

## Transcriptional Analysis of Latent and Inducible Kaposi's Sarcoma-Associated Herpesvirus Transcripts in the K4 to K7 Region†

Jennifer L. Taylor, Heather N. Bennett, Beth A. Snyder, Patrick S. Moore, and Yuan Chang\*

Molecular Virology Program, University of Pittsburgh Cancer Institute, University of Pittsburgh, Pittsburgh, Pennsylvania 15213

Received 22 April 2005/Accepted 23 June 2005

**Kaposi's sarcoma-associated herpesvirus (KSHV) is a gamma-2 herpesvirus with a genome containing a long unique coding region (LUR) flanked by GC-rich terminal repeat sequences. The LUR encodes approximately 90 annotated open reading frames (ORFs) with complex patterns of gene expression during viral latency, reactivation, and de novo infection. To identify unannotated KSHV genes, we examined the region between 21,500 and 30,000 bp of the KSHV LUR, representing approximately 8.5 kb of sequence. This region encodes seven known single-exon ORFs (*K4*, *K4.1*, *K4.2*, *K5*, *K6*, *K7*, and *PAN*), but previous computer analyses have failed to identify additional likely genes in the remaining 5.2 kb. We identified four novel transcripts using Northern blotting, phage library screening, and 5' rapid amplification of cDNA ends analysis in the region between ORFs *K4.2* and *K7*. In vitro analysis of KSHV-infected primary effusion lymphoma cell lines in the presence of 12-*O*-tetradecanoylphorbol-13-acetate and phosphonoformic acid suggests that one latent transcript is coterminal with the previously annotated *K3* gene encoding an ubiquitin-ligase known to downregulate major histocompatibility complex class I expression. This alternatively spliced transcript may contribute to KSHV adaptive immune evasion during latent infection. Other transcripts are inducible, including a 6.1-kb transcript that is the largest transcript found in the KSHV genome to date.**

Kaposi's sarcoma-associated herpesvirus (KSHV [or human herpesvirus 8; HHV-8]) was initially sequenced and annotated in 1996 using cosmid and phage genomic libraries from the KSHV-infected primary effusion lymphoma (PEL) cell line BC-1 (25). The KSHV genome is ~165 kb in length, with a ~145-kb long unique coding region (LUR), containing all known viral genes, flanked by a GC-rich terminal repeat sequence. Similar sequencing analyses of KSHV obtained from a KS lesion and a PEL suggest that the virus is highly conserved between isolates (17, 19), although hot spots of strain-specific variation have been found both at the ends of the LUR (11, 35) and in the glutamine-rich repeat region of *ORF73/LANA1* (34).

Initial annotation using conservative assumptions identified at least 81 potential open reading frames (ORFs) in the LUR (25). Those ORFs with sequence homology to herpesvirus saimiri (HVS) genes were given corresponding numeric ORF designations. Fifteen ORFs not homologous to HVS genes were numbered in consecutive order with a K prefix. Since the original annotation of KSHV, additional ORFs have been discovered, including *K4.1/vMIP-III* and *K4.2* (18); genes that are spliced, including *K8.1*, have contributed additional and revised ORFs (5, 22).

KSHV gene expression varies during lytic and latent virus replication. Sarid et al. (27) defined three classes of mRNAs in the KSHV-infected PEL cell line BC-1 related to expression patterns during latency and lytic replication. Class I KSHV mRNAs are

expressed during latency and not upregulated by agents such as 12-*O*-tetradecanoylphorbol-13-acetate (TPA). Class II mRNAs are detected in untreated cell cultures but are substantially increased after TPA treatment. Finally, class III mRNAs are detected only in TPA-treated cultures and correspond to "traditional" lytic gene classes. Genes induced by either sodium butyrate or TPA have also been classified by Sun et al. (30) as immediate-early, early, or late genes. Immediate-early genes are those genes that do not require de novo expression of viral transactivator proteins during lytic replication and are resistant to cycloheximide treatment. Early gene transcription is sensitive to cycloheximide but resistant to phosphonoacetic acid (PAA), a viral DNA polymerase inhibitor. Late genes are characterized by sensitivity to both cycloheximide and PAA and include many structural and replication genes. Similar analyses using DNA microarrays have allowed simultaneous comparisons of KSHV transcription patterns (6, 15, 20).

