JOURNAL OF VIROLOGY, Oct. 1986, p. 114–123 Vol. 60, No. 1
0022-538X/86/100114-10$02.00/0
Copyright © 1986, American Society for Microbiology

Organization of the Thymidylate Synthase Gene of Herpesvirus Saimiri

WALTER BODEMER,† HANS HELMUT NILLER, NIKOLAUS NITSCHE, BRIGITTE SCHOLZ, AND BERNHARD FLECKENSTEIN*

Institut für Klinische Virologie der Universität Erlangen-Nürnberg, D-8520 Erlangen, Federal Republic of Germany

Received 20 March 1986/Accepted 15 June 1986

Herpesvirus saimiri codes, unlike most other herpesviruses, for a thymidylate synthase (TS). The TS gene of herpesvirus saimiri is unusual in structure and regulation of expression. It is transcribed into a nonspliced mRNA of 2,190 nucleotides. The single open reading frame of the viral TS gene, instructing a polypeptide of 33.5 kilodaltons, has extensive sequence homology with the corresponding TS coding sequences of human cells and of various procaryotes; the putative polypeptide derived from the nucleotide sequence of the herpesvirus saimiri TS gene is 70% identical with the human enzyme. The untranslated regions of the herpesvirus saimiri TS gene do not share homology with the other characterized eucaryotic or bacterial TS genes. The 5′ untranslated sequence has 22 ATG triplets shortly followed by stop codons. The herpesvirus saimiri TS gene, which may be weakly transcribed during immediate early and early times of virus replication, is maximally expressed at the late phase. Various parameters suggest that the TS gene has been acquired in virus evolution by an ancestral herpesvirus from the cellular genome.

Herpesvirus saimiri, a virus of squirrel monkeys (*Saimiri sciureus*), induces malignant lymphoproliferative diseases in other New World primate species and in rabbits (reviewed in references 15 and 16). Lymphoid T-cell lines derived from tumors contain multiple copies of nonintegrated covalently closed circular viral DNA molecules (25, 40). The protein-coding DNA (L-DNA) of virus particles has 112 kilobase pairs (kb); however, large parts of L-DNA sequences can be missing in the viral genomes persisting in nonproducer lymphoid cell lines (10, 40). Only L-DNA sequences from the left and right termini (about 30 kb) appear always to be preserved in the nonintegrated viral genomes (34). We found a single polyadenylated viral transcript of ~2,500 nucleotides in cultured nonproducer lymphoma cells (27). This transcript was mapped within the *Kpn*I D fragment of herpesvirus saimiri virion L-DNA between genome coordinates 0.89 and 0.93. Small amounts of the same mRNA were also found in permissive owl monkey kidney (OMK) cell cultures when protein synthesis was blocked with cycloheximide before virus infection (6). However, the initial conclusion that this genomic region represents the immediate early gene of herpesvirus saimiri needed revision. The 2.5-kb transcription unit is unusual in its regulation. The initial presence of the 2.5-kb mRNA could have been due to inoculation conditions, and the RNA is most abundantly found late in virus replication (6). Transfection of fusion genes with the promoter onto cultured cells did not result in transient expression (R. Rüger and W. Bodemer, unpublished results). In addition, another herpesvirus saimiri gene was reported as immediate early by P. Smith and R. Honess (personal communication). It thus seemed doubtful that the 2.5-kb transcription unit would be the functional equivalent of the regulatory immediate early genes in other herpesviruses. Further experiments showed that a long open reading frame (882 base pairs [bp], encoding a ~33.5-kilodalton [kDa] polypeptide) of the herpesvirus saimiri 2.5-kb gene has extensive sequence homology with the thymidylate synthase (TS) genes from the bacteria *Escherichia coli*, *Lactobacillus casei*, and *Bacillus subtilis*, from bacteriophage T4, and from human cells (22). The amino acid sequence derived from the herpesvirus saimiri gene was 70% identical with human TS. Results of experiments with monospecific rabbit antisera indicated that a TS activity could be precipitated from infected cells with a late viral polypeptide of about 30 kDa, leaving little doubt that the herpesvirus gene encodes a functional viral TS (22). This study gives the nucleotide sequence of the entire viral TS gene, reports mapping of the mRNA, and structurally compares the herpesvirus saimiri TS L-DNA region with other known TS genes. The results suggest that the coding DNA of the herpesvirus saimiri TS gene, which appears expressed, unlike all other virus genes, has been acquired from the cellular genome, probably without flanking cellular regulatory sequences.

* Corresponding author.
† Present address: Immuno GmbH, A-1011 Vienna, Austria.

MATERIALS AND METHODS

Cloning procedures. *Hin*dIII subfragments of existing L-DNA clones (27, 35) were subcloned in the plasmid vector pUC-8 and isolated from *E. coli* HB101 by standard procedures (21).

RNA isolation. Total cellular RNA was isolated from uninfected OMK cells, OMK cells infected with herpesvirus saimiri in the presence of cycloheximide (100 μg/ml), or OMK cells in the late stages of herpesvirus saimiri infection. The RNA was prepared by the guanidinium-isothiocyanate procedure with slight modifications (18, 27).

DNA sequencing. Cloned DNA fragments were digested with restriction endonucleases *Bgl*II, *Dde*I, *Dra*I, *Hin*dIII, *Hin*fI, *Kpn*I, and *Xba*I (Boehringer GmbH, Mannheim, Federal Republic of Germany) and labeled at the 5′ and 3′ termini with T4 polynucleotide kinase (Boehringer) and terminal transferase (Bethesda Research Laboratories, Inc., Gaithersburg, Md.), respectively. Subsequent cleavage with a second restriction endonuclease generated fragments with one ^{32}P-labeled terminus that were sequenced by the protocol of Maxam and Gilbert (30). A few DNA fragments were sequenced by the dideoxy-chain termination method (33) with bacteriophage derivatives M13mp18/19 (41).

