than 80% over the central 450 bases of the coding region. A complete comparison of the chicken and human thymidine kinase proteins is presented in Fig. 4. The homology exceeds 74% throughout the entire gene with the only major devergences at both the amino and carboxy termini. There is ca. 82% homology over the central 172 amino acids of the proteins. Deletions to the left of this terminator in chicken have been shown to destroy the activity of this gene (21), suggesting that this is the functional terminator. This comparison is further evidence that the human thymidine kinase cDNA sequence is correct.

We searched for nucleotide and amino acid sequence homologies between the human cDNA and the herpes simplex virus type 1 thymidine kinase gene with the Intelligentics SEQ program, but we have found no major homologies. This is in agreement with nucleic acid hybridization (unpublished data) and antibody neutralization (24) studies. Analysis of the large unused open reading frame has also shown no homologies to any of the known thymidine kinase genes at the level of either the nucleotide sequence or the potential translation products.

DISCUSSION

Several features of the human thymidine kinase gene have been examined with the expressible cDNA clone. The ability of the simian virus 40-promoted cDNA to transform tk^- cells to tk^+ verifies that the 1.5-kb mRNA identified by hybridization to the 1.6-kb *Eco*RI-*Xho*I fragment of our genomic clone does, in fact, encode thymidine kinase. Analysis of the nucleotide sequence of the human thymidine kinase cDNA reveals that the thymidine kinase mRNA has an apparently normal structure, with a 5′ untranslated region of at least 57 bp, a coding region of 702 bp, and a rather long 3′ untranslated region of 662 bp. A long 3′ untranslated region has been observed in the mRNA of another salvage pathway enzyme, human hypoxanthine-guanine phosphoribosyltransferase (13). One unusual property of the human thymidine kinase cDNA is the large, open reading frame overlapping the thymidine kinase coding region. The possibility that alternative mRNA splicing mechanisms may permit translation of this second open reading frame is being explored. The nucleotide sequence data predict a thymidine kinase protein with a maximum molecular weight of 25,504. This closely agrees with that predicted for the chicken and vaccinia virus thymidine kinase genes (14, 20). The catalytically active human thymidine kinase has an estimated molecular weight of 90,000 and migrates as a single band of M_r = 44,000 when subjected to SDS-polyacrylamide gel electrophoresis (10). We can only conclude from these conflicting data that either the migration of purified human placental thymidine kinase on an SDS-polyacrylamide gel is anomalous or that there are fundamental differences between the translation product of the thymidine kinase cDNA we have isolated and the enzyme purified from placenta. The existence of isozymic forms of thymidine kinase or covalent modification of the protein after translation are possible explanations for the difference in the calculated protein molecular weight.

The conservation of thymidine kinase nucleotide sequence homology among vertebrates is somewhat surprising, since the chicken thymidine kinase gene has been reported to show no cross hybridization to rodent DNA (24). Of the vertebrates examined, only the alligator DNA lacked detectable hybridizing bands when probed with the 1.25 kb *Sma*I-*Bam*HI fragment of pTK11. All of the DNA samples tested showed a light background smear in addition to any hybridizing bands, but this smear is not observed in Southern blot hybridizations of human DNA when the final wash is carried out at 65°C (unpublished data). It is likely that the high G+C content (60%) of the human thymidine kinase cDNA leads to nonspecific annealing at the lower stringency.