There are two KSHV genomic regions with sparse gene density. Recently, one of these regions between *ORF72/v-cyclin* and *K12/Kaposin* was found to encode 10 to 11 microRNAs that may target cellular mRNAs for degradation (3, 21). The second region, between 23.0 kb and 28.6 kb, spans an area from the end of *K4.2* to the beginning of *K7*. While two ORFs are encoded in this region, large regions of featureless but highly conserved sequence are present. To identify novel transcripts from this region, we screened cDNA libraries from PEL and KS cell lines and performed scanning Northern blot analysis. Three transcripts from this region were identified, including a large unspliced, untranslated message and a latent transcript encoding the E3-ligase MIR1 (modulator of immune recognition 1) responsible for major histocompatibility complex class I (MHC-I) downregulation in PEL cells.

\* Corresponding author. Mailing address: Hillman Cancer Center, Molecular Virology Program, University of Pittsburgh Cancer Institute, 5117 Centre Avenue, Suite 1.8, Pittsburgh, PA 15213. Phone: (412) 623-7721. Fax: (412) 623-7715. E-mail: yc70@pitt.edu.

† Supplemental material for this article may be found at <http://jvi.asm.org/>.

TABLE 1. PCR primer sequences for consecutive 300-bp products from KSHV nt 21500 to nt 30000

| Probe (KSHV nt)  | Primer sequence               |                               |
|------------------|-------------------------------|-------------------------------|
|                  | Forward                       | Reverse                       |
| A (21500–21801)  | 5'-GTATCGTTAGAGATCCGTCG-3'    | 5'-TGCTGACTGCCTTGCTTTG-3'     |
| B (21800–22098)  | 5'-CACAGCGACGAGCAGGAT-3'      | 5'-GTCTGGTTACCTTGCAAGTTA-3'   |
| C (22101–22392)  | 5'-ACATCTAGGAAGCGTAAATATC-3'  | 5'-ATTGCCACCGGATAAGCT-3'      |
| D (22393–22708)  | 5'-CTGGCACTGGACAGGGAAT-3'     | 5'-CGATGCGCAATATTTTGGG-3'     |
| E (22709–23000)  | 5'-ATAAGGTTTGCCAAAACACCCGA-3' | 5'-GGTGCTGTTAAGCGGGTTCTGC-3'  |
| F (23002–23306)  | 5'-AATCCACTTGCGGTCCAGAA-3'    | 5'-CGCCCATGAAATGGCCAAAA-3'    |
| G (23308–23611)  | 5'-CGCCATCTGAATTTGCTAATC-3'   | 5'-CAGCTGCGCTAGGATTTAAAGGA-3' |
| H (23600–23900)  | 5'-CTAGCGCAGCTGAGTCATCGCAG-3' | 5'-CCCAGTAGAGTGACCCAGCTG-3'   |
| I (23891–24196)  | 5'-CTTTATAATAGGGCCGATGAGTC-3' | 5'-TTCCCGGCGAAGGGGTGTA-3'     |
| J (24197–24498)  | 5'-CGCTATAAAAACGAGGGACAGCA-3' | 5'-CTGCCGGGCTCCCTGCGGCT-3'    |
| K (24500–24806)  | 5'-ACCCAGGAGCCCCGGCA-3'       | 5'-AGGGTGGCGCGCCGGGAT-3'      |
| L (24802–25068)  | 5'-ACCCTCCCGGAGGGGGAT-3'      | 5'-GACGCGTGGCTGTGCTTCTC-3'    |
| M (25075–25388)  | 5'-CTTTAAAAACATCGTTATCCAAT-3' | 5'-GAAAAGTGGACGGGAATAG-3'     |
| N (25389–25695)  | 5'-CCCAGCTATTCTGTCCAGCAAT-3'  | 5'-CTGCGGGGATGACGTGACGCC-3'   |
| O (25700–26000)  | 5'-CACCCAAGTGGTTGTTCAC-3'     | 5'-GACCGTACTGCCATACCGA-3'     |
| P (26000–26275)  | 5'-CGGTACCCCGGTTTATTG-3'      | 5'-GCCACAGGTTAAGGCGACTAC-3'   |
| Q (26256–26556)  | 5'-CCACAGGTTAAGGCGACTAC-3'    | 5'-GGAGACTAACTGGGCACTCT-3'    |
| R (26470–26720)  | 5'-TCCTTAGACGCCATCTGAC-3'     | 5'-GTTTACCGACAGGATGCAACG-3'   |
| S (26743–27029)  | 5'-GCTCTCCGCCTGTGTTTGA-3'     | 5'-CTTAGAAGACCTGTGAGGAAGAA-3' |
| T (27030–27320)  | 5'-TTGCGCGTCTCCTAGAAATGGC-3'  | 5'-CGCCTAATAGCTGCTGCTACGG-3'  |
| U (27330–27630)  | 5'-GTGTACGACAGGAGTACCC-3'     | 5'-TCACTTTTGTCTGCAACCCG-3'    |
| V (27630–27930)  | 5'-TTGTTGGGGCGCTTACTAG-3'     | 5'-CACAATAACAGATGACGATAC-3'   |
| W (27931–28228)  | 5'-GCTTGC CGGAGACAACATA-3'    | 5'-CTGGCGGACGTGTGTGATT-3'     |
| X (28229–28527)  | 5'-CGATTTGCAGCGCAGCTA-3'      | 5'-ACTCCATACTCTACCAAATATGG-3' |
| Y (28531–28825)  | 5'-CTTATGGATTATTAAGGGTACGT-3' | 5'-GGAAAACCTAGCCGAAAGC-3'     |
| Z (28826–29128)  | 5'-GTCCTACTTTTCCACATGG-3'     | 5'-CCGAAAACAACGATGAGCA-3'     |
| AA (29129–29432) | 5'-TTCTGTGTTTGTCTGATTCTAG-3'  | 5'-AAGAGCGTCCCAGCTG-3'        |
| BB (29432–29719) | 5'-TTCAATGTTAATGTTTAAATGTG-3' | 5'-TTATGTTGTAAGTTGCATTACG-3'  |
| CC (29725–30000) | 5'-GGTCAATGTTTAAATCCATTTCC-3' | 5'-GACGGTGGGGCAAATGT-3'       |