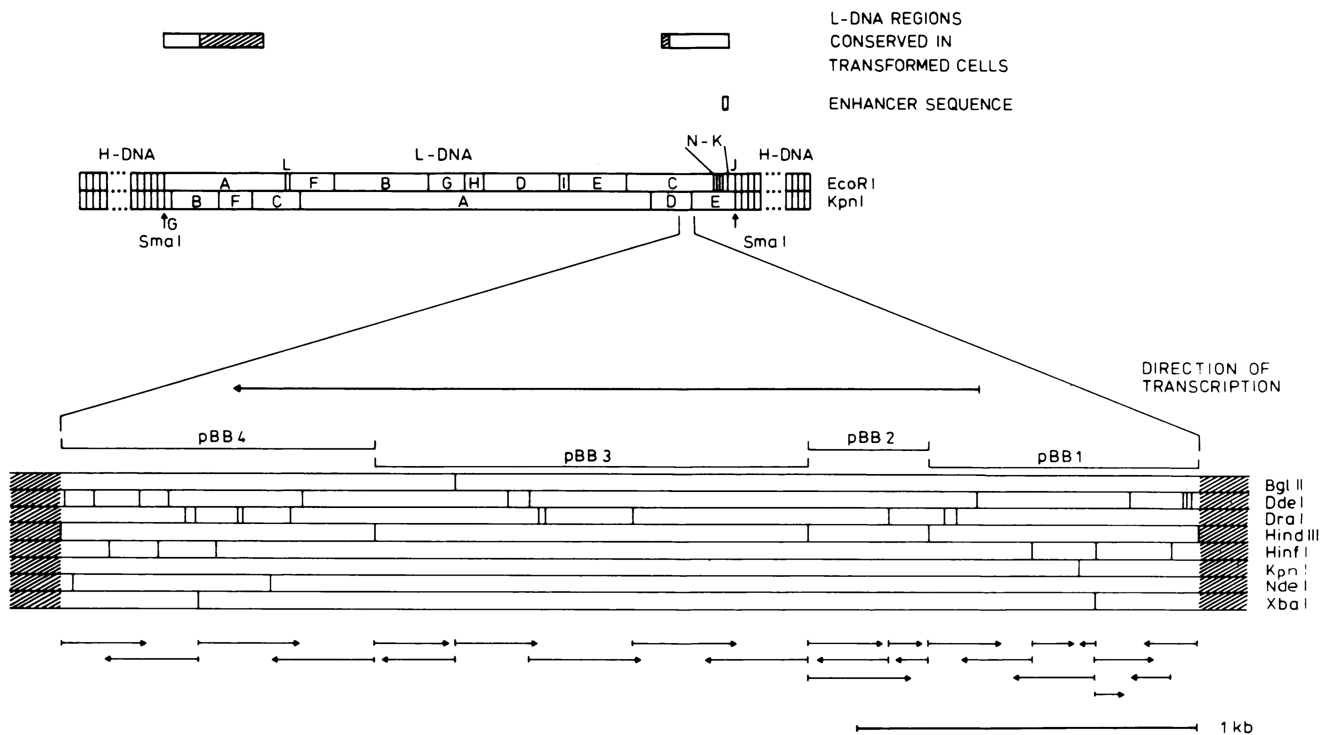

FIG. 1. Restriction endonuclease maps of the herpesvirus saimiri strain 11 L-DNA and of the 3.3-kb segment encompassing the entire TS gene. Functionally important genome regions are indicated: The L-DNA sequences preserved in all lymphoid tumor cell lines and in vitro-transformed cells (10, 16, 34, 40) and the enhancer sequence of herpesvirus saimiri (35). The strategy for sequencing the TS gene by Maxam-Gilbert degradation is symbolized by arrows at the bottom of the figure. The direction of transcription from the TS gene is indicated by an arrow.

Nuclease mapping of mRNA. DNA fragments were end labeled with T4 polynucleotide kinase or with the Klenow fragment of DNA polymerase I. After further cleavage with appropriate restriction endonucleases, the fragments were precipitated with ethanol. The DNA was denatured, frozen in liquid N_2, and dissolved with mRNA in a small volume (30 μl) of hybridization buffer (40 mM PIPES [piperazine-*N*,*N*′-bis(2-ethanesulfonic acid)], 1 mM EDTA, 0.4 M NaCl, and 80% purified formamide) (14). Hybridization of DNA and RNA was performed for 2 to 12 hours at 25 to 37°C, dependent on the G+C content of the DNA fragment. After hybridization, 300 μl of ice-cold S1 buffer (30 mM sodium acetate, 50 mM NaCl, 1 mM $ZnSO_4$, 20 μg of salmon sperm DNA per ml, 5% [vol/vol] glycerol [pH 6.7]) was added. Endonuclease S1 (Bethesda Research Laboratories) was used for 1 h at 25°C with a final activity of 1,000 U/ml. The endonuclease digestion was stopped by a phenol extraction. Labeled protected DNA fragments were precipitated with ethanol from the aqueous phase and analyzed on nondenaturing and denaturing 1.5% (wt/vol) agarose or 4% (wt/vol) polyacrylamide gels. The gels were exposed to X-ray film (X-Omat; Eastman Kodak Co., Rochester, N.Y.).

Primer extension analysis. Oligonucleotide primer synthesis was performed in a DNA synthesizer (model 380A; Applied Biosystems, Foster City, Calif.) by a solid-phase phosphoamidite method; the product was purified by polyacrylamide gel electrophoresis. Primer DNA was annealed with mRNA in hybridization buffer under the conditions of S1 mapping. The hybrids of DNA primer and RNA were precipitated in the presence of 0.3 M sodium acetate with 2.5 volumes of ethanol. After freeze-drying, the pellet was suspended in 50 μl buffer consisting of 24 mM Tris hydrochloride (pH 8.3), 16 mM $MgCl_2$, 8 mM dithiothreitol, 0.5 mM dGTP, 0.5 mM dTTP, 0.5 mM dCTP, and 40 μCi of [α-^{32}P]dATP. cDNA was synthesized at 42°C for 45 min with 5 U of reverse transcriptase (Boehringer). After the addition of 40 mM cold dATP, the reaction mixture was incubated for additional 45 min at 42°C. The radioactively labeled primer extension product was precipitated with ethanol, and its size was determined on denaturing 7.5% (wt/vol) polyacrylamide gels with known DNA sequence ladders.

Computer analysis. The complete nucleotide sequence was analyzed with a computer (type VAX 11/780; Digital Equipment), using programs of the University of Wisconsin Genetics Computer Group with data banks of nucleotide sequences (GenBank; European Molecular Biology Laboratory) and proteins (National Biological Research Fund) (11).

