SHRIDHARA C. S. MURTHY, JOHN J. TRIMBLE, AND RONALD C. DESROSIERS*

New England Regional Primate Research Center, Harvard Medical School, One Pine Hill Drive, Southborough, Massachusetts 01772

Received 6 March 1989/Accepted 21 April 1989

Analysis of a 5,549-base-pair sequence at the left end of herpesvirus saimiri unique (L-) DNA revealed two open reading frames and genes for five small nuclear U RNAs (herpesvirus saimiri U RNAs). Replicationcompetent deletion mutants were constructed in order to assess the importance of these genetic features for transformation by this oncogenic herpesvirus. Although not required for replication, one of the open reading frames was found to be required for immortalization of marmoset T lymphocytes into continuously growing cell lines. The protein predicted by this reading frame (STP; saimiri transformation-associated protein) has a highly hydrophobic stretch of 26 amino acids sufficient for a membrane-spanning domain near its carboxy terminus; this domain is immediately preceded by a sequence appropriate for formation of a metal-binding domain (His X_2 His X_6 Cys X_2 Cys, where Xs are other amino acids). One of two poly(A)⁺ RNAs that could encode STP is bicistronic, while the other has a long 5' untranslated region (~1.5 kilobases). Although some analogies can be drawn between STP and LMP (lymphocyte membrane protein) of Epstein-Barr virus, STP is not related in sequence to LMP.

Herpesviruses can be classified into three subfamilies according to biological and molecular criteria (29). A human herpesvirus serves as the prototype for each of the subfamilies: herpes simplex virus for the alphaherpesviruses, human cytomegalovirus for the betaherpesviruses, and Epstein-Barr virus (EBV) for the gamma- or lymphotropic herpesviruses. Among the herpesviruses, only members of the gamma group have been shown to be oncogenic and to possess genes whose expressions are associated with transformation. EBV expresses several nuclear antigens (EBNAs) and a lymphocyte membrane protein (LMP) in tumor and growth-transformed B cells; these products appear to be consistently expressed in continuously growing B cells immortalized by EBV (7). Furthermore, a gene construct expressing LMP is able to transform RAT-1 cells in vitro, and these transformed cells induce tumors in nude mice (33).

Herpesvirus saimiri, another member of the gamma subfamily, produces no signs of disease in its natural host, the squirrel monkey (25). Infection of other species of New World primates results in rapidly progressing, fatal lymphoproliferative diseases including lymphomas, leukemias, and lymphosarcomas (for review, see reference 12). The virus is also able to immortalize lymphocytes in vitro (10). The target cells for herpesvirus saimiri persistence in squirrel monkeys and for oncogenic transformation in marmosets and owl monkeys are T lymphocytes (28, 36). The cell lines derived from immortalization of common marmoset lymphocytes with herpesvirus saimiri represent a distinct minor subpopulation of T lymphocytes, a $T12^+$ T8⁺ NKH1⁺ T4⁻ T lymphocyte with NK activity (10, 19). The target cell for transformation by herpesvirus saimiri is thus very different from that of EBV.

Previous studies have identified a region of the herpesvirus saimiri genome required for immortalization of marmoset lymphocytes in vitro and for oncogenicity in vivo; this region is not required for replication of the virus (8, 10). This identification was greatly facilitated by the ability to grow herpesvirus saimiri permissively in monolayer cell lines. This feature allows for construction of site-specific mutants via recombination of virion DNA in an infected cell with cloned, mutated DNA (9). Replication-competent mutants of herpesvirus saimiri with deletions in this region no longer immortalized marmoset lymphocytes in vitro and were no longer oncogenic in vivo; restoration of the deleted sequences restored these abilities (8, 10).

We describe in this report the nucleotide sequence and unusual genomic organization in this region required for oncogenicity. The results of mutational analyses provide strong evidence for the requirement for one of two open reading frames in this region for transformation.

MATERIALS AND METHODS

Molecular cloning and sequencing. The starting parent cloned DNAs, pT7.4 and pTp8, of the oncogenic herpesvirus

TABLE 1. Replication-competent deletion mutants of herpesvirus saimiri

| Deletion mutant | Nucleotides deleted ^a | Restriction sites used ^b | Ability to immortalize | Onco- genicity | | |
|----------------------------|-------------------------------------|--|---------------------------|-------------------|--|--|
| 11att ^d | -310 to +1810 | | - | _ | | |
| $\mathbf{K}\mathbf{H}^{d}$ | +965 to +1516 | KpnI-HpaI | + | + | | |
| Δ3 | +3328 to +5546 | Scal-Scal | + | NT | | |
| Δ1B | -30 to $+457$ | SstI-SstI | _ | NT | | |
| Δ1Α | +457 to +1264 | SstI-XbaI | - | NT | | |
| Δ2B* | 1 to +5200 | | _ | NT | | |
| Δ2Α | +1472 to +2851 | BglII-BglII | + | NT | | |
| S4 ^d | +457 to +4275 | SstI-SstI | - | - | | |
| MR | +965 to +4275 | | + | NT | | |

^a Nucleotide numbers can be found in Fig. 1. Negative numbers are nucleotides leftward into H-DNA.

^b 11att and $\Delta 2B^*$ are deletions that arose spontaneously. The limits of the 11att deletion were determined by DNA sequencing. The limits of the $\Delta 2B^*$ deletion are based on restriction endonuclease mapping and are approximate. ^c Oncogenicity determined previously (8). NT, Not tested.

 d The 11att, KH, and S4 deletion mutants have been described previously (8, 31).