The 3.0-kb chicken DNA *Hin*dIII fragment which hybrid-

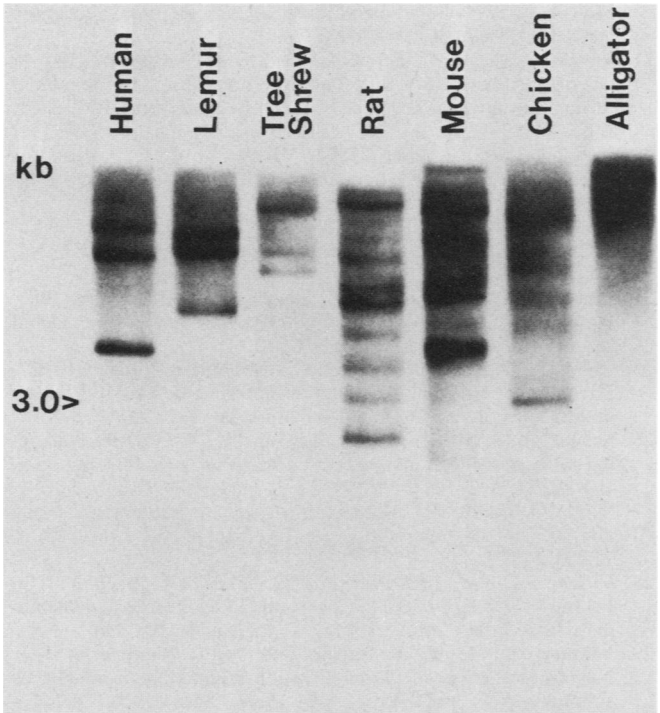

FIG. 3. Southern blot hybridization analysis of vertebrate DNA sequences homologous to the human thymidine kinase cDNA. DNA from various vertebrate tissues or cell lines was digested with *Hin*dIII, fractionated by agarose gel electrophoresis, transferred to nitrocellulose, and hybridized to the ^{32}P-labeled 1.25-kb *Sma*I-*Bam*HI fragment of pTK11 as described in the text. *Hin*dIII fragments of DNA served as size markers.

```
        10         20         30         40         50         60         70
 N LTV G H     GRP        F                    RL   R   LV         CTTGVS  R
MSCINLPTVL PGSPSKTRGQ IQVILGPMFS GKSTELMRRV RRFQIAQYKC LVIKYAKDTR YSSSFC*THDR

        80         90        100        110        120        130        140
     R  A  Q  Y       S              V     K   T                    A  S
NTMEALPACL LRDVAQEALG VAVIGIDEGQ FFPDIMEFCE AMANAGKTVI VAALDGTFQR KPFGAILNLV

       150        160        170        180        190        200        210
         N     G Y    S     A R                  A    Q RPQ *L S E    V MGV
PLAESVVKLT AVCMECFREA AYTKRLGTEK EVEVIGGADK YHSVCRLCYF KKASGQPAGP DNKENCPVPG

       220        230
 QLDMP S   I  S.----- ----
KPGEAVAARK LFAPQQILQC SPAN.
```

FIG. 4. Comparison of the predicted amino acid sequences of the human and chicken thymidine kinase proteins. The complete amino acid sequence of the human thymidine kinase protein is shown. The chicken thymidine kinase amino acid sequence is identical except where substitutions are indicated. Asterisks (*) represent spaces required for optimum alignment. Translation terminators are shown as periods (.). The numbering system refers to the human thymidine kinase.

izes to the human thymidine kinase cDNA probe is the same fragment which harbors the chicken thymidine kinase gene. The striking conservation of nucleotide and predicted amino acid sequence between these two vertebrates may help in identifying portions of the protein crucial to its function.

The molecular cloning of a human thymidine kinase cDNA opens a number of previously inaccessible areas for investigation. For the first time, a detailed transcription map of a eucaryotic thymidine kinase gene can be constructed. Though the human thymidine kinase gene is several times larger than the chicken thymidine kinase gene (3, 15, 18) the availability of the human thymidine kinase cDNA probe will make identification and characterization of genomic flanking regions and intron-exon boundaries relatively straightforward. Promoter switch experiments like those used to identify transcriptional control signals in the herpes simplex virus type 1 thymidine kinase gene (19) can be performed to define the elements that regulate the cell cycle-specific expression of thymidine kinase activity. These experiments will be greatly facilitated by the complete restriction map provided by the nucleotide sequence analysis. The goal of this line of inquiry is to identify cell cycle control mechanisms and their relationship to cellular differentiation and development.

ACKNOWLEDGMENTS

We thank Hiroto Okayama and Paul Berg for permission to use their cDNA expression library, and Doug Jolly and Ted Friedmann for providing us with aliquots of this library, and Abby Esty and Ted Friedmann for providing the Ltk$^-$ cell line. We also thank Gary Merrill for sharing the chicken thymidine kinase sequence data with us before its publication and for many helpful discussions. Expert technical assistance was rendered by Jenny Diniak.

During part of this work, H.D.B. was a predoctoral fellow of the National Science Foundation, and he is now a trainee of the National Institutes of Health (HL07098). This work was supported by Public Health Service grant GM29848 from the National Institutes of Health.

