Expression of collagenlike sequences by a tumor virus, herpesvirus saimiri

Peter Geck
University of Massachusetts Medical School

Et al.

Let us know how access to this document benefits you.
Follow this and additional works at: https://escholarship.umassmed.edu/oapubs

Part of the Life Sciences Commons, and the Medicine and Health Sciences Commons

Repository Citation

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in Open Access Articles by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.
Expression of Collagenlike Sequences by a Tumor Virus, Herpesvirus Saimiri

PETER GECK,1++ SCOTT A. WHITAKER,1† MARIA M. MEDVECZKY,1† AND PETER G. MEDVECZKY1,2†

Departments of Pharmacology1 and Molecular Genetics and Microbiology,2 University of Massachusetts Medical School, Worcester, Massachusetts 01655

Received 20 December 1989/Accepted 2 April 1990

Sequencing demonstrates that the oncogenic regions of a group A strain and a group C strain of herpesvirus saimiri are nonhomologous. A bicistronic viral mRNA from this region is transcribed in tumor cells transformed by a highly oncogenic group C virus. The first open reading frame is homologous to collagen; no such sequences were found in group A or B strains. This is the first report that a virus encodes for sequences similar to that of a connective tissue protein.

Unlike alphaherpesvirus (e.g., herpes simplex virus) and betaherpesvirus (e.g., cytomegaloviruses), members of the gammaherpesvirus subfamily (Epstein-Barr virus, herpesvirus saimiri, herpesvirus atelines, and herpesvirus syilvagius) are able to induce lymphoproliferative diseases (39, 40). Herpesvirus saimiri is used as a model for studying human lymphoproliferative malignancies, since its genetic structure and molecular biology are related to those of the human Epstein-Barr virus (14). Herpesvirus saimiri shows high oncogenicity in New World monkeys (13) and New Zealand White rabbits (9, 30). The transformed cells of the acute lymphomas and leukemias induced by the virus are CD8-positive T lymphocytes with NK activity (21, 24) and carry a circular form of the originally linear virion DNA with 110 kilobase pairs (kbp) of unique sequences (L-DNA, 36% G+C) flanked by tandem repeats (H-DNA, 71% G+C) (3).

The rightmost L-DNA region (polarity analogous to that of Epstein-Barr virus DNA) is thought to contain the oncogene(s). This hypothesis is based on four lines of evidence: (i) the region is always retained in immortalized cell lines (23, 44), (ii) deletions in this region result in loss of oncogenicity without any effect on virus replication (11, 25), (iii) variability of this sequence forms three homology groups (A, B, and C) with different oncogenic potencies (30, 32), and (iv) in intergroup recombinants the characteristic oncogenic trait is conveyed by this region (30). In strain 11 of group A, unique features established so far in this region involve five genes for U-type small RNAs (27, 51), a dihydrofolate reductase gene (50), and an open reading frame with no homology to known sequences but with a presumed function in immortalization (35), as shown by deletion mapping. Only the U-type small RNAs are transcribed, and no corresponding mRNAs have been detected in strain 11-transformed cells.

Efforts in this laboratory were focused on the rightmost L-DNA region of group C strain 484-77, which is highly oncogenic in rabbits. Several lymphoblastoid tumor-derived cell lines have been established from thymuses and spleens of tumor-bearing animals (30) and cultured in RPMI medium supplemented with 20% fetal calf serum and 10 μg of gentamycin sulfate per ml. The cell line used in these experiments (484-77T) had been continuously cultured for approximately 1 year prior to RNA isolation and was shown to carry no linear viral genomes but approximately 50 copies of circular viral genomes per cell. No virus production was detected by cocultivation on owl monkey kidney cells, so lytic-cycle genes are probably not expressed.

Cloning the 6.4-kbp 484-77 right-hand terminal L-DNA. The 6.4-kbp clone (containing 6.2 kbp of L-DNA) was generated by PstI digestion of a clone consisting of the rightmost terminal 14-kbp L-DNA in a 21-kbp insert of 484-77 cloned into the vector pHyg, which confers resistance to hygromycin B (30, 47). The original 21-kbp 484-77 clone was generated by Sau3AI partial digestion of the full-length 484-77 viral genome followed by sucrose gradient size fractionation and cloning into the pHyg vector (14). The 6.4-kbp fragment was subcloned into the PstI site of the Bluescript M13-SK vector (Stratagene). A set of small subclones was then created by taking advantage of the mapped restriction endonuclease sites, and a library of clones with "nested deletions" was also generated by standard protocols (18, 19) (Fig. 1).