^{*} Corresponding author.

| -180 | GCGCTGGGCT | CTGGGGCTGG | GGGGGCTTGA | ACAGTTGTGG | GACCETTACT | CTAGCAGCGC | |
|--------------|--------------|-------------|------------|--------------------------|----------------------------|-------------------|------------------|
| -120 | CTCGGCCTAG | CCAGGOCTCT | BEGEACTESC | TCTAAGCACA | BGBBCACABC | BCCCCCGGGC | |
| 1 | GGTCGCGAGG | BTCTAGCGCC | TCGAAACCGG | CTCBGAGCAC | AABCAGACTC | TABCCCCCTC | |
| 61 | CCCTAGTACA | CAGAGCCCAG | CAGGCAGCTA | CAGCCOCTCA | ACGCGAGTCC | CTCCCCTTGC | |
| 181 | GACATTTACT | GCAGCACTAT | GTGATTCACT | TTGATTCTTT | TACATTTTTT | TAAACATAAT | |
| 241 | TACTAGCATT | AAACCAATTA | TEATTAATAS | CAAAACAATA | ATAACTAGCA | GCAATABGAT | |
| 361 | BACTTCTBCB | TABACATOT | CTICACTICC | TECTOCICS | CARCCACTER | CACGTACTEC | STP |
| 421 | TGATAAGCCT | ACTOGOGTOC | TIAAAIGTBA | TRAGCTCCOT | GARCCAGATO | GIGITEGIAA | 164 ANTHO ACIDS |
| 481 | BECTACTOCT | CCCGATAGTA | CIBITABICI | TCCTBGBCAT | CCALITCH | <u>ACACIAGATA</u> | 104 141110 10100 |
| 601 | TAAGTATOST | GTAGETCCAT | CATCTCCATC | ACTICITICA | TCASTATTOT | STOGAGGATC | |
| 661 791 | TCCGTTGCTT | TCATCOTTT | CTTOTOGOTC | TCCTTCACCT | ABACCTCTTB | CCAITTTCTT | |
| 781 | TTATATAGTA | GCTTBTTACT | TCTTGGAAAS | TCCABCAAGA | TGGTGTCCTG | TTTAACAGCT | |
| 841 | TGACCACATG | TTTTACAGGA | CTTAAAAATT | TANATTTTAA | CCTTTTBACA | AAGAGCAAAA | |
| 961 | GTGTTAACTT | AUTTTATTAT | AAGTTBATBC | THECTATION | ABTOCTTATA | BCABCTITAC | |
| 1021 | TATCAGCTTT | TAGTAGTTAT | TOCTAOCTIT | ATCTAGCTTT | OCTCTCAATG | ABCTATAAAT | |
| 1081 | TATCTACABT | TTAATTATA | GATTACTATA | CAAGACAATG | TACCAACTTA | ABAACTACAT | |
| 1201 | GACATGTGTT | TTCTAGATTC | TACATGATAS | TABAATCTTB | AABCTTTACA | BETCTTCTET | |
| 1261 | CTAGAAGGCA | TATAGTAAAG | ATAAAGCTTA | TCTAGTGTCT | TCAAAAGTAA | GTAGTAGAGA | |
| 1381 | TTCATABAGA | TTAAACAGTS | AGTAABTTAG | TCAGTTTAAC | AGCTAGTGAT | ATAAAATGAT | |
| 1441 | ACACITGAGA | ACAGAAGTTC | TETETTTOCA | GAGATCTCGG | TAAATACCCA | AAGTCTACTT | NCIID 1 |
| 1561 | AGCAAGTAAC | GGGTTGTTTG | AGACTTTTAG | TTTAGGITAT | CACAGATITA | AGTTCCAGTA | 107 HT |
| 1621 | AACATTACTA | AGAAATAAAT | AAATAAATAT | BTAGTGTACT | TTAGCAGCTC | TATTTGACTA | 142 11 |
| 1741 | TTACAATTTA | TOCACTACAT | GACTACACAT | TAAAAGTABA | AGCATGTTGT | GTTTTCTCTT | |
| 1801 | TAAAGTACAG | CATAACCAGA | GATTGTTTTT | AACATATTTA | TETAATATEC | AAATTTCTBA | |
| 1861 | ATCACTTICA | ATTAATCAAG | GACTCAAAAA | TATTATTATA | GTAATATATT TAAAATATATAT | AGCTATOCTC | |
| 1981 | GTTTBTBCAB | TATGTTTGTG | AAATAATGTG | TAATTGTACA | CTACAATTTA | TCATTGTTAT | |
| 2041 | TTACTTACAA | AGCAAACTCA | CAAGTTTACT | ATTTAGCTAC | AAGTTTCTTA | AAAACAATAT | |
| 2161 | TTCTTGACTT | TTBAGAAGTT | TATECTTAAT | TECAAGACAT | CATBAAACTT | TATTACTTAA | |
| 2221 | ATTTTGTTTA | AATAGETACA | AATTTATTAA | GCATAGAGCA | TTOCTTTATA | AACAATTTTT | |
| 2341 | GAGCACAGAT | ACCTTTTTTT | AAACTTAAGG | CAGCOCTOGT | TTTTAAATAT | GTACACCCAG | HSUR 2 |
| 2401 | COCTTTCAAA | AATTAGTAGA | AGCEATACCT | CGTGTGTGAT | TAAAGCAGCA | TTATCAGETA | 115 NT |
| 2961 | TTAACATATT | TTAAAGTTAG | GGTGGTGGTT | TTATCTTTCA | CATTTTTACCC | TTTTTABTTC | 223 41 |
| 2581 | CAAAGAACCG | CETACATCAA | AAGTTAGGGC | ATTTTGCTGA | AAAGGTCATA | GTTAAAACBT | |
| 2641 | TITCACITIC | GATETTTEE | ATRIACTIT | TTCTGAGTAA | TATTCAAATT | TRACATATAA | |
| 2761 | AGTGCAAGCA | TTAGTAGTGT | ATATATACAG | TAAAGTCAGC | TAATTTTAAG | AAABTACATT | |
| 2821 | AAAAAGCTAG | CAAATTÍGTT | TTTAACTTAC | AGATCTTAAA | CATTTACATE | CATACGTTAA | |
| 2941 | GAAGTTCTAA | ATAATGTAGC | AGCGCTCTCA | CTOCTOTAAC | TTCTTGTTGA | AGTATCAABC | |
| 3001 | AGTAAGAGCG | CGGACTTTCT | GTTTTTGTAC | TGACTTAGGC | GOCTCTGGTT | BTTABTAACA | hsur 5 |
| 3121 | AGCTTTGCGA | GCGAAAAATA | AATATGTAGT | GTTCCCTTTA | AGCAATATTT | TCTTCTTATA | 114 NT |
| 3181 | AGTTTCTATT | ACTAGTTTTT | AAGTTAGGGT | ATGCOGTAGT | TTTCTTACAA | TTTGTCTACT | |
| 3241 3301 | AAAGATAGCA | CAATATTTTC | CCTAAGTACT | TATATTCTBA | BTAATATTCA | GCATTACTTA | |
| 3361 | AAAGETETAA | TTTCCTGCTG | TTTGCTGCAC | TOTTCATACT | TACTGTGTTT | AAABCTATTT | |
| 3421 | GCTACATAAA | AGCTIGIGIT | TTAAACAAAG | TTAATAATAA | CACTAATATA BTABAGCTTC | CATTICIANA | |
| 3541 | AGAAAGTATC | TTTATATGCG | TTTGTAAAAA | ACAGTTATTT | ACTAGTTAGT | GABTAATTTT | |
| 3601 | TGTTATAAAA | GTTTTAGTTA | TTAAGTGTAT | BTAAAACTTC | ACAATAAGTC | AGTATTABCT | |
| 3721 | AGCACTTTCA | TTTTCT6GCC | GCTTTTAAAT | BCTTABCTCA | CATATATITT | ATATATAACT | |
| 3781 | TETATCTETA | TTAGAATCTT | TTCTTATACT | TTAGCTTACT | ATAATTTAAA | ATTAGAAATA | |
| 3901 | GTGAATTTAT | AATGTATATG | TAAGTTTTCA | CAAATACATG | ACATCTTCAC | CTETAAAAGA | |
| 3961 | CATTTOTATL | AATGATICIT_ | CICATATACT | TCANATTIST | ACTTAATACT | TTTCTCTTCT | |
| 4021 | TCTGGAAAAA | AGGIGICACA | TICAAABTCT | TECATEATCE | TAGTCACAAA | CASTTAASA | DHFR |
| 4141 | TEGCATOGAT | AACTCATAGC | TTCTTTATAC | ACAGAACTAC | CTCCAATTAT | CCAGACCATS | 187 AMINO ACIDS |
| 4201 | TCTACTITAT | GAGCTCTATE | TOBCTOTICS | STAASTTTTA TTTASTTCCT | TECTROBAC | TAAACTETTA | |
| 4321 | CTOCCTTTTA | AAGGCCGGTT | CTTCTCAGGA | ATTGAGAACC | AABTCTTTTT | ACCCATAATC | |
| 4381 | ACTAGATICT | BTTTATCTBB | TACABAAGAT | GTAGTAGTCA | TICTITODAA | ATSTITAAAA | |
| 4501 | ACAGCAACAA | TECAGITIAG | TOCTTOAACC | ATAATACTAA | CTAAAGACTT | CTTCTTCTCT | |
| 4561 | AGTAATAAAA | AGTAAATAAT | CTTOCTOTAT | TITTCTABTE | TATTTTTT | TIGTOTITIA | uend 7 |
| 4681 | ACTEGTTTEG | ACCTAAABAC | CCAGAGCCTT | CANAAATTAT | TCTTOGETTE | TTAATCTCCT | 76 JUGIN |
| 4741 | ATAGCAAGTC | TTAAATATAC | ATECTATTIE | TITATAACAT | ATAACACTAT | CTATTATTTA | 70 NT |
| 4861 | AAGTGCAAGC | AAGCTGTTGT | AAGATTCTTT | TATAAATTTT | TATCOTAATT | TTTTTTTACAA | |
| 4921 | AAAGBACABA | BAATACTCAA | ATTACTTCAS | ABCTTTCATT | TTAAAAAATT | ACATCCTCTT | |
| 4781 5041 | TOCATOTICAAT | AGTTTATAR | CTBTAGCAAC | ACOSTITIT | TICTOCCCAC | TUCTATAAAC | |
| 5101 | AGTOGCACTO | STTTOGACTA | CCCCAGAGCC | TTCAAAAATT | OCTOAOTOCT | TETBAAACTG | HSUR 4 |
| 5161 | AGAAGCTCCA | ACTOTATOAA | CCTBAAGAGT | AACTCTCTBG | CTOTOGOCAG | TCTCTGTTCA | 106 NT |
| 5281 | TOTCTOTAAA | ACAGATTAAT | BCAAAATAGC | AAAATAAAAG | TGTATETOAA | CAATTETTAC | |
| 5341 | TACTATATAS | ATATACASTA | BTTATTTTT | ATAAGAABTT | TAATTAAGCT | BTATTCAAAT | |
| 3401 5461 | ATGCAATATT 1 | TTTTAGTACA | CBAATACATT | TTCAAAATTA | TTGTTGATTT | TTTAAAGCTA | |
| 5521 | AGTTTTAAAA | GCTTOTTTG | TTTAGTACT | | | | |