RESULTS

Determination of the primary structure. The entire TS transcription unit is contained within four adjacent *Hin*dIII fragments totaling up to about 3.3 kb (Fig. 1). Restriction maps were generated for each cloned *Hin*dIII fragment by single or multiple cleavages with the restriction endonucleases *Bgl*II, *Dde*I, *Dra*I, *Hin*dIII, *Hin*fI, *Kpn*I, and *Xba*I. The relative orientation of the *Hin*dIII fragments was determined by the physical mapping of the *Kpn*I D fragment and of a cloned 2.6-kb *Xba*I fragment containing the entire transcription unit (Fig. 1). Partial digestion experiments did exclude the presence of smaller *Hin*dIII fragments. The nucleotide sequencing strategy is indicated in Fig. 1. Individual restriction fragments were radioactively labeled from 5′ ends (for *Bgl*II, *Dde*I, *Hin*dIII, *Hin*fI, and *Xba*I) or from 3′

```
HinDIII
AAGCTTGTCTTTAGTATGATCTTAGCATGATCTTAGCATGTTCTTAGCATGTTGCTACACTGCATCAGTGAAATTTTGACTCATGTCAACTATGTTGTTG  100
CATGTAGTTGTATGTTAGCACATGAAGTTCACGTGTTTGAGGCATGAAGTGATCTATTTTAGATCACTTCATGCTTAAACATACGTTATAATCCTTGCTC  200
                                                                                            XbaI
TGAGTTTCTCGAACTGTGATTGGATAACTCATTTTTAAGCATTCTGATTGGATCAGAGCGTAAGTGATCTACTAGAGGCGAGCTTTTGATTCGTCTAGAG  300
                                             KpnI
CAGCAGCCAATAGAAATGCAGACAGTGATCTTATATAGTATAGTTATTCGGTACCGAATATTCGGTACTGAAAAATCTATTTTTTAATTTAATTTATTTT  400
ATATTAATTTACCTATACTTGTTATCGGTATAGAGGTATCGGTATATGTATTTCTATACCGATATATCGGTATAGAGACGAAATGACTCATCTGTTTAGT  500
ATTTTTAGAAAATGCAGCACTGTTTTATGTGGTGTGTTTATATAAAGCATGCTAGCTACTGACTGTGTTTTTAATTGAGCAAAACAGCTTTTGTGGCTGC  600
ATTTTTGAGTG[TATATAA]GGCAAGCAAGCTAGAGCTGCTTGGCTCACTTGCTCAGTGAAGCCCGAAGAAGCTGGTCATGATTTTGAGCTCTCCCAGCTTA  700
                                                                                       HinDIII
AAAATTTAAACCTGCTTAAGGTTGTGACAAAAAAATTAAAATTTAAAAAAATTTTAGAATGCTTGTGGATAAAAGCTGAAAAATACAAGCTTTATTGTGT  800
TGCAAAATTTATACTGCAATTTAAGCACAAAAAACACACTGAATAATTATTTGCTACAGTGAGCTATATTTCCAGCTAAAAATTACTTGCTTAAAATTAA  900
TGCTTTAAAATTGTGAATAATTTATTATGTAAAGAAAATATTTTGTGTGATTTTATTGAGCCATTGAAATAATTTTTTGTATTTAATTGCTGCGAAAACA  1000
ATTGTAGAATTTAATTAATTCATGGAAAAACAATTTGAAATATTTAATGGTTACAATCATTGTTGTGAAAACAATTGATATTTAATTCAGCTATGAAAAA  1100
                                         HinDIII
TATTTGAATATTACATGGTTGTAAGAATATTGTAGCATTTAAGCTTTCAGCCTATGGAGATGTTTAATGGCAATGTGAATATTTTAGCATATTGTATGGT  1200
TGTGAGAATATATTGTGGCATTCAAGTCAGCTATCAAAAAATTGATATTTAATTACAGCATGTATATTTTATGACTTTTAATTTAGCTGTGAAAATATAT  1300
CAGTGTAAAAATATTGACATTTAATTAAGCTATCAAAATATTTCATAGTTTATTTGTTCAGTGAAAATACATTGCATTTTTATTGTTGTAAGAAAATCTG  1400
TATCATTAATTGTTGTAAACATTGCATTTTATTAATTTAGTATTGTGTAAGACACATTTACAATGCATCTTGGATTGAAGTGGTTTGTTTCTTTTCTACT  1500
TACGTATTTATTTGGTCACTTCATATTCATACCTTGATTTATTTCTTTTCATATTTTTGAACTTTAATTCTTATATTTTTTCATATTTTTTATGTGCACA  1600
ATTTTACCGCTCTTTAGTTTATGGTTGCACAATTTTTACTGAACTATTAAATTTAAAAAAAATAAGTTTATAAAGCTGTATAATACATTTTTAATGTGCT  1700
CAAAAACGTAACTTTTGCACAATATATTAATTTTAATTTTGGTTTTTACTTTATTTGCAGATAAAAATGACAAAATTTAAGTTTTATGTTTTATTTGTAT  1800
TATGGACACTTTATACAGTATATTAACTAAAACTTGACAAATAATAAGATGTCAACACACACAGAAGAACAGCATGGAGAACACCAGTACCTCTCACAGG  1900
                                                    MetSerThrHisThrGluGluGlnHisGlyGluHisGlnTyrLeuSerGlnV
TACAGCACATTTTAAATTATGGGTCATTTAAAAATGACAGAACAGGGACTGGAACACTGAGTATTTTTGGAACACAGTCTAGGTTTAGTTTAGAAAATGA  2000
alGlnHisIleLeuAsnTyrGlySerPheLysAsnAspArgThrGlyThrGlyThrLeuSerIlePheGlyThrGlnSerArgPheSerLeuGluAsnGl
ATTTCCACTTTTAACTACTAAGAGAGTATTTTGGAGAGGTGTTGTTGAAGAACTGTTGTGGTTTATCAGAGGATCTACTGACAGCAAAGAGCTATCAGCA  2100
uPheProLeuLeuThrThrLysArgValPheTrpArgGlyValValGluGluLeuLeuTrpPheIleArgGlySerThrAspSerLysGluLeuSerAla
                               XbaI
GCTGGCGTGCACATTTGGGATGCGAATGGATCTAGATCATTTTTAGATAAACTTGGTTTTTATGACAGAGATGAAGGAGATCTTGGACCTGTGTATGGAT  2200
AlaGlyValHisIleTrpAspAlaAsnGlySerArgSerPheLeuAspLysLeuGlyPheTyrAspArgAspGluGlyAspLeuGlyProValTyrGlyP
TTCAGTGGAGGCATTTCGGAGCAGAATATAAAGGTGTTGGGCGCGATTATAAAGGAGAAGGAGTTGACCAATTAAAACAGTTGATTGATACTATAAAAAC  2300
heGlnTrpArgHisPheGlyAlaGluTyrLysGlyValGlyArgAspTyrLysGlyGluGlyValAspGlnLeuLysGlnLeuIleAspThrIleLysTh
AAACCCTACAGATAGAAGGATGTTAATGTGTGCTTGGAATGTTTCAGACATTCCTAAAATGGTGTTGCCGCCTTGTCATGTATTAAGTCAATTTTATGTC  2400
rAsnProThrAspArgArgMetLeuMetCysAlaTrpAsnValSerAspIleProLysMetValLeuProProCysHisValLeuSerGlnPheTyrVal
    HinDIII
TGTGATGGAAAGCTTTCCTGTCAACTGTATCAAAGATCAGCTGATATGGGGTTAGGAGTGCCATTTAACATTGCTAGCTATTCTCTTTTAACTTGCATGA  2500
CysAspGlyLysLeuSerCysGlnLeuTyrGlnArgSerAlaAspMetGlyLeuGlyValProPheAsnIleAlaSerTyrSerLeuLeuThrCysMetI
TTGCTCATGTCACTAATCTAGTGCTTGGAGAGTTTATCCATACTATAGGAGATGCTCACATCTATGTAGATCACATTGATGCTCTTAAAATGCAGTTGAC  2600
leAlaHisValThrAsnLeuValLeuGlyGluPheIleHisThrIleGlyAspAlaHisIleTyrValAspHisIleAspAlaLeuLysMetGlnLeuTh
GCGAACTCCTAGACCGTTTCCAACACTTAGATTTGCTAGAAATGTTTCATGCATTGATGACTTTAAAGCAGATGACATTATATTAGAAAACTATAACCCA  2700
rArgThrProArgProPheProThrLeuArgPheAlaArgAsnValSerCysIleAspAspPheLysAlaAspAspIleIleLeuGluAsnTyrAsnPro
                NdeI
CACCCTATAATTAAAATGCATATGGCTGTTTAAGAAAAATAACATGTAAAACGTATCATAAACTAAAACACTTACATTGTAACTACATTTTTCATAGCTG  2800
HisProIleIleLysMetHisMetAlaValEnd
TTTTAAAA[AATAAA]AGTTTTAAAACGTATTTACTTGTCTGTTGAATGATTTCTGGGGC[CATTG]GGACTTAAACCAAGTGAATCATTAAGTTTTAGGTAGT  2900
                                XbaI
ATCTTTAAGTTTTAGACTTACAATACCACTGTCTAGAATTTTAAAAGAAAGACAACAATCATATAAATTTTAAATATCTTTTATTTGAAGAGCATTTATG  3000
CCTTTTACAATTTCTTCTCTGAGCTCATCAGTACAATCAAGAGTAGACACAGATTCACATATCGTGTCTTCTGCGATAGAGATGTTTCTCTGTATTGCAA  3100
AACACTGAGTAAATGTCATTGCATAAACATCTTTGATGTAATCAATTGAAATTTTATAGTTTGGTAAATTGTCCATAGGCATTCTTAAGCAAGATTCGCC  3200
AATGTGCAACGTCTGGCACTTAAATGTCAGACAAATACTGAGGCAATCATCTTTTTCAAAAAAAATTGGAAATACGTTGTGTCCAGCTTCTTTTACTACA  3300
NdeI                          HinDIII
ACATATGCGTATAAGAGCTCTGGCTGAGTAAAGCTT  3336
```