## MATERIALS AND METHODS

**Cell culture.** BC-1 (PEL-derived B-cell line coinfecting with KSHV and Epstein-Barr virus [EBV]) (4), BCP-1 (PEL-derived B-cell line infected with KSHV alone) (10), and BJAB (KSHV- and EBV-negative) cell lines were maintained in RPMI 1640 medium (Cellgro, Herndon, VA) supplemented with 10% (BC-1 and BJAB) or 20% (BCP-1) fetal bovine serum (Gibco-BRL, Grand Island, NY) containing penicillin-streptomycin (100 U/ml), L-glutamine (2 mM), sodium pyruvate (1 mM), HEPES (10 mM), and glucose (4.5 g/liter) at 37°C in the presence of 5% CO<sub>2</sub>. To induce lytic gene transcription, cells were treated with 20 ng/ml of TPA (Sigma Chemical Co., St. Louis, MO) and harvested after 48 h. To inhibit virus DNA replication, phosphonoformic acid (PFA) (Sigma) at a concentration of 0.5 mM was added alone or in the presence of 20 ng/ml TPA for 48 h.

**PCR.** PCR primers were designed to yield consecutive 300-bp products from nucleotide (nt) 21500 to 30000 of KSHV (Table 1). Additional primers were designed to amplify part of *ORF3* (nt 19258 to 19567) and *ORF70* (nt 20094 to 20301) along with a 70-bp oligonucleotide representing 70 bp within probe H (nt 23770 to 23840). Primers were as follows: K3(S), 5'-CGTGATCTATCTCGGGTCATG-3'; K3(AS), 5'-GAGCTCGGAAATGAGAGATTAG-3'; ORF70(S), 5'-TACTGC CATTTCATACGAATGG-3'; and ORF70(AS), 5'-ACCCGGGGAGTTTAT TCACAC-3'. All primers were synthesized by QIAGEN (Alameda, CA).