FIG. 1. Sequence of herpesvirus saimiri DNA at the left junction of H- and L-DNA. The first nucleotide of L-DNA, nucleotide +1, is the first nucleotide to digress from the standard H-DNA repeat unit. By this convention, the first 145 nucleotides (nt) of L-DNA actually represent rearranged H-DNA sequences. Nucleotide -1 in this convention corresponds to H-DNA nucleotide 845 in the H-DNA repeat sequence of Bankier et al. (2). "L-DNA" nucleotides 1 to 38 correspond to H-DNA nucleotides 1425 to 18 of Bankier et al. (2), with 38 out of 38 matching nucleotides 39 to 83 correspond to H-DNA nucleotides 1131 to 1175 in the numbering system of Bankier et al., with 44 of 45 matching nucleotides. "L-DNA" nucleotides 84 to 116 correspond to H-DNA nucleotides 102 to 134 of Bankier et al., with 26 of 33 matching nucleotides. Nucleotide 145 and beyond contain low G+C, true L-DNA sequences. There is high concordance between the nucleotide sequence and the 20+ restriction sites that were previously mapped (9, 20). The *TaqI* site we had determined to be within 100 nucleotides of the H-L-DNA border (20) was actually located 21 nucleotides from the border. A sequence for *XbaI* cleavage (TCTAGA) at 4069 is not recognized in cloned DNA presumably because of a contiguous GATC recognition site for *dam* methylation. What had been mapped previously as single sites for *XbaI* and *HindIII* around 1.2 and 1.3 kbp turned out to be two closely spaced sites for *HpaII* and *FnuDII* near the H-DNA border.