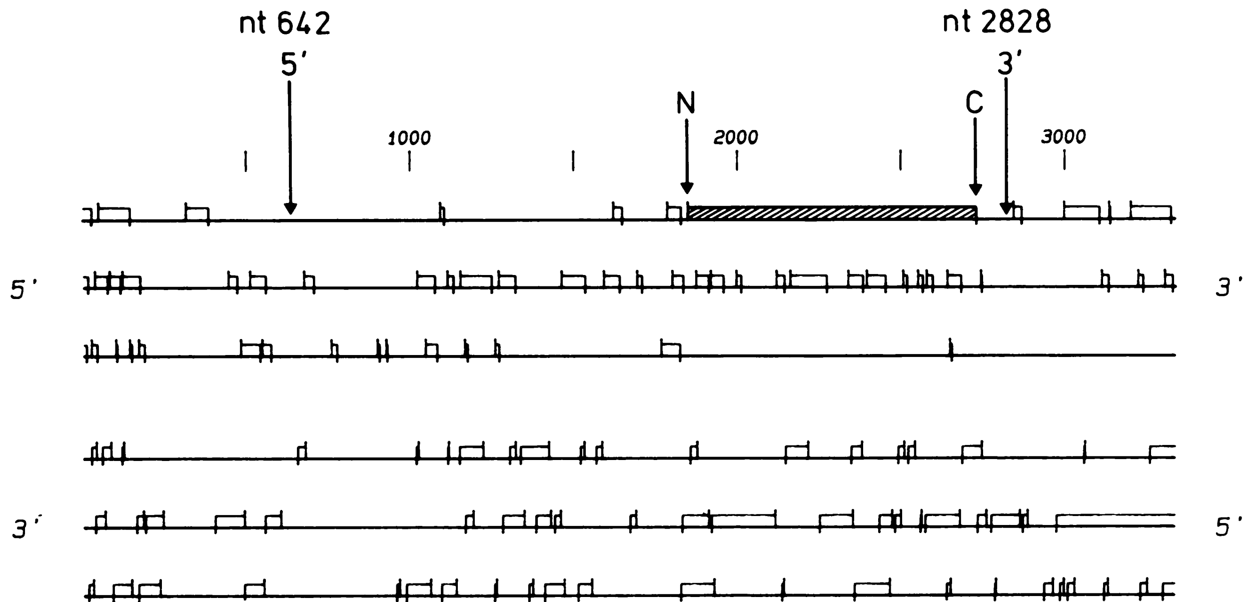

FIG. 3. Reading frame analysis of the four *Hin*dIII-DNA fragments composing the herpesvirus saimiri TS gene. The reading frame encoding TS is symbolized by a dashed box. Bars above the lines indicate ATG codons, bars crossing the lines are stop codons. Transcription initiation (5') and polyadenylation sites (3') are indicated by arrows.

ends (for *Dra*I) and sequenced by chemical degradation. When a single strand was used for sequencing, the reaction was performed independently and read identically at least twice. Two fragments that left ambiguities were inserted into the multiple cloning site of the bacteriophage derivative M13 mp 18/19 for sequencing by the dideoxynucleotide chain termination procedure. The entire sequence of 3,336 nucleotides is presented in Fig. 2. Reading frame analysis indicated a single open reading frame of 882 nucleotides. This DNA sequence can code for a polypeptide of 33,498 daltons (Fig. 3). The protein has striking structural homologies to procaryotic and human cellular TS, and in a previous study a functional TS has actually been demonstrated as it binds 5' fluoro-dUMP and 5', 10'-methylenetetrahydrofolic acid (22). No other translational reading frame for a polypeptide of ≥45 amino acids was found within the full length of the four sequenced *Hin*dIII fragments in the direction of transcription. This was consistent with the results of in vitro translation experiments with hybrid-selected RNA: when RNA from OMK cells late after virus infection was hybridized with the immobilized cloned viral DNA, eluted from solid phase, and translated in a rabbit reticulocyte lysate, a single virus-specific polypeptide with an apparent molecular mass of about 33 kDa was detected in denaturing polyacrylamide gels (Fig. 4). Thus, the single open reading frame encodes a 33.5-kDa protein with TS activity.

Transcript mapping. Transcription initiation and polyadenylation sites of the 2.5-kb RNA were localized by nuclease protection analyses. poly(A)$^+$ RNA extracted from OMK cells 40 h after infection was hybridized with the 5'-labeled *Hin*dIII fragment of 786 nucleotides; DNA-RNA hybrids were treated with nuclease S1 and then sized in nondenaturing agarose and polyacrylamide gels (Fig. 5A). A protected hybrid fragment of about 158 ± 5 nucleotides was detected, placing the start site for transcription at around nucleotide 642, if there were no splicing in the formation of late 2.5-kb mRNA. The putative cap site of this transcript was more precisely mapped by primer extension synthesis. A synthetic oligonucleotide of 30 residues, corresponding to the transcribed DNA strand between positions 689 and 660 (5'-AGCTCAAAATCATGACCAGCTTCTTCGGGC-3'), was synthesized by solid-phase-phosphoamidite chemistry. It was hybridized with the 2.5-kb RNA, and the 3' terminus of the DNA primer was elongated by reverse transcription in the presence of [α-^{32}P]dATP. The synthetic DNA primer was mostly extended by 18 nucleotides (Fig. 6). Thus, apparently, nucleotide 642 corresponds to the 5' terminus of viral mRNA and to the initiation of transcription. The autoradiogram in Fig. 6 shows, in part, a number of weak bands corresponding to a radioactive extension product of 36 to 44 nucleotides. Though this could be due to secondary transcription initiation sites or RNA degradation, it may also be caused by artificial premature termination of reverse transcription. Repeated experiments with immediate early and late RNA gave the same pattern consistently, suggesting that secondary structure of RNA may be responsible for premature termination. The oligonucleotide with 49 bases, seen in low molarities in lanes 1 and 2 of Fig. 6, is probably the extension artifact that can be observed with capped RNA (28). These results thus indicate that the 5' end of the 2.5-kb mRNA corresponds to nucleotide 642, whether isolated under a cycloheximide block or at late times of virus replication. It excluded splicing within the first 145 nucleotides of the primary transcript. The nucleotide motif TATATAA precedes the transcription initiation point by 30 bases; it represents a canonical TATA sequence, as it is

FIG. 2. Complete nucleotide sequence of four *Hin*dIII DNA fragments composing the herpesvirus saimiri TS gene. Signals for initiation and termination of transcription are boxed. The cap site and the polyadenylation site are indicated by arrows. The sequence of the oligonucleotide used for primer extension experiments is underlined. The amino acid sequence corresponds to the single long open reading frame in the direction of transcription.

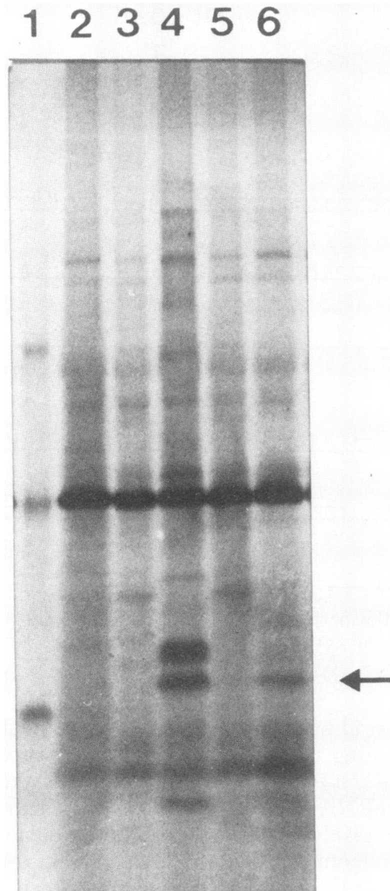

FIG. 4. Hybrid-selected in vitro translation of TS RNA. Late RNA was isolated from lytically infected OMK cells, hybridized with cloned viral DNA, and eluted from filters, and ^{35}S-labeled proteins were synthesized in vitro with rabbit reticulocyte lysates. Lane 1, Molecular weight markers (68,000, 44,000, 30,000); lane 2, translation of RNA from uninfected OMK cells; lanes 3 and 5, translation of RNA from transformed lymphoid tumor cell line 1670; lane 4, translation of late RNA from infected cells at 30 h postinfection; lane 6, late RNA selected with the cloned *Kpn*I D fragment of herpesvirus saimiri L-DNA and translated in vitro.

frequently found as a promoter element at this distance from transcription initiation points.

The 3′ terminus of the viral RNA was mapped first by nuclease protection experiments with the *Hin*dIII fragment of 921 bp. The recessed DNA strands were ^{32}P labeled by extension of 3′ ends with Klenow fragment of DNA polymerase I; they were hybridized with the late 2.5-kb RNA, treated with S1 nuclease, and the sizes of RNA-DNA hybrids were determined in nondenaturing agarose gels. This placed the 3′ end of the transcript between nucleotides 2825 and 2835 (about 400 nucleotides from the *Hin*dIII cleavage site) (Fig. 2 and 5D). The position of the polyadenylation site was more closely determined by S1 analysis with the 3′-labeled *Nde*I fragment between nucleotides 2720 and 3303 (Fig. 2) and electrophoresis in a denaturing 7.5% (wt/vol) polyacrylamide gel. The polyadenylation site was found 10 ± 4 nucleotides downstream of the signal sequence AATAAA. Considering the known consensus sequences of polyadenylation sites and their minimal spacing to the AATAAA motif (5, 31), we assumed that the polyadenylation site for the TS transcript of herpesvirus saimiri is nucleotide 2828 (Fig. 2). The polyadenylation site of the herpesvirus saimiri TS gene is in the immediate vicinity of the sequence motif CAYTG that is known to flank 3′ ends of eucaryotic transcription units (17). Several GT-rich motifs are located immediately downstream of the polyadenylation site, as frequently found in eucaryotic genes (5).