LITERATURE CITED

1. **Bello, L. J.** 1974. Regulation of thymidine kinase synthesis in human cells. Exp. Cell Res. **89**:263–274.
2. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
3. **Bradshaw, H. D., Jr.** 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.
4. **Cheng, Y.-C., K. C. Chadha, and R. G. Hughes, Jr.** 1977. Biochemical and immunological characterization of deoxythymidine kinase of thymidine kinaseless HeLa cells biochemically transformed by herpes simplex virus type 1. Infect. Immun. **16**:486–492.
5. **Claycomb, W. C.** 1979. DNA synthesis and DNA enzymes in terminally differentiating cardiac muscle cells. Exp. Cell Res. **118**:111–114.
6. **Cohen, S. N., A. C. Y. Chang, and L. Hsu.** 1973. Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA. Proc. Natl. Acad. Sci. U.S.A. **69**:2110–2114.
7. **Deininger, P. L.** 1983. Random cloning of sheared DNA: applications to shotgun DNA sequence analysis. Anal. Biochem. **129**:216–223.
8. **Deininger, P. L.** 1983. Approaches to rapid DNA sequence analysis. Anal. Biochem. **135**:247–263.
9. **Enquist, L. W., G. F. Van de Woude, M. Wagner, J. R. Smiley, and W. C. Summers.** 1979. Construction and characterization of a recombinant plasmid encoding the gene for the thymidine kinase of herpes simplex type 1 virus. Gene **7**:335–342.
10. **Gan, T. E., J. L. Brumley, and M. B. Van der Weyden.** 1983. Human thymidine kinase: purification and properties of the cytosolic enzyme of placenta. J. Biol. Chem. **258**:7000–7004.
11. **Girvitz, S. C., S. Bachetti, A. J. Rainbow, and F. L. Graham.** 1980. A rapid and efficient procedure for the purification of DNA from agarose gels. Anal. Biochem. **106**:492–496.
12. **Hanahan, D., and M. Meselson.** 1980. A protocol for high density screening of plasmids in X1776. Gene **10**:63–67.
13. **Jolly, D. J., H. Okayama, P. Berg, A. C. Esty, D. Filpula, P. Bohlen, G. G. Johnson, J.-E. Shively, T. Hunkapillar, and T. Friedmann.** 1983. Isolation and characterization of a full-length expressible cDNA for human hypoxanthine phosphoribosyltransferase. Proc. Natl. Acad. Sci. U.S.A. **80**:477–481.
14. **Kwoh, T. J., and J. A. Engels.** 1984. The nucleotide sequence of the chicken thymidine kinase gene and the relationship of its predicted polypeptide to that of the vaccinia virus thymidine kinase. Nucleic Acids Res. **12**:3959–3971.
15. **Lau, Y.-F., and Y.-W. Kan.** 1984. Direct isolation of the functional human thymidine kinase gene with a cosmid shuttle vector. Proc. Natl. Acad. Sci. U.S.A. **81**:414–418.
16. **Lewis, J. A., K. Shimuzu, and D. Zipser.** 1983. Isolation and preliminary characterization of the Chinese hamster thymidine kinase gene. Mol. Cell. Biol. **3**:1815–1823.
17. **Lin, P.-F., M. Yamaizumi, P. D. Murphy, A. Egg, and F. Ruddle.** 1982. Partial purification and characterization of the mRNA for human thymidine kinase and hypoxanthine/guanine phosphoribosyltransferase. Proc. Natl. Acad. Sci. U.S.A. **79**:4290–4294.
18. **Lin, P.-F., S.-Y. Zhao, and F. H. Ruddle.** 1983. Genomic cloning and preliminary characterization of the human thymidine kinase gene. Proc. Natl. Acad. Sci. U.S.A. **80**:6528–6532.
19. **McKnight, S. L.** 1982. Functional relationships between transcriptional control signals of the thymidine kinase gene of herpes simplex virus. Cell **31**:355–365.
20. **Merrill, G. F., R. E. Harland, M. Groudine, and S. L. McKnight.** 1984. Genetic and Physical analysis of the chicken *tk* gene. Mol. Cell. Biol. **4**:1769–1776.
21. **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.
22. **Messing, J., and J. Vieira.** 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene **19**:269–276.
23. **Okayama, H., and P. Berg.** 1983. A cDNA cloning vector that permits expression of cDNA inserts in mammalian cells. Mol. Cell. Biol. **3**:280–289.
24. **Perucho, M., D. Hanahan, L. Lipsich, and M. Wigler.** 1980. Isolation of the chicken thymidine kinase gene by plasmid rescue. Nature (London) **285**:207–210.
25. **Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. J. Berg.** 1977. Labeling deoxyribonucleic acid to high specific activity *in vitro* with DNA polymerase I. J. Mol. Biol. **113**:237–251.
26. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**:5463–5467.
27. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
28. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:269–276.
29. **Wagner, M. J., J. A. Sharp, and W. C. Summers.** 1981. Nucleotide sequence of the thymidine kinase gene of herpes simplex virus type 1. Proc. Natl. Acad. Sci. U.S.A. **78**:1441–1445.
30. **Wieslander, L.** 1979. A simple method to recover intact high molecular weight RNA and DNA after electrophoretic separation in low gelling temperature agarose gels. Anal. Biochem. **98**:305–309.