saimiri strain 11 have been described previously (9, 20). Smaller subclones were constructed based on detailed restriction maps. In addition, restriction enzymes RsaI and DraI, which cleave the cloned herpesvirus saimiri DNA frequently, were used to generate series of smaller overlapping clones. For sequencing of some regions, series of nested deletions were constructed by the procedures of Henikoff (15) from one end of cloned herpesvirus saimiri DNA, using exonuclease III and S1 nuclease. Subclones generated by the above methods fully covered the 5.6kilobase (kb) left-end region of herpesvirus saimiri DNA. The double-stranded plasmid DNAs were alkaline denatured and sequenced by the dideoxy procedure of Sanger et al. (30), using either Klenow or reverse transcriptase enzymes. In some cases sequencing reactions were performed using Sequenase brand T7 DNA polymerase (United States Biochemicals, Cleveland, Ohio). Both strands of herpesvirus saimiri L-DNA in this region were sequenced at least twice to assemble the sequence data. The limits of the deletion in 11att were determined by sequencing the appropriate region of the pT6.0 clone derived from strain 11att virion DNA (20). The sequences were recorded with a digitizer gel reader (IBI, New Haven, Conn.) and stored in an IBM XT computer.

Construction of mutant viruses. The 11att, S4, and KH deletion mutants of herpesvirus saimiri were described previously (9, 31). The Δ 3, Δ 1A, Δ 1B, and Δ 2A deletion mutants

were isolated by using the same basic procedure as that used for the isolation of S4 and KH. Sequences were eliminated in cloned herpesvirus saimiri DNA by excision with the restriction endonucleases indicated in Table 1. $p\Delta 3$ was constructed by complete digestion of pT7.4 (cloned herpesvirus saimiri L-DNA sequences +21 to approximately +7400) with *ScaI*. Elimination of sequences for $\Delta 1B$ was achieved by partial digestion of pTp8 with *SstI*. Similarly, elimination of sequences for $p\Delta 2A$ was achieved by partial digestion of pT7.4 with *BglII*. $p\Delta 1A$ was constructed by excision of sequences between the *XbaI* (1212) and *SstI* (457) sites in clone pHp1.4 (cloned L-DNA sequences +27 to +1520) and blunting of the ends. After transformation of *Escherichia coli*, appropriate clones were isolated and characterized.

Recombinant plasmid pMR was constructed by adding 508 nucleotides of L-DNA sequences from 457 to 964 to plasmid pS4 at its *SstI* site. This restored the complete coding sequences for STP, but nucleotides 965 to 4275 remained deleted. Recombinant plasmid pMR was checked for continuity and integrity of the sequences by restriction enzyme analysis and sequencing across the junction, using synthetic oligonucleotides as primers.

The limits of the deletions in 11att, KH, and MR cloned DNAs were determined by DNA sequencing. Cloned DNAs p3, p1B, p1A, p2A, and pS4 were analyzed by restriction endonuclease mapping and by Southern blot hybridization. Southern blot hybridization used fragments from within the



FIG. 2. Analysis of open reading frames. A, Left to right; B, right to left (reverse and complement); C, scale in kilobase pairs; D, relative locations of restriction sites. Horizontal lines in A and B represent stop codons, derived from the sequence shown in Fig. 1. Code for restriction sites: t, *TaqI*; m, *HpaII*; p, *PstI*; s, *SstI*; d, *HindIII*; v, *PvuII*; h, *HpaI*; x, *XbaI*; b, *Bg/II*; k, *KpnI*; c, *ScaI*.



FIG. 3. Sequences required for immortalization by herpesvirus saimiri. The code for restriction endonuclease sites is the same as in the legend of Fig. 2; a, *SmaI*. Hatched bars represent deleted sequences.

deleted sequences to verify the absence of hybridization as well as larger fragments to confirm the size of appropriate restriction fragments.

To isolate virus with these deletions, owl monkey kidney cells (OMK 637 cell line) were transfected with full-length, infectious herpesvirus saimiri strain 11 virion DNA together with the mutated plasmid DNA; Ca^{2+} coprecipitation was used. Recombinant viruses were identified by a limiting dilution-spot hybridization assay, purified, and characterized as previously described (9). Characterization included hybridization of Southern blots of restricted virion DNA with fragments from within the deleted sequences to verify the absence of hybridization and with larger fragments to confirm the size of appropriate restriction products. Except for MR, recombinant viruses showed no evidence of plasmid vector sequences in their virion DNA.

The $\Delta 2B^*$ mutant arose spontaneously in a transfected cell culture. The limits of the deletion were determined approximately by restriction endonuclease mapping of virion DNA.

5'-End analysis by primer extension. Cytoplasmic RNA was extracted from herpesvirus saimiri strain 11-infected OMK cells by lysing the cells using Nonidet P-40 in hypotonic buffer as described (23). DNA primer extension analysis was performed as described (1). Two 35-mer oligonucleotides (L-DNA nucleotides 4536 to 4570 and 4613 to 4647)

were ³²P labeled at their 5' termini, using T4 polynucleotide kinase. After overnight hybridization of 10^5 dpm of ³²Plabeled oligonucleotide with 40 µg of herpesvirus saimiriinfected OMK cell RNA at 30°C in 80% formamide, the oligonucleotide-RNA hybrids were precipitated with 0.3 M sodium acetate and ethanol. Extension reactions of the labeled oligonucleotide primers hybridized to RNA were performed using avian myeloblastosis virus reverse transcriptase at 42°C for 90 min. After RNase A digestion, the samples were precipitated and suspended in 4 µl of formamide loading buffer. The samples were then boiled for 3 min and loaded onto an 8% polyacrylamide–8 M urea sequencing gel along with pGEM-2 *Hpa*II DNA size markers end labeled with [³²P]dGTP, [³²P]dCTP, and Klenow polymerase.