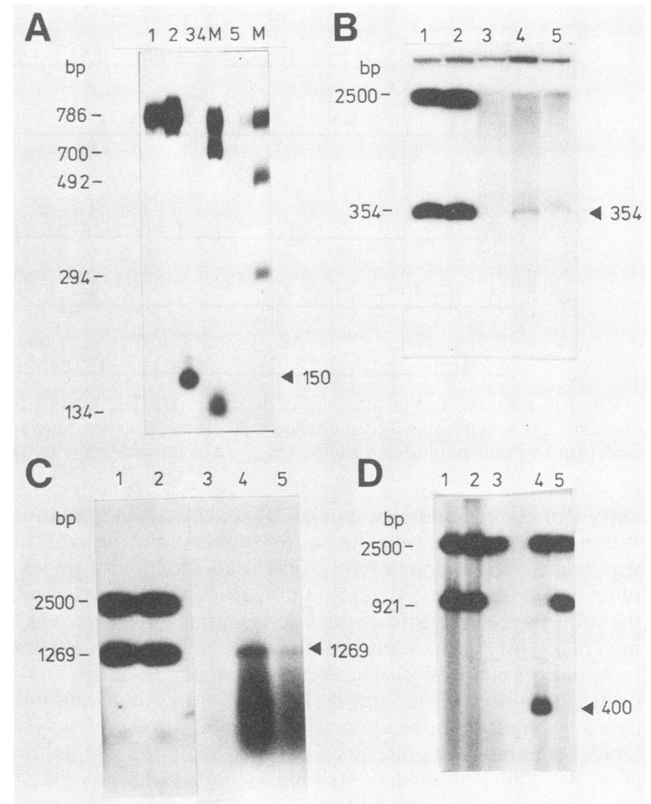

FIG. 5. Endonuclease S1 protection analyses of TS mRNA with the four *Hin*dIII DNA fragments covering the entire gene. (A) Protection analysis of the 786-bp fragment (nucleotides 1 through 786 in Fig. 2). Lane 1, 5′-labeled DNA fragment; lane 2, labeled DNA fragment after denaturation and renaturation; lane 3, labeled DNA fragment hybridized with RNA from infected cells and digested with S1 nuclease; lane 4, labeled DNA fragment denatured and digested with S1 nuclease; lane 5, labeled DNA denatured, incubated with yeast RNA under reannealing conditions, and digested with S1; lane M, size markers. (B) Protection analysis of the internal 354-bp fragment (nucleotides 787 through 1141 in Fig. 2). Lane 1, 3′-labeled DNA fragment; lane 2, labeled DNA fragment after denaturation and renaturation; lane 3, unprotected single strands digested with S1; lane 4, labeled DNA fragment hybridized with RNA from infected OMK cells; lane 5, labeled DNA fragment denatured, incubated with yeast RNA under reannealing conditions, and digested with S1. (C) Protection analysis of the internal 1,269-bp fragment (nucleotides 1142 through 2410 in Fig. 2). Lane 1, 3′-labeled DNA; lane 2, labeled DNA fragment after denaturation and renaturation; lane 3, unprotected single strands degraded by S1 nuclease; lane 4, labeled DNA fragment hybridized with RNA from infected OMK cells and digested with S1 nuclease; lane 5, control with yeast RNA. (D) Protection analysis of the 921-bp fragment (nucleotides 2411 through 3331 in Fig. 2) containing the polyadenylation site. Lane 1, 3′-labeled DNA fagment; lane 2, labeled DNA fragment after denaturation and renaturation; lane 3, unprotected single-stranded DNA degraded by S1; lane 4, labeled DNA hybridized with RNA from infected cells and incubated with S1 nuclease; lane 5, yeast control. In panels B through D, the fragments of about 2.5 kb are 3′-labeled DNA of cloning vector pUC8.

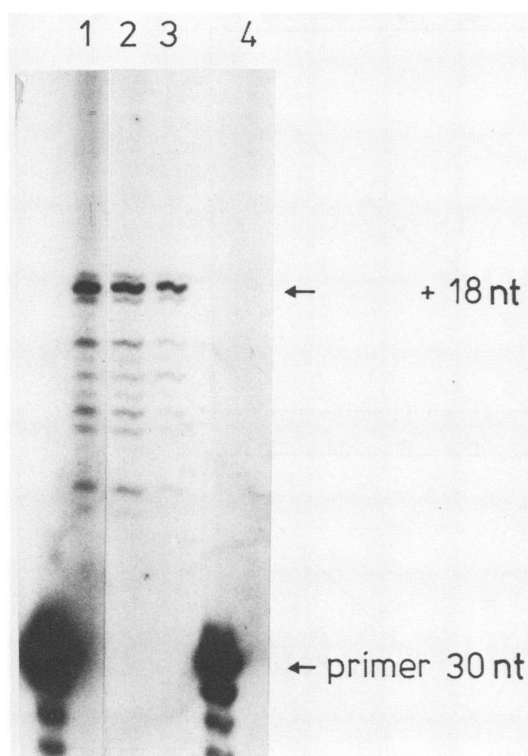

FIG. 6. Primer extension analysis to determine the 5′ end of the TS mRNA. RNA extracted from OMK cells under a cycloheximide block (lane 1) or late times after infection (lanes 2 and 3) was hybridized with a synthetic oligonucleotide corresponding to the coding strand from nucleotides 660 through 689 (Fig. 2). The unlabeled primer was extended by reverse transcriptase with α-^{32}P-labeled dATP. The large radioactive spots of 30 nucleotides (nt) are 5′-labeled primers (lane 4). The labeled nucleotides were sized by running them in parallel to a sequence ladder.

Nuclease S1 protection analysis of the 2.5-kb mRNA with the internal *Hin*dIII fragments of 354 and 1,269 bp under denaturing agarose gel conditions did not give evidence for transcript splicing (Fig. 5B and C). The protected bands appeared rather weak; this obvious technical problem was probably due to the very high A+T content (up to 80%) of those DNA fragments, resulting in a low RNA-DNA hybrid stability. The absence of RNA splicing was confirmed by digestion of RNA-DNA hybrids with exonuclease VII (data not shown). This altogether indicates that the TS transcription unit does not contain an intron; the mature 2.5-kb RNA is formed without splicing. It is consistent with the results of earlier Northern blotting experiments in which very limited space was left for RNA splicing (6).