A virus-specific mRNA transcribed in tumor cell lines is mapped in a region of the genome involved in oncogenicity. Cytoplasmic RNA was isolated by Nonidet P-40 lysis, as described previously (12), from the cell line (484T) derived from strain 484-77-induced tumor, from the strain 11-transformed cell line (1670), and from uninfected lymphocytes. The preparations were poly(A) selected and probed on Northern (RNA) blots by using standard protocols (31, 38, 49) with the 6.4-kbp sequence containing the 6.2-kbp rightmost L-DNA region. The probe did not hybridize to strain 11-specific or uninfected cytoplasmic RNAs, but in the cells derived from 484-77-induced tumor, a 1.2-kilobase (kb) RNA with poly(A) tail was detected (Fig. 2) (four small RNAs in total unselected RNA preparations were also detected; data not shown).

Mapping and analysis of the mRNA was carried out by RNase protection assays and primer extension. Nearly 50 overlapping subclones were used for strand-specific RNA

* Corresponding author.
† Present address: Department of Medical Microbiology and Immunology, College of Medicine, University of South Florida, Box 10, Tampa, FL 33612-4799.
endonuclease sites used for probe synthesis by established protocols (33). The highly labeled, single-stranded RNA probes were hybridized in solution to the RNA preparations from 484T cells. RNase treatment by the method of Zinn et al. (55) reduced the sizes of the probes, and only those fragments of the probes which became protected by the mRNA remained intact. The total region of protection finally identified the mRNA-coding DNA sequences. The results of such RNase protection assays with some of the subclones are shown in Fig. 3. The data show that the transcribed region is an intronless sequence between nucleotide positions 1336 and 140 (position +1 is the first nucleotide leftward from the last intact repetitive unit of the right-hand H-DNA; see Fig. 5). Rightward transcription was indicated by the strandedness of protection assays.

The exact 5' end was determined by primer extension analysis by standard protocols (22, 29). An oligonucleotide primer complementary to a sequence close to the 5' end of the mRNA (as estimated by preliminary data) between map positions 1248 and 1268 (5' CCGTTTCTTCAGTGGAT3') was synthesized, labeled with [γ-32P]ATP by T4 polynucleotide kinase, hybridized at 22°C to 4 µg of oligo(dT)-selected RNA, and extended by reverse transcriptase toward the 5' end of the mRNA to serve as template. Electrophoresis of the extended primer with a sequencing marker identified the number of incorporated nucleotides (Fig. 4). The data demonstrate that the primer was extended with 68 nucleotides from the 3'-terminal nucleotide of the primer at map position 1268. In addition to the major band at 68 nucleotides, minor bands at 67 and 66 nucleotides refer to the irregular 5' end of the mRNA (5' cap structure), where the reverse transcriptase occasionally preterminates.

Nucleotide sequence and predicted amino acid sequence of ORF1 of the 1.2-kb mRNA are homologous to collagen. The gene and flanking regions of the 1.2-kb mRNA were sequenced by the dideoxynucleotide-chain termination method by standard protocols (43, 48) and analyzed with the DNASTAR computer program (DNASTAR, Inc., Madison, Wis.) (Fig. 5). The established map positions of the mRNA were confirmed by the consensus sequences for transcription regulation. A typical TATA box was found 28 nucleotides upstream of the first nucleotide of the mRNA, at position 1335; a primary-transcript terminal cleavage signal

---

**FIG. 1.** Map of subclones of the rightmost 2-kbp L-DNA region. Exact positions (Map position) of the subclones and restriction endonuclease sites used for subcloning are indicated. Two sets of clones created by nested deletions are shown. One set (pND1.0 to pND0.5) was obtained by unidirectional deletions only, and the other (pHBND1.9 to pHBND1.3) had the BamHI-PstI fragment deleted also.