Immortalization of common marmoset lymphocytes. Immortalization of lymphocytes from common marmosets (*Callithrix jacchus*) was attempted in 25-cm² flasks and in 24-well Linbro tissue culture plates as previously described (10). Media contained 5 mg of 2-mercaptoethanol per liter. Immortalization was reproducible with each immortalization-competent deletion mutant. For each deletion mutant, we verified that the virus being produced by the immortalized cell line had the appropriate deletion in its virion DNA. The deletion mutants that were unable to immortalize com-



FIG. 4. Mapping of 5' end of 4.9-kb RNA by primer extension. Cytoplasmic RNA from herpesvirus saimiri-infected OMK cells was hybridized to ${}^{32}P$ -5'-end-labeled oligonucleotides representing L-DNA nucleotides 4536 to 4570 (lane 1) or L-DNA nucleotides 4613 to 4647 (lane 2). Labeled, hybridized oligonucleotide primers were extended by reverse transcriptase and analyzed on a denaturing polyacrylamide sequencing gel. DNA size markers are shown in lane M.

mon marmoset peripheral blood lymphocytes were tested on at least six different occasions.

RESULTS

Nucleotide sequence and genomic organization. The herpesvirus saimiri genome contains a centrally located stretch of 110 kilobase pairs (kbp) of unique sequence DNA called L-DNA because it is low in G+C content (36%). This is flanked at each end by a variable number of 1,444-bp repeat units called H-DNA that is high in G+C content (71%) (2, 4). Sequences previously identified as required for transformation and oncogenicity are located near the left junction of unique L- and repetitive H-DNA.

Our nucleotide sequencing has identified the transition

sequences at this junction (Fig. 1). Nucleotide 1 of L-DNA is defined as the first nucleotide to digress from the standard H-DNA repeat unit. Using this convention, however, the first 145 bp of L-DNA actually represents rearranged H-DNA sequences (Fig. 1, legend). The nucleotide sequence in this region agrees with the location of the junction previously determined by restriction endonuclease mapping (20).

Analysis of the first 5,549 nucleotides of L-DNA sequence revealed only two open reading frames larger than 70 amino acids (Fig. 2). Both open reading frames are present in a right-to-left orientation. The rightmost open reading frame in this region was previously shown to correspond to coding sequences for dihydrofolate reductase (DHFR) (32). The other open reading frame predicts a protein that we will refer to as the saimiri transformation-associated protein (STP) for reasons that are described below.

This 5,549-bp region of herpesvirus saimiri also encodes five small nuclear U RNAs referred to as herpesvirus saimiri U RNAs (HSURs) (22, 27, 35). HSURs are found in tumor cells and are 143 (HSUR 1), 115 (HSUR 2), 76 (HSUR 3), 106 (HSUR 4), and 111 to 114 (HSUR 5) nucleotides in length. The HSURs assemble into nuclear ribonucleoprotein particles of low abundance; they bind proteins with Sm determinants and possess a 5' trimethyl-G cap structure. The HSURs are distinct from any previously characterized cellular U RNAs. Coding sequences for these HSURs are derived from L-DNA nucleotides 1657 to 1515 (HSUR 1), 2485 to 2371 (HSUR 2), 4752 to 4677 (HSUR 3), 5208 to 5103 (HSUR 4), and 3153 to 3040 (HSUR 5) (Fig. 1 and 3). The viral genes encoding the HSURs possess conserved enhancer, promoter, and 3'-end formation signals typical of U RNA genes (22, 35). These are the first virus-encoded U RNAs to be identified. Their role in the virus life cycle remains to be determined.

Two $poly(A)^+$ RNAs of 4.9 and 2.3 kb were previously identified from this region in infected cells (18). Both RNAs are transcribed from right to left, i.e., an orientation appropriate for expression of the two open reading frames (18; Fig. 3). The two RNAs were shown previously to be 3' coterminal, with their 3' ends about 150 bp from the left border of unique L- and repetitive H-DNA (18). The sequence AAUAAA, located at nucleotides 166-161 in L-DNA, likely serves as the signal for poly(A) addition. Although S1 nuclease and exonuclease VII mapping procedures revealed the presence of two small introns in these $poly(A)^+$ RNAs (18; Fig. 3), these mapping experiments revealed no evidence for extensive splicing that could result in the creation of additional open reading frames. Using primer extension, the 5' end of the 4.9-kb RNA was mapped to the 4716-4728 region in L-DNA (Fig. 4), consistent with the previous S1 and exonuclease VII experiments. Although we have not determined the 5' end of the 2.3-kb RNA in infected cells by primer extension, the previous S1 and exonuclease VII mapping studies suggest that it is derived from the 2000–2300 region in L-DNA.

Mutational analysis. Replication-competent deletion mutants of herpesvirus saimiri were constructed in order to assess the importance of these genetic features for T-cell transformation. Deletions were constructed in cloned DNA, permissive OMK cells were cotransfected with virion DNA plus mutant cloned DNA, and recombinant viruses were isolated from the lysed cells. Mutant viruses were then tested for their ability to immortalize common marmoset peripheral blood lymphocytes (10).