Structural comparisons. Mapping of the transcription initiation site (nucleotide 642) indicated that the TS mRNA of herpesvirus saimiri has a 5′ untranslated sequence of 1,207 nucleotides. This sequence is exceptionally high in average A+T composition (76%), significantly higher than the protein-coding sequence (62%) and the 3′ untranslated region (69%) (Fig. 7). A striking feature of the 5′ untranslated region in the thymidylate synthase gene of herpesvirus saimiri is its extreme length. The 5′ untranslated sequence of the TS gene in herpesvirus saimiri contains 22 ATG triplets that are shortly followed by stop codons. This forms 19 very short open reading frames in front of the long TS frame (Fig. 3). The sequence conservation between the TS gene of herpesvirus saimiri and its procaryotic or eucaryotic counterparts was found to be confined to the corresponding long open reading frames. Figure 8 shows dot plot analyses of the amino acid sequences in TSs of herpesvirus saimiri with *E. coli* (4), *L. casei* (29), bacteriophage T4 (7), and human cells (37). These computer analyses essentially confirmed the sequence comparisons presented in the previous report (22). Evolutionary divergence appears to have introduced two short segments of divergence (amino acids 92 through 109 and 122 through 139 of herpesvirus saimiri TS), resulting in some expansion or contraction of the proteins. It left three precisely colinear blocks of amino acids in the herpesvirus

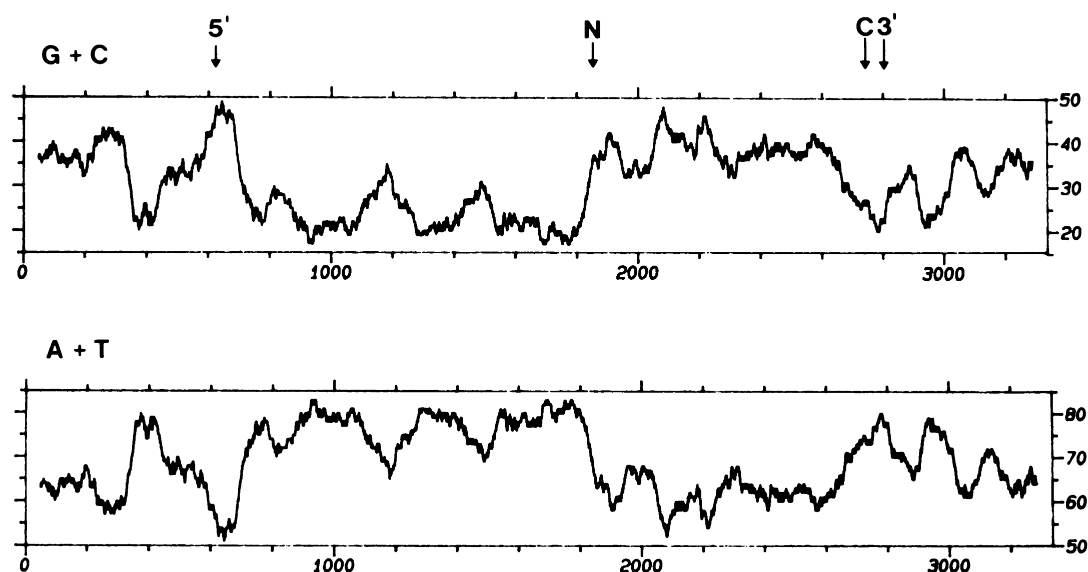

FIG. 7. Distribution of the G+C and A+T content within the herpesvirus saimiri TS gene. Arrows indicate the transcription initiation site (5′) and the polyadenylation signal (3′) for the mRNA and also the amino (N) and carboxy (C) termini of the TS-coding reading frame.

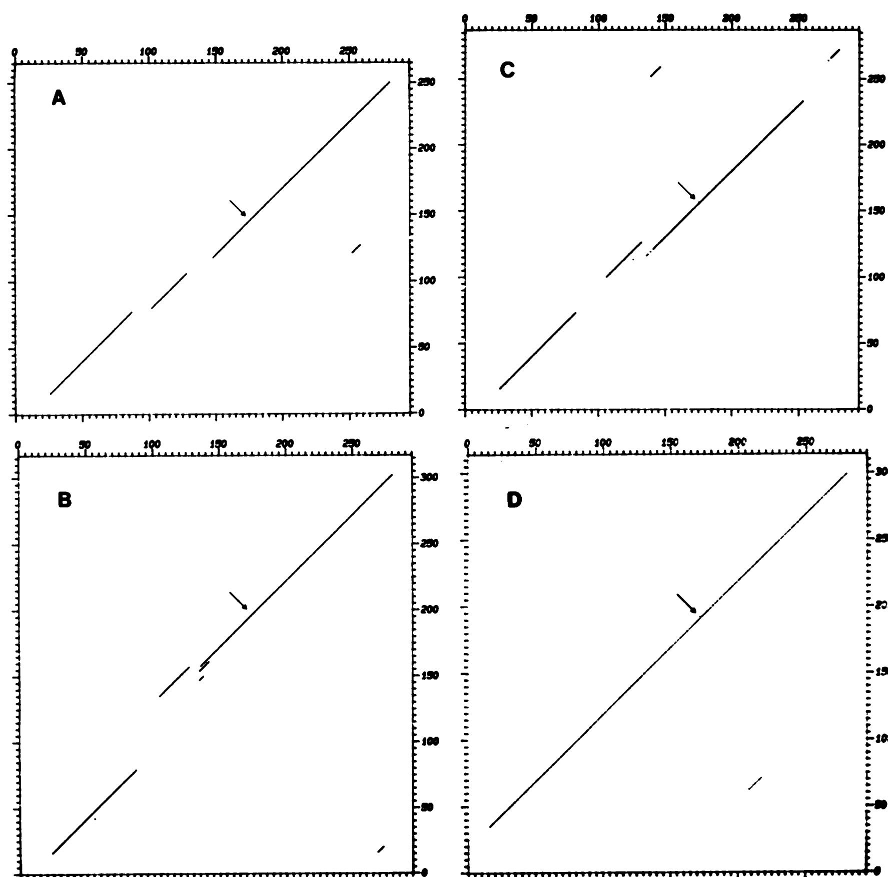

FIG. 8. Computer-based dot matrix comparisons of the amino acid sequences of the TS gene in herpesvirus saimiri with TS genes from *E. coli* (A), *L. casei* cells (B), bacteriophage T4 (C), and human cells (D). The analysis was performed to indicate a homology of 8 amino acids within a window of 30 amino acids. The herpesvirus saimiri sequence is delineated on the abscissa (294 amino acids).

saimiri TS gene, having 47% identity with the *E. coli* and *L. casei* genes and having 42% identity with the bacteriophage T4 gene. The enzymes from human cells and herpesvirus saimiri were found to be 70% identical without a single amino acid inserted or deleted within the entire sequence of homology. The human TS is slightly higher in molecular mass (35.675 kDa), as it has an additional amino-terminal sequence of 19 amino acids. Nucleotide sequence composition in the TS genes did reflect the patterns of amino acid identity. These computer studies, however, did not recognize any DNA sequence homology in the untranslated regions of the transcription units or in the transcriptional regulatory sequences of the various TS genes (Fig. 9).

DISCUSSION

Herpesvirus saimiri encodes a TS of 33.5 kDa which closely resembles in primary structure the equivalent enzyme of human cells and has striking similarities with the TSs from various procaryotes (*E. coli*, *L. casei*, *B. subtilis* bacteriophage 3T, and phage T4) (22). The TS gene of herpesvirus saimiri appeared to be very weakly transcribed into a 2.5-kb mRNA at conditions of immediate early expression; it is strongly expressed during the late phase of replication (6). The same transcripts were found in cultured lymphoid tumor cells in the absence of any other detectable virus mRNA (27). This paper describes the highly unusual structural organization of the TS gene in herpesvirus saimiri. DNA sequencing of the entire viral gene and nuclease protection analyses revealed a nonspliced mRNA, consisting of 2,187 nucleotides from the 5′ end to the polyadenylation site. The 5′ untranslated sequence of this RNA is unusually long; it measures 1,207 nucleotides and has multiple translational start and stop codons in series. This seems to be unique for the herpesvirus saimiri TS gene. In contrast to the open reading frames, nucleotide sequences of 5′ and 3′ untranslated RNA did not show homology with the untranslated parts of the characterized human or procaryotic TS genes.

Long 5′ untranslated regions had been found in regulatory genes for differentiation (mouse c-*myc*) (9), for homone release (luteinizing-hormone-releasing hormone [36]), and for metabolic control (general control nonderepressible gene of yeast [GCN4]) (38) and in the genes for growth factors such as insulin-like growth factor (IGF II/MSA) (12) and platelet-derived growth factor, β-chain (PDGF) (8). However, none of these genes exceeds the herpesvirus saimiri TS gene in the length of the 5′ untranslated sequence. Fully characterized regulatory (1) and structural protein (39) genes of herpesviruses did not have comparable 5′ untranslated RNA sequences. The long 5′ untranslated regions of regulatory genes appear to be of functional importance; it was shown that deletion variants of the noncoding DNA of the mouse c-*myc* gene and GCN4 in yeast cells are more effectively translated in vitro (9) and in living cells (20).