**FIG. 2.** Northern blot hybridization of poly(A)+ RNAs from uninfected and tumor-derived cell lines. RNA (4 μg) was applied in each lane for formaldehyde–1% agarose gel electrophoresis. Numbers at left indicate sizes in kilobase pairs. Lane M, RNA size marker (six 32P-labeled synthetic poly(A)-tailed RNAs [Bethesda Research Laboratories, Gaithersburg, Md.]); lane 11, cytoplasmic poly(A)-selected RNA of a group A strain (strain 11)-transformed monkey cell line (1670); lane 484, cytoplasmic, poly(A)-selected RNA of a group C strain (strain 484-77)-transformed, thymus-derived, rabbit tumor cell line (484T); lane Uninfected, cytoplasmic, poly(A)-selected RNA of an uninfected human lymphoblastoid cell line (Jurkat). The nick-translation DNA probe represented cloned rightmost L-DNA sequences between nucleotides 1 and 6200, with a portion of the adjoining H-DNA (−1 to −254).
TGTGTGTG sequence was localized at the 3' end of the mRNA; and a poly(A) addition signal AATAAAA sequence was identified 40 nucleotides upstream. Two major open reading frames (ORF1 and ORF2) were revealed. ORF1 codes for 99 amino acids between nucleotides 1290 and 994 (Fig. 2). The coding capacity of ORF2 is 214 amino acids from nucleotides 812 to 171. The latter is not discussed further, since no significant homology was found in the GenBank data.

In the sequence of ORF1, however, a remarkable structure has been revealed. Of the total 297 nucleotides, 162 code for a repetitive sequence with an element of 18 nucleotides repeated in tandem nine times, with minor irregularities in the first and last units (Fig. 5A). High similarity was found between ORF1 and different forms of collagen of diverse origin (not shown). The greatest similarity to ORF1 was displayed by a collagen sequence of a nematode, *Caenorhabditis elegans* (26), for which 116 of 162 nucleotides in the repetitive region were identical (71.6%); the flanking nonrepetitive sequences of ORF1 also showed similarity to collagen (Fig. 5B). High homology was demonstrated when amino acid sequences of various known collagens and the putative herpesvirus saimiri protein sequence were compared (Fig. 6).

**Collagenlike sequences are present in group C strains of herpesvirus saimiri and absent in group A and B strains.** To obtain virion DNA for Southern hybridization, OMK cells

**FIG. 3.** Mapping of mRNA by RNase protection assay. (A) Autoradiogram from a denaturing 3% polyacrylamide gel electrophoresis displaying the larger protected fragments. Numbers at left indicate lengths in nucleotides. Lane M, RNA size marker synthesized by T7 RNA polymerase on templates of determined size; lanes 1 through 5, protected RNA fragments (map positions of probes and protected fragments are depicted in panel C). (B) Autoradiogram from a denaturing 6% polyacrylamide gel electrophoresis displaying the smaller protected fragments. Lane M, DNA size marker (pUC19 *HpaII* digestion, labeled with [32P]dCTP by using the Klenow fragment of *Escherichia coli* DNA polymerase); lanes 1 and 2, protected RNA fragments (map positions of probes and protected fragments are depicted in panel C). (C) Map of probes and protected fragments. Numbers on the map indicate kilobase pairs. The exact map positions of probes (—) and protected fragments (—) are also shown, with references to the corresponding lanes in panels A and B.

**FIG. 4.** Primer extension analysis of the 5' end of the mRNA. Autoradiogram from a denaturing 6% polyacrylamide gel electrophoresis showing the extended primer together with a size marker sequencing ladder. Only a portion of the sequencing gel is shown. The sequence itself is unrelated and is used merely as a marker. Lane Extension, Extended primer; lanes G, A, T, and C, part of a size marker sequencing ladder, shown only from nucleotide (N) 54. G, A, T, and C label the individual chain termination reactions.
were infected with the following strains representing the three subgroups: strains 484-77, 487-77, and 505-77 (all group C); strain 11 (group A); and strain S295C (group B). DNA was isolated from the virions as described previously (30). Restriction endonuclease digestions, electrophoresis, vacuum blotting, and hybridization with nick-translated DNA probe were performed by using standard protocols (31, 38, 49). The ORF1 sequence was present in the three group C strains tested, with no variability in length of the ORF1-containing fragments (Fig. 7).

As expected, no hybridization was detected in a representative from group A or B. No homology was found between ORF1 or ORF2 nucleic acid or predicted protein sequence and the published right-end sequence of strain 11 (group A) (35) with ALIGN or ALIGN programs (DNASTAR, Inc.) or FASTA programs ALIGN, RD2, or PLFASTA (37) (data not shown).