The herpesvirus saimiri deletion mutant $\Delta 3$ is missing the entire DHFR open reading frame as well as the genes for

J. VIROL.

| | 710 * | 700 | | | | 690 * | | | | 680 * | | | 670 * | | | 660 # | | | |
|---------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| ATG Met | GCA Ala | AGA Arg | GGT Gly | CTA Leu | GGT Gly | GAA Glu | GGA Gly | GAC Asp | CCA Pro | CAA Gln | GAA Glu | AAC Asn | GAT Asp | GAA Glu | AGC Ser | AAC Asn | GGA Gly | GAT Asp | CCT Pro |
| | 650 | 640 | | | | 630 | | | | 620 | | | 610 | | | 600 | | | |
| CCA Pro | CAC | AAT Asn | ACT Thr | GAT Asp | GAA Glu | AGA Arg | AGT Ser | GAT Asp | GGA Gly | GAT Asp | GAT Asp | GGA Gly | CCT Pro | ACA Thr | CCA Pro | TAC Tyr | TTA Leu | CCT Pro | GTA Val |
| | 590 | 580 | | | | 570 | | | 560 | | | 550 | | | 540 | | | | |
| ACG Thr | CTT Leu | TTA Leu | AAT Asn | GCA Ala | GGG Gly | CCC Pro | TTT Phe | GGG Gly | CCT Pro | TAT Tyr | AAT Asn | CCC Pro | TAC Tyr | TGC Cys | TTG Leu | CTT Leu | GGC Gly | CAC His | CCA Pro |
| | 530 | 520 | | | | 510 | | | 500 | | | | 490 | | | 480 | | | |
| GTG Val | CAA Gln | GAA Glu | AGC Ser | GGA Gly | TGC Cys | CCA Pro | GGA Gly | ÂGA Arg | CCA Pro | ACA Thr | GCA Ala | CTA Leu | TCG Ser | GGA Gly | GCA Ala | GTA Val | GGC Gly | TTA Leu | CCA Pro |
| | 470 * | 460 | | | | 450 | | | 440 | | | | 430 * | | | 420 | | | |
| ACA Thr | CCA Pro | TCT Ser | GGC Gly | TCA Ser | CGG Arg | AGC Ser | TCA Ser | TCA Ser | CAT His | TTA Leu | AGC Ser | ACC Thr | CCA Pro | GTA Val | GGC Gly | TTA Leu | TCA Ser | GCA Ala | GTA Val |
| | 410 | | 400 | | | | 390 | | | 380 | | | 370 * | | | 360 | | | |
| CGT Arg | GTC Val | AGT Ser | GGC Gly | TGC Cys | GGA Gly | GGA Gly | GCA Ala | GGA Gly | AGT Ser | GAA Glu | GAA Glu | CAT His | GTC Val | TAC Tyr | GCA Ala | GAA Glu | GTC Val | GGA Gly | AGC Ser |
| | 350 | 340 | | | | 330 | | | 320 | | | 310 | | | 300 | | | | |
| CTG Leu | CAC His | AGT Ser | GAA G1u | CAC His | GAG Glu | CAA Gln | GAA Glu | GGT Gly | GAC Asp | AAA Lys | TGC Cys | ACA Thr | GAC Asp | TGT Cys | TCT Ser | GTA Val | ACT Thr | ATC 11e | CTA Leu |
| | 290 | 280 | | | | 270 | | | | 260 | | | 250 | | | 240 | | | |
| TTG L e u | CTG Leu | CTA Leu | GTT Val | ATT Ile | ATT Ile | GTT Val | TTG Leu | CTA Leu | TTA Leu | ATC Ile | ATA Ile | ATT Ile | GGT Gly | TTA Leu | ATG Met | CTA Leu | GTA Val | ATT Ile | ATG Met |
| | 230 | | | 220 |) | | | | | | | | | | | | | | |
| TTT Phe | AAA Lys | AAA Lys | ATG Met | TAA | | | | | | | | | | | | | | | |

FIG. 5. Predicted amino acid sequence of STP. The hydrophobic stretch is underlined. Boxes identify histidine and cysteine residues appropriate in location and spacing for a putative Zn²⁺-binding domain.

HSUR 3 and HSUR 4. The Δ 3 mutant nonetheless retains its ability to immortalize common marmoset lymphocytes. Herpesvirus saimiri produced by $\Delta 3$ -immortalized cell lines was verified to contain the 3328-5546 deletion in its virion DNA. These results demonstrate that DHFR and HSURs 3 and 4 are not necessary for immortalization of common marmoset lymphocytes by herpesvirus saimiri. Enhancer sequences for HSUR 5 are also deleted in the $\Delta 3$ virus. Similarly, $\Delta 2A$ deletes the genes for HSURs 1 and 2, and it too still immortalizes. Cell lines established by $\Delta 2A$ grow considerably more slowly than those established by wild-type herpesvirus saimiri strain 11 and by the $\Delta 3$ and KH deletion mutants. Herpesvirus saimiri produced by the $\Delta 2A$ -immortalized cell lines was verified to contain the 1472-2851 deletion in its virion DNA. These results demonstrate that HSUR 1 and HSUR 2 are not required for immortalization of common marmoset lymphocytes by herpesvirus saimiri. In contrast, deletions that interrupted the STP open reading frame yielded virus that was no longer able to immortalize. These include not only large deletions (11att, S4, and $\Delta 2B^*$) but also smaller deletions confined to the STP region ($\Delta 1A$ and $\Delta 1B$).

We sought further evidence for the involvement of STP in transformation by attempting to convert a nontransforming strain into a transforming one by adding back the necessary sequences. For this, we inserted 508 bp into the 3,818-bp S4 deletion mutant to create deletion mutant MR; this addition was just sufficient to reopen the STP reading frame (Fig. 3).

Addition of these sequences restored the ability of virus to immortalize lymphocytes. The growth phenotype of cells immortalized with the MR strain was slow, similar to that of $\Delta 2A$

STP. The STP open reading frame is 168 amino acids, 164 amino acids from the first methionine (Fig. 5). The protein predicted by the STP open reading frame has a highly hydrophobic stretch of 26 amino acids near its carboxy terminus (Fig. 5 and 6). This stretch is immediately preceded by the sequence His X_2 His X_6 Cys X_2 Cys, where Xs are other amino acids. The amino terminus is acidic and hydrophilic. No sites for N-linked glycosylation were found.