Herpesviruses with a high G+C content in their protein-

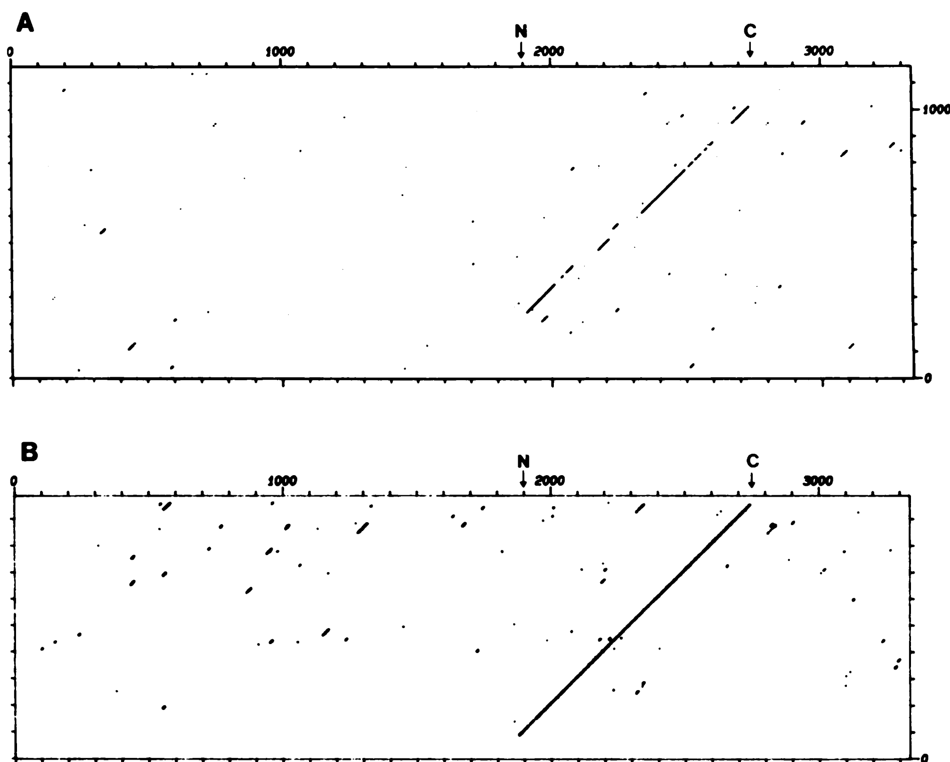

FIG. 9. Computer-based dot matrix comparison of the nucleotide sequence in the TS gene of herpesvirus saimiri with TS sequences from *E. coli* (A) and human cells (B). The herpesvirus saimiri sequence is on the horizontal axis (3,336 nucleotides).

coding DNA, such as herpes simplex virus type 1 (67%) (32), pseudorabies virus (69%) (19), and human cytomegalovirus (56%) (13), did not induce TS activity in permissive cells during virus replication (22), and no nucleotide or amino acid sequence homology was detected in a comparison of the TS gene of herpesvirus saimiri with the entire sequence of Epstein-Barr virus B95-8 (59%) (3). On the other hand, the varicella-zoster virus (VZV), which has a low G+C genome (46%) (23), does encode a TS protein sharing amino acid sequence homology with the TS of herpesvirus saimiri and human cells (R. Honess and A. Davison, personal communication). Thus, so far, TS genes were found only in herpesviruses with low G+C content in their viral DNA; however, there is no correlation with taxonomic subgrouping. This raises questions concerning the evolutionary origin of TS genes in herpesvirus genomes and their functional importance for viral replication. As concentrations of deoxynucleotide triphosphates, first of all dTTP, are considered rate limiting in cellular DNA synthesis (2), the increase of the dTTP pool in herpesvirus-infected cells by a virus-instructed enzyme may accelerate replication of A+T-rich herpesviruses, contributing to their high demand for dTTP. However, a virus-encoded TS gene is not the sole determinant of replication velocity; herpesvirus saimiri and varicella-zoster virus, in general, replicate rather poorly in culture systems, while herpes simplex and pseudorabies viruses grow exceedingly fast.

It could be hypothesized that a progenitor herpesvirus had a functional viral TS gene. The gene may have been lost independently in different herpesvirus subgroups with the evolution of some of their members to very high G+C content. The pronounced shift to high G+C content in some herpesviruses may not have been compatible with the conservation of an enzyme that requires extreme evolutionary stability in amino acid sequence to preserve its functions. However, another hypothesis that may be compatible with all observations is that a few herpesviruses of different subgroups have acquired their TS genes independently from the host genome. So far, TS genes are restricted to very few herpes group members. They are only found in varicella-zoster virus (alpha herpesvirus) (R. Honess and A. Davison, personal communication) and herpesvirus saimiri with its close relatives (gamma herpesviruses) (22) (I. Puchtler and B. Fleckenstein, unpublished results). The majority of herpesviruses in all subgroups (including herpes simplex virus and pseudorabies virus as alpha herpesviruses, human cytomegalovirus as a beta herpesvirus, and Epstein-Barr virus as a gamma herpesvirus) do not have the gene (22). The idea of gene acquisition may also be supported by the discoordinate gene expression of the TS gene found in herpesvirus saimiri replication. Other herpesvirus genes are coordinately regulated in transcription. For instance, immediate early genes of herpesviruses are tightly repressed later in virus replication; only a few examples of late transcription of those genes have ben reported (24, 26). The TS gene of herpesvirus saimiri appeared weakly expressed at early times, and maximal concentrations of TS mRNA are found late in virus replication, when DNA synthesis is believed to have terminated (6). The atopic position of the translational reading frame within the TS transcription unit, resulting in an excessively long 5′ untranslated sequence, could be the consequence of an event in which the cellular TS-coding DNA sequence was nearly precisely transposed into an unrelated viral sequence environment. Future DNA sequencing work revealing the colinearity in gene organization of the TS-coding herpesviruses will decide whether the TS

genes were gained by randomly positioned insertion of a cellular gene. So far, there is no other comparable example known of cellular gene acquisition by a DNA virus.

Transcripts identical with the TS mRNA of lytically infected permissive cells were found in lymphoid tumor cell lines (27), and other polyadenylated transcripts were not detected. Nonintegrated circular herpesvirus saimiri genomes persisting in those cell lines can lose viral sequences totaling up to 73% of the coding DNA; however, the TS gene of herpesvirus saimiri has always been preserved (34). This could hint at a functional role of viral TS in lymphocyte transformation. It had been demonstrated that dTTP depletion in TS^- cultured cells can result in DNA strand breaks (2) and chromosomal rearrangement (37). Heritable disorders of the human TS gene have been reported to correlate with fragile chromosomal sites and mental retardation (37). Possibly, the continuous disregulated expression of an aberrant herpesvirus saimiri TS protein in persistently infected lymphoid cells could interfere with the synthesis of normally functioning cellular TS, induce mutations, and thus contribute to the neoplastic phenotype. Future studies aimed at the deletion or replacement of the herpesvirus saimiri TS gene may help to substantiate these hypotheses.

ACKNOWLEDGMENTS

We thank Thomas Iftner and Walter Dietrich for support in the computer analyses of nucleotide sequences, Bodo Plachter for synthesizing an oligonucleotide primer with the help of Wolfgang Hillen, Ronald C. Desrosiers for critical reading of the manuscript.

The contributions of H. H. Niller and N. Nitsche were made in partial fulfillment of the requirements for a doctoral degree.

This work was supported by Deutsche Forschungsgemeinschaft.