**RNA preparation and Northern analysis.** Total RNA from unstimulated and TPA-treated BC-1 cells was extracted by TRIzol reagent (Invitrogen) according to the manufacturer's instructions. Poly(A)<sup>+</sup> mRNA was isolated with a NucleoTrap mRNA purification kit (BD Biosciences Clontech, Palo Alto, CA). Northern blotting was performed on 1 µg mRNA per lane by using a 1% formaldehyde-agarose gel and transfer onto nylon membranes (GeneScreen Plus, NEN Research Products, Boston, MA). DNA probes were synthesized by PCR, TA cloned into pCR2.1-TOPO (Invitrogen), sequence confirmed, and then <sup>32</sup>P labeled by random priming (RediPrime DNA labeling system, Amersham, Buckinghamshire, England). Hybridization was performed at 42°C in 3× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–50% formamide–5× Denhardt's solution–5% dextran sulfate–0.1 M Tris (pH 7.4)–100 mg/ml of sonicated salmon sperm DNA (Stratagene, La Jolla, CA). β-Actin was used to standardize loading. pCR2.1 vectors containing the 300-bp DNA probes in both orientations were used to synthesize riboprobes using a T7 Riboprobe system (Promega) and [ $\alpha$ -<sup>32</sup>P]rCTP (Amersham).

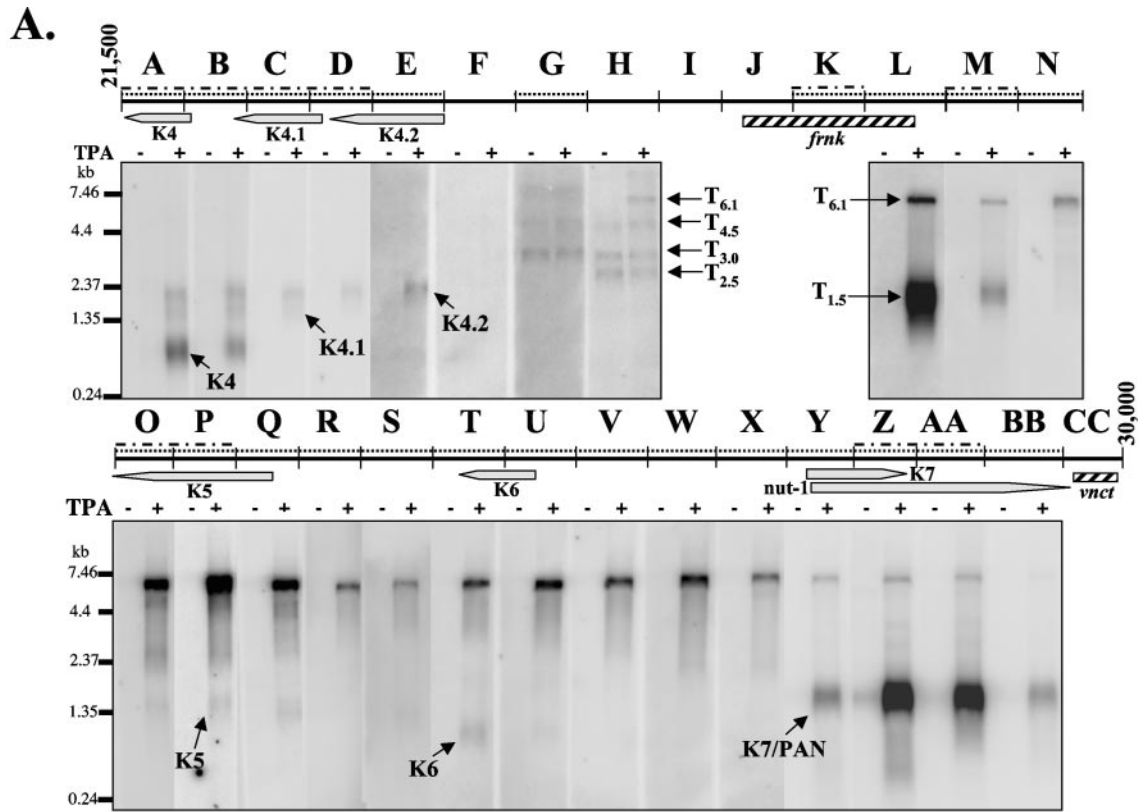
**Lambda phage screening.** A cDNA phage library of TPA-treated BC-1 cell line, one from TPA-treated BCP-1 cell line, and a third generated from a KS lesion were constructed as per the manufacturer's protocol in a ZAP Express™ vector (Stratagene, La Jolla, CA) (28). Clones were plaque purified. Positive phages were converted into phagemids by employing an ExAssist helper phage (Stratagene), and inserts were sequenced by automated DNA sequencing (ABI 377; Perkin-Elmer, Foster City, CA) with T7 and M13 reverse primers.