DISCUSSION

A number of unique features are peculiar to this 5,549-bp region of the herpesvirus saimiri genome. The gene for DHFR is the first such gene to be identified in a mammalian virus (32). The herpesvirus saimiri DHFR gene is highly conserved as compared with its mammalian cell counterparts. It is devoid of intron sequences, suggesting that herpesvirus saimiri DHFR was acquired by some process involving reverse transcription. This region also contains genes for the first virus-encoded U RNAs to be identified (22, 27, 35). Although clear similarity exists among all five HSURs, they are distinct from any previously characterized cellular U RNAs. The organization of the HSUR genes is also unusual in that they are interspersed within the tran-



FIG. 6. Hydropathy plot of STP.

scriptional units containing the DHFR and STP open reading frames. In mammalian cells, genes for U RNAs can be clustered (6), but we are not aware of any evidence for poly(A)⁺ RNA transcripts overlapping those for U RNAs. Synthesis of STP from the two abundant poly(A)⁺ RNAs found in infected cells would require translation either from a bicistronic mRNA (4.9 kb) or from an mRNA (2.3 kb) with a long 5' untranslated sequence. Long 5' untranslated sequences are most commonly found among the mRNAs of cellular oncogenes (21). Bicistronic mRNAs are rare in mammalian cells, but of the few known examples, one is from EBV (34).

Five separate deletions that interrupted or eliminated the STP reading frame yielded virus not able to immortalize marmoset T lymphocytes. Addition of 508 bp to a nontransforming 3,818-bp deletion mutant reopened the STP reading frame and restored the ability of virus to immortalize T cells. Our results provide strong evidence for the requirement of the STP open reading frame for transformation; it is not,

ever, required for replication of the virus. The slow growth phenotype of the $\Delta 2A$ and MR constructed viruses could possibly be explained by the lack of transcriptional control signals for the 2.3-kb RNA in these mutants and inefficient expression of STP from the bicistronic 4.9-kb RNA.

Gompels et al. (14) have shown that open reading frames at 13 separate sites along the herpesvirus saimiri genome were unambiguously similar to EBV gene products. These sequences were more closely related to EBV gene products than to those of other prototypic herpesviruses. Furthermore, the organization, order, and orientation of these 13 herpesvirus saimiri sequences resembled those of their homologs on the EBV genome. The STP open reading frame of herpesvirus saimiri is located in a region of the genome that corresponds to the location of LMP in EBV. Although it is possible that STP may be the functional equivalent to LMP, STP is not related or similar in sequence to LMP. In fact, STP, DHFR, and the five HSURs all appear to have no homologs in EBV.

The unique existence of DHFR and U RNA genes in herpesvirus saimiri suggests the possibility that this whole region of the genome may have been captured from host cell sequences. STP is not significantly related to sequences available in GenBank or NBRF. We have not been able to detect hybridization of STP sequences to New World primate cellular DNA at moderate stringency. If STP does represent a captured cellular gene, either the cellular homolog has not been sequenced yet or the viral gene has diverged to the extent that similarity is no longer recognizable.

The role of STP in immortalization can only be speculated upon at this time. Most but not all membrane-spanning proteins are made with an amino-terminal signal sequence that is cleaved upon insertion of the protein into the lipid bilayer. Although STP has a predicted membrane-spanning domain near its carboxy terminus, it has no signal peptide at its amino terminus. This is similar in some respects to the polyomavirus middle T antigen, which has no signal sequence at its amino terminus, yet its association with cellular membranes is believed to occur via a distinct, predominantly hydrophobic 22-amino-acid domain near the carboxy terminus (24). STP also has a predicted metal-binding domain similar to the consensus sequence of a zinc finger (3, 11, 13). However, several aspects of the sequence suggest to us that it may not be a classic zinc finger involved with DNA binding and transcription regulation. Zinc finger proteins such as TFIII A. SP1, and steroid hormone receptors (16, 17, 26) possess not only two cysteines and two histidines with the appropriate spacing, they also contain a specific phenylalanine and aromatic residues; STP is lacking these latter characteristics. Furthermore, the transcriptional regulatory proteins described above have multiple zinc fingers that chelate zinc and bind DNA; STP has only one putative metal-binding domain. The E1A protein of adenovirus does have a single zinc finger motif that binds zinc in a region that is important for transactivation (5). Further studies will need to focus on amino acids essential for STP transforming activity, on the localization and biochemical activities of STP, and on the mechanisms by which STP brings about altered T-cell growth.

ACKNOWLEDGMENTS

This work was supported by Public Health Service grants 31363, from the National Cancer Institute, and RR00168, from the Division of Research Resources.

We greatly appreciate the technical assistance of Anke Bakker and Daniel Silva. We also thank Joan Steitz, Susanna Lee, and Beverly Blake for critical reading and Nancy Adams and Joanne Newton for preparation of the manuscript.