LITERATURE CITED

1. **Akrigg, A., G. W. G. Wilkinson, and J. D. Oram.** 1985. The structure of the major immediate early gene of human cytomegalovirus strain AD 169. Virus Res. **2**:107–121.
2. **Ayusawa, D., K. Shimizu, K. Hideki, T. Keiichi, and T. Seno.** 1983. Accumulation of DNA strand breaks during thymineless death in thymidylate synthase-negative mutants of mouse FM3A cells. J. Biol. Chem. **258**:12448–12454.
3. **Baer, R., A. T. Bankier, M. D. Biggin, P. L. Deininger, P. J. Farrell, T. J. Gibson, G. Gatfull, G. S. Hudson, C. Satchwell, P. S. Tuffnell, and B. G. Barrell.** 1985. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. Nature (London) **310**:207–211.
4. **Belfort, M., G. Maley, J. Pedersen-Lane, and F. Maley.** 1983. Primary structure of the *Escherichia coli thyA* gene and its thymidylate synthase product. Proc. Natl. Acad. Sci. USA **80**:4914–4918.
5. **Birnstiel, M. L., M. Busslinger, and K. Strub.** 1985. Transcription termination and 3′ processing: the end is in site! Cell **41**:349–359.
6. **Bodemer, W., E. Knust, S. Angermüller, and B. Fleckenstein.** 1984. Immediate-early transcription of herpesvirus saimiri. J. Virol. **51**:452–457.
7. **Chu, F. K., G. F. Maley, F. Maley, and M. Belfort.** 1984. Intervening sequence in the thymidylate synthase gene of bacteriophage T4. Proc. Natl. Acad. Sci. USA **81**:3049–3053.
8. **Collins, T., D. Ginsburg, J. M. Boss, S. H. Orkin, and J. S. Pober.** 1985. Cultured human endothelial cells express platelet-derived growth factor β-chain: cDNA cloning and structural analysis. Nature (London) **316**:748–750.
9. **Darveau, A., J. Pelletier, and N. Sonenberg.** 1985. Differential efficiencies of in vitro translation of mouse c-*myc* transcripts differing in the 5′-untranslated region. Proc. Natl. Acad. Sci. USA **82**:2315–2319.
10. **Desrosiers, R. C.** 1981. Herpesvirus saimiri DNA in tumor cells—deleted sequences and sequence rearrangements. J. Virol. **39**:497–509.
11. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
12. **Dull, T. J., A. Gray, J. S. Hayflick, and A. Ullrich.** 1984. Insulin-like growth factor II precursor gene organization in relation to insulin gene family. Nature (London) **310**:777–781.
13. **Ebeling, A., G. Keil, B. Nowak, B. Fleckenstein, N. Berthelot, and P. Sheldrick.** 1983. Genome structure and virion polypeptides of the primate herpesviruses herpesvirus aotus types 1 and 3: comparison with human cytomegalovirus. J. Virol. **45**:715–726.
14. **Favaloro, J., R. Treisman, and R. Kamen.** 1980. Transcription maps of polyoma virus-specific RNA: Analysis by two-dimensional nuclease S1 gel mapping. Methods Enzymol. **65**:718–749.
15. **Fleckenstein, B.** 1979. Oncogenic herpesviruses of non-human primates. Biochim. Biophys. Acta **560**:301–342.
16. **Fleckenstein, B., and R. C. Desrosiers.** 1982. Herpesvirus saimiri and herpesvirus ateles, p. 253–332. *In* B. Roizman (ed.), The herpesviruses, vol. 1. Plenum Publishing Corp., New York.
17. **Gil, A., and N. J. Proudfoot.** 1984. A sequence downstream of AAUAAA is required for rabbit β-globin mRNA 3′ end formation. Nature (London) **312**:473–474.
18. **Glisin, V., R. Crkvenjakov, and C. Byus.** 1974. Ribonucleic acid isolated by cesium chloride centrifugation. Biochemistry **13**:2633–2637.
19. **Gustafsohn, D. P.** 1970. Pseudorabies, p. 337–355. *In* H. W. Dunne (ed.), Diseases of swine, 3rd ed. Iowa State University Press, Ames, Iowa.
20. **Hinnebusch, A. G.** 1985. A hierarchy of *trans*-acting factors modulates translation of an activator of amino acid biosynthetic genes in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **5**:2349–2360.
21. **Holmes, D. S., and M. Quigley.** 1981. A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. **114**:193–197.
22. **Honess, R. W., W. Bodemer, K. R. Cameron, H. H. Niller, B. Fleckenstein, and R. E. Randall.** 1986. The (A+T)-rich genome of the oncogenic gammaherpesvirus, herpesvirus saimiri, contains a highly conserved gene for thymidylate synthase. Proc. Natl. Acad. Sci. USA **83**:3604–3608.
23. **Iltis, J. P., J. E. Oakes, R. W. Hyman, and F. Rapp.** 1977. Comparison of the DNAs of varicella zoster viruses isolated from clinical cases of varicella and herpes zoster. Virology **82**:345–352.
24. **Jahn, G., E. Knust, H. Schmolla, T. Sarre, J. A. Nelson, J. K. McDougall, and B. Fleckenstein.** 1984. Predominant immediate-early transcripts of human cytomegalovirus AD 169. J. Virol. **49**:363–370.
25. **Kaschka-Dierich, C., F. J. Werner, I. Bauer, and B. Fleckenstein.** 1982. Structure of nonintegrated, circular herpesvirus saimiri and herpesvirus ateles genomes in tumor cell lines and in vitro-transformed cells. J.Virol. **44**:295–310.
26. **Keil, G. M., A. Ebeling-Keil, and K. U. Koszinowski.** 1984. Temporal regulation of murine cytomegalovirus transcription and mapping of viral RNA synthesized at immmediate early times after infection. J. Virol. **50**:784–795.
27. **Knust, E., W. Dietrich, B. Fleckenstein, and W. Bodemer.** 1983. Virus-specific transcription in a herpesvirus saimiri-transformed lymphoid tumor cell line. J. Virol. **48**:377–383.
28. **Lee, D. C., and R. G. Roeder.** 1981. Transcription of adenovirus type 2 genes in a cell-free system: apparent heterogeneity of initiation at some promoters. Mol. Cell. Biol. **1**:635–651.
29. **Maley, G. F., R. L. Bellisario, D. U. Guarino, and F. Maley.** 1979. The primary structure of *Lactobacillus casei* thymidylate synthetase. J. Biol. Chem. **254**:1301–1304.
30. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
31. **McDevitt, M. A., M. J. Imperiale, H. Ali, and J. R. Nevins.** 1984. Requirement of a downstream sequence for generation of a poly

(A) addition site. Cell **37**:993–999.

32. **Roizman, B.** 1979. The structure and isomerization of herpes simplex virus genomes. Cell **16**:481–494.
33. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
34. **Schirm, S., I. Müller, R. C. Desrosiers, and B. Fleckenstein.** 1984. Herpesvirus saimiri DNA in a lymphoid cell line established by in vitro transformation. J. Virol. **49**:938–946.
35. **Schirm, S., F. Weber, W. Schaffner, and B. Fleckenstein.** 1985. A transcription enhancer in the herpesvirus saimiri genome. EMBO J. **4**:2669–2674.
36. **Seeburg, P. H., and J. P. Adelman.** 1984. Characterization of cDNA for precursor of human luteinizing hormone releasing hormone. Nature (London) **311**:666–668.
37. **Takeishi, K., S. Kaneda, D. Ayusawa, K. Shimizu, O. Gotoh, and T. Seno.** 1985. Nucleotide sequence of a functional cDNA for human thymidylate synthase. Nucleic Acids Res. **13**:2035–2043.
38. **Thireos, G., M. D. Penn, and H. Greer.** 1984. 5'-Untranslated sequences are required for the translational control of a yeast regulatory gene. Proc. Natl. Acad. Sci. USA **81**:5096–5100.
39. **Watson, R. J., J. H. Whis, J. S. Selstrom, and L. W. Enquist.** 1982. Herpes simplex virus type 1 glycoprotein D gene: nucleotide sequence and expression in *Escherichia coli*. Science **218**:381–384.
40. **Werner, F.-J., G. W. Bornkamm, and B. Fleckenstein.** 1977. Episomal viral DNA in a herpesvirus saimiri-transformed lymphoid cell line. J.Virol. **22**:794–803.
41. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13 mp 18 and pUC19 vectors. Gene **33**:103–119.