**5' RACE.** 5' rapid amplification of cDNA ends (5' RACE) was performed by using a GeneRacer kit (Invitrogen, Carlsbad, CA). RACE cDNA was produced from TPA-treated BC-1 mRNA. GeneRacer 5' primer and GeneRacer 5' nested primer were used in conjunction with the following transcript-specific un-nested (U) and nested (N) primers: K3(U).nt19580, 5'-CCAGCAGACGGAACATCCTC ATCT-3'; ORF70(N).nt20154, 5'-AGGTGTAACCTTCCATGGAAGAA-3'; H(N).nt23797, 5'-CCAGCGGACTACCCCACT GTAT-3'; H(U).nt23820, 5'-ATACAGTTGGGGGTAGTCCGCTGG-3'; and H(N).nt23758, 5'-GGAAAT ACCACCTGAGGACCCCGT-3'. The last two primers, H(U).nt23820 and H(N).nt23758, were for identification of the 5' end of the 6.1-kb transcript.

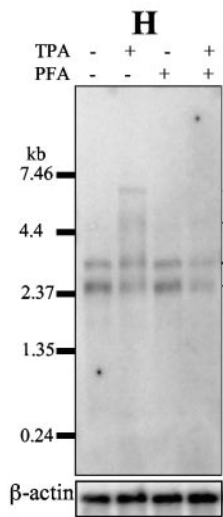
**In vitro translation.** The *K3* gene was subcloned into pCR-Script SK(+) (Stratagene) by PCR amplification using K03F1 and K03R1 primers, resulting in the plasmid P140 (K03F1.nt18568, 5'-GGTAAACACCACCAACCACAG-3'; K03R1.nt19627, 5'-TGGGTCTCTCTCAGGATATGGAAG-3'). In vitro transcription and translation were carried out with T7 polymerase for the *ORF3* plasmid (P140) and T3 polymerase for the H phage plasmid (P1520) using a T7/T3 in vitro transcription and translation kit (Promega) along with [<sup>35</sup>S]methionine to radiolabel the in vitro-synthesized proteins according to the manufacturer's protocol. Proteins from each sample were then electrophoresed in a 10% sodium dodecyl sulfate-polyacrylamide gel, dried, and exposed to film.

## RESULTS

**Transcript screening of the 21.5- to 30.0-kb LUR sequence.** Scanning RT-PCR analysis was performed using 29 primer pairs spanning the 21.5- to 30.0-kb region to amplify 300-bp products on PEL cell and KS lesion cDNA libraries. For purposes of this analysis, the 29 consecutive regions have been



**B.**



**C.**

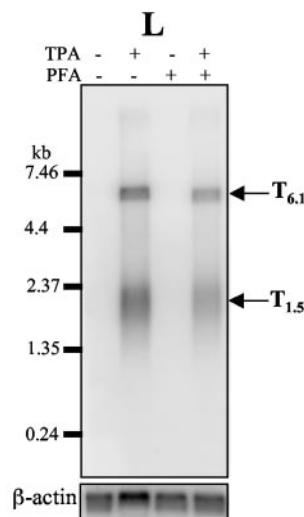


FIG. 1. PCR and Northern blot analyses identify previously unannotated transcripts. (A) Map of KSHV genome nt 21500 to 30000 with letter designations (A to CC) for consecutive 300-bp products. Positive PCR amplifications from a BCP-1 TPA-stimulated cell library (dotted line) and a KS tumor cDNA library (dotted-dashed line) are shown and were used as probes for Northern blot analyses. The remaining probes were generated from PCR products using genomic DNA, and primer pairs that did not amplify with either cDNA library. Northern blot analyses were performed using poly(A) RNAs isolated from BC-1 cells that were untreated or treated with TPA as indicated. Previously annotated viral genes with their orientation along with *frnk* and *vnc1* direct repeat regions are noted. (B) mRNA from BC-1 cells untreated or treated with TPA, PFA, or a combination of TPA and PFA using Probe H to determine directionality and expression pattern. (C) Expression pattern for T<sub>6.1</sub> and T<sub>1.5</sub> transcripts using Probe L with mRNA from BC-1 cells untreated or treated with TPA, PFA, or a combination of TPA and PFA.