LITERATURE CITED

- Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. A. Smith, J. G. Seidman, and K. Struhl (ed.). 1987. Current protocols in molecular biology. John Wiley & Sons, Inc., New York.
- Bankier, A. T., W. Dietrich, R. Baer, B. G. Barrel, F. Colbère-Garapin, B. Fleckenstein, and W. Bodemer. 1985. Terminal repetitive sequences in herpesvirus saimiri virion DNA. J. Virol. 55:133–139.
- Berg, J. M. 1986. Potential metal-binding domains in nucleic acid binding proteins. Science 232:485–487.
- 4. Bornkamm, G., H. Delius, B. Fleckenstein, F.-J. Werner, and C. Mulder. 1976. Structure of *Herpesvirus saimiri* genomes: arrangement of heavy and light sequences in the M genome. J. Virol. 19:154-161.
- Culp, J. S., L. C. Webster, D. J. Friedman, C. L. Smith, W.-J. Huang, F. Y.-H. Wu, M. Rosenberg, and R. P. Ricciardi. 1988. The 289-amino acid E1A protein of adenovirus binds zinc in a region that is important for trans-activation. Proc. Natl. Acad. Sci. USA 85:6450-6454.
- 6. Dahlberg, J. E., and E. Lund. 1988. The genes and transcription of the major small nuclear RNAs, p. 38–71. *In* M. L. Birnstiel (ed.), Small nuclear ribonucleoprotein particles. Springer-Verlag, New York.
- 7. Dambaugh, T., K. Hennessy, S. Fennewald, and E. Kieff. 1986. The EBV genome and its expression in latent infection, p. 13-46. In M. A. Epstein and B. G. Achong (ed.), The Epstein-Barr virus: recent advances. William Heinemann Medical Books Ltd., London.
- Desrosiers, R. C., A. Bakker, J. Kamine, L. A. Falk, R. D. Hunt, and N. W. King. 1985. A region of the *Herpesvirus saimiri* genome required for oncogenicity. Science 228:184–187.
- 9. Desrosiers, R. C., R. L. Burghoff, A. Bakker, and J. Kamine. 1984. Construction of replication-competent *Herpesvirus* saimiri deletion mutants. J. Virol. **49**:343–348.
- Desrosiers, R. C., D. P. Silva, L. M. Waldron, and N. L. Letvin. 1986. Nononcogenic deletion mutants of herpesvirus saimiri are defective for in vitro immortalization. J. Virol. 57:701-705.
- Evans, R. M., and S. M. Hollenberg. 1988. Zinc fingers: gilt by association. Cell 52:1–3.
- 12. 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.
- Frankel, A. D., and C. O. Pabo. 1988. Fingering too many proteins. Cell 53:675.
- Gompels, U. A., M. A. Craxton, and R. W. Honess. 1988. Conservation of gene organization in the lymphotropic herpesviruses herpesvirus saimiri and Epstein-Barr virus. J. Virol. 62:757-767.
- 15. Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28: 351–359.
- Hollenberg, S. M., V. Giguere, P. Segui, and R. M. Evans. 1987. Colocalization of DNA-binding and transcriptional activation functions in the human glucocorticoid receptor. Cell 49:39–46.
- Kadonaga, J. T., K. R. Carner, F. R. Masiarz, and R. Tjian. 1987. Isolation of cDNA encoding transcription factor SP1 and functional analysis of the DNA binding domain. Cell 51:1079–

1090.

- Kamine, J., A. Bakker, and R. C. Desrosiers. 1984. Mapping of RNA transcribed from a region of the *Herpesvirus saimiri* genome required for oncogenicity. J. Virol. 52:532–540.
- Kiyotaki, M., R. C. Desrosiers, and N. L. Letvin. 1986. Herpesvirus saimiri strain 11 immortalizes a restricted marmoset T8 lymphocyte subpopulation in vitro. J. Exp. Med. 164:92–93.
- Koomey, J. M., C. Mulder, R. L. Burghoff, B. Fleckenstein, and R. C. Desrosiers. 1984. Deletion of DNA sequences in a nononcogenic variant of *Herpesvirus saimiri*. J. Virol. 50:662-665.
- Kozak, M. 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. 15:8125– 8148.
- Lee, S. I., S. C. S. Murthy, J. J. Trimble, R. C. Desrosiers, and J. A. Steitz. 1988. Four novel U RNAs are encoded by a herpesvirus. Cell 54:599–607.
- 23. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Markland, W., S. H. Cheng, B. A. Oostra, and A. E. Smith. 1986. In vitro mutagenesis of the putative membrane-binding domain of polyomavirus middle-t antigen. J. Virol. 59:82–89.
- 25. Melendez, L. V., M. D. Daniel, R. D. Hunt, and F. G. Garcia. 1968. An apparently new herpesvirus from primary kidney cultures of the squirrel monkey (*Saimiri sciureus*). Lab. Anim. Care 18:374–381.
- Miller, J., A. D. McLachlan, and A. Klug. 1985. Repetitive zincbinding domains in the protein transcription factor IIIA from *Xenopus* oocytes. EMBO J. 4:1609–1614.
- Murthy, S., J. Kamine, and R. C. Desrosiers. 1986. Viralencoded small RNAs in herpes virus saimiri induced tumors. EMBO J. 5:1625-1632.
- Neubauer, R. H., C. J. Briggs, K. B. Noer, and H. Rabin. 1983. Identification of normal and transformed lymphocyte subsets of nonhuman primates with monoclonal antibodies to human lymphocytes. J. Immunol. 130:1323–1329.
- Roizman, B., L. E. Carmichael, F. Deinhardt, G. de-The, A. J. Nahmias, W. Plowright, F. Rapp, P. Sheldrick, M. Takahashi, and K. Wolf. 1981. Herpesviridae. Intervirology 16:201-217.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.
- Schaffer, P. A., L. A. Falk, and F. Deinhardt. 1975. Brief communication: attenuation of *Herpesvirus saimiri* for marmosets after successive passage in cell culture at 39°C. J. Natl. Cancer Inst. 55:1243-1246.
- Trimble, J. J., S. C. S. Murthy, A. Bakker, R. Grassmann, and R. C. Desrosiers. 1988. A gene for dihydrofolate reductase in a herpesvirus. Science 239:1145–1147.
- Wang, D., D. Liebowitz, and E. Kieff. 1985. An EBV membrane protein expressed in immortalized lymphocytes transforms extablished rodent cells. Cell 43:831-840.
- 34. Wang, F., L. Petti, D. Braun, S. Seung, and E. Kieff. 1987. A bicistronic Epstein-Barr virus mRNA encodes two nuclear proteins in latently infected, growth-transformed lymphocytes. J. Virol. 61:945-954.
- Wassarman, D. A., S. I. Lee, and J. A. Steitz. 1989. Nucleotide sequence of HSUR 5 RNA from herpesvirus saimiri. Nucleic Acids Res. 17:1258.
- Wright, J., L. A. Falk, D. Collins, and F. Deinhardt. 1976. Brief communication: mononuclear cell fraction carrying Herpesvirus saimiri in persistently infected squirrel monkeys. J. Natl. Cancer Inst. 57:959–962.