Sequencing the oncogenic DNA region of a highly oncogenic herpesvirus saimiri strain and studying its expression by transcription analysis resulted in three major and potentially important observations. (i) Earlier hybridization data that the rightmost L-DNA sequences are nonhomologous among different groups of herpesvirus saimiri have been confirmed by sequencing. (ii) This is the first report in which an mRNA from the oncogenic region in transformed cells has been detected and characterized. (iii) This is also the first
report in which a viral mRNA has been shown to have an open reading frame that codes for a collagenlike sequence. To explain the lack of homology between strains 11 and 484-77 in the right-end 1.8-kbp L-DNA region, we favor the hypothesis that different cellular sequences have been acquired by these strains. There is evidence that herpesvirus saimiri readily integrates cellular sequences into its genome. The virally coded thymidylate synthase (27, 50), and U-type small RNAs (52, 53) are highly homologous to corresponding human sequences and were probably acquired from the cellular genome. The difference in G+C content suggests that the collagenlike sequence may also be of cellular origin (52%) G+C, in contrast to 36% G+C in the surrounding viral L-DNA sequences). However, we assume that these cellular sequences then further evolved, since no homology between cellular and viral sequences can be detected by Southern hybridization (32; P. Medveczky, unpublished data). The right-end region of L-DNA of strain 11 (group A) has one open reading frame with no homology to ORF1 or ORF2 of 484-77, as shown, but deletion of the open reading frame results in loss of oncogenicity (35). These results suggest that the two viruses probably transform T cells by different molecular mechanisms.

The viral message in the tumor-derived cell line is 1.2 kb long and contains a 3' poly(A) tail, and our data indicate no splicing, unless sequences outside the studied 6.2-kbp region are involved. The mRNA sequence contains two open reading frames, which is unusual but not unprecedented among eucaryotes.

The structure of the saimiri collagenlike (SCOL) sequence is coherent with basic patterns in collagen structure (4, 53). In collagen, a nine-nucleotide element codes for Gly-X-Y amino acid triplets (X and Y are frequently proline and hydroxyproline, respectively) which are repeated several hundred times in individual polypeptides of the triple helical structure (26, 36; for a review, see reference 6). A characteristic element of collagen evolution is a 54-nucleotide unit. The relics of this evolutionary stage are still present in the exons of a chicken collagen, in which most of the 50 exons contain 54 nucleotides (6, 36). In herpesvirus saimiri, the 162-nucleotide sequence appears to be the result of exactly three reiterations of a 54-nucleotide unit. The absence of splicing may indicate a reverse transcription step prior to integration into the viral genome.

The perfect invariable repeats are one of the intriguing features of the saimiri collagenlike sequence, since in collagen the X and Y positions are highly variable. The presence of the sequence was detected without fragment variability in all tested group C strains by Southern analysis. This demonstrates the conservation of the sequence within the subgroup, which may indicate the biological significance of the saimiri collagenlike sequence.

No evidence is available on the function of the saimiri collagenlike sequence, and the expression of the polypeptide has yet to be demonstrated. It has been suggested that collagen plays a role in cellular growth regulation, usually by down regulating cell division (7, 28, 46). On the other hand, very small (15,000 to 17,000 M,) collagen polypeptides that have stimulating effects on embryonic tissues (42) and bone marrow stem cell culture (52) have been detected. Mitogenic stimulation and cell transformation after a fibronectin fragment with affinity to collagen is bound (10, 20) seem to indicate a collagenlike cell surface receptor, which may act in cell growth regulation. Integrin receptors on cytotoxic T lymphocytes (nonspecific adhesion receptors, e.g., LFA-1, LFA-3, VLA-1, and VLA-2) (5, 16, 17, 34, 41) may also be involved in a collagen-related growth regulation.

The cellular and retroviral oncogenes are derivatives of...
normal cellular counterparts with physiological functions (1, 45). These proto-oncogenes act at different stages of signal transduction in the inter- and intracellular pathways of cell growth regulation (8, 15, 54). No experimental evidence so far has indicated, however, that sequences related to connective tissue proteins are involved in the process. This report raises the possibility of a growth regulatory function for a virally encoded collagenlike polypeptide.

We thank Larry Hardy, Carel Mulder, Rosemary Vassallo, and Douglas Waud for critical reading of the manuscript. We thank Daniel Mullen for the excellent photographic work.

This work was supported by Public Health Service grants ROI CA43264, ROI CA59653, and AI 21129 from the National Institutes of Health and biomedical research support grant SOTR05712 from the National Institutes of Health to the University of Massachusetts Medical School.

LITERATURE CITED