labeled A through CC (Fig. 1A). Using a TPA-treated BCP-1 cell line cDNA library, transcripts corresponding to all known annotated genes in this area (*K4*, *K4.1*, *K4.2*, *K5*, *K6*, *K7*, and *PAN*) were found (dotted lines). A similar analysis using a KS cDNA library amplified only a subset of these genes, including *K4*, *K4.1*, *K5*, *K7*, and *PAN* (dotted-dashed lines). PCR products were also amplified from areas between annotated ORFs by using both libraries, suggesting that additional, unannotated transcripts exist in this genomic block. Additional PCR prod-

ucts amplified from the TPA-treated BCP-1 cDNA library (dotted lines) include nt 23300 to 23600 (G), nt 24500 to 25695 (K, L, M, and N), nt 26470 to 27029 (R and S), and nt 27330 to 28527 (U, V, W and X). Similarly, PCR products were amplified from the KS cDNA library (dotted-dashed lines) at nt 24500 to 24806 (K) and nt 25075 to 25388 (M). These PCR products were used as probes in Northern blot analyses. The remaining probes were generated from PCR products using genomic DNA and primer pairs that did not amplify with

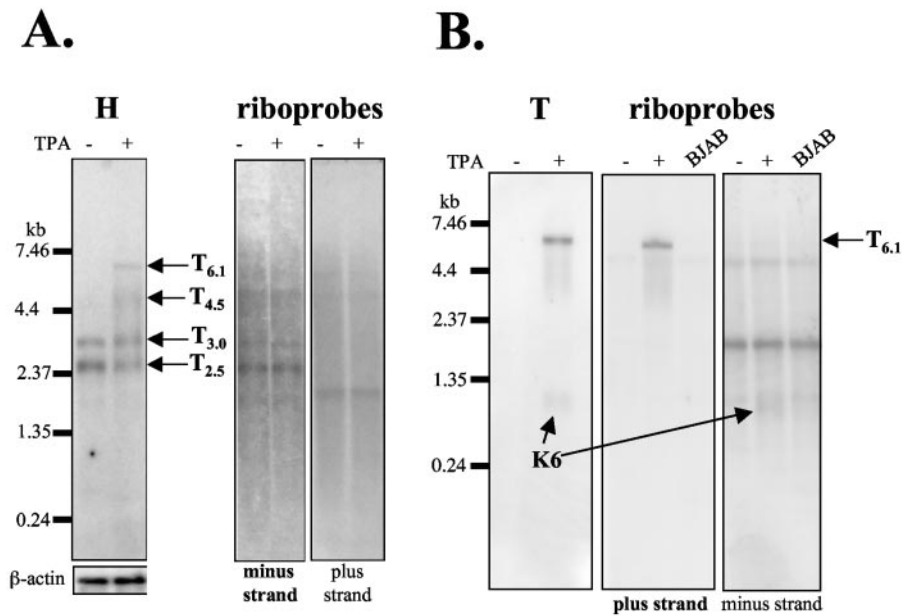


FIG. 2. Directionality of transcripts determined by riboprobe Northern blotting. (A) mRNA from BC-1 cells untreated or treated with TPA with riboprobes from Probe H to determine directionality of transcripts. (B) BC-1 mRNA from untreated or TPA-treated cells with riboprobes from Probe T to determine directionality of T<sub>6.1</sub> transcript.

either cDNA library. PCR products were not obtained from the areas designated I (nt 23891 to 24196), J (nt 24197 to 24498), K (nt 24500 to 24806), and CC (nt 29725 to 30000), even with the addition of betaine and dimethyl sulfoxide to the PCR mixture, due to the presence of *frnk* (20- and 30-bp repeats in the region from nt 24285 to 24902) or *vnc1* (13-bp repeat between bases 29775 and 29942). *Frnk* and *vnc1* are direct repeat regions of KSHV named by the original authors of the genome annotation after the American composer Frank Vincent Zappa (25).

To confirm the RT-PCR analysis, Northern blotting with each amplifiable product on BC-1 cell mRNA was performed (Fig. 1A). Appropriate size transcripts were identified from *K4*, *K4.1*, *K4.2*, *K5*, *K6*, *K7*, and *PAN*, which are induced by TPA treatment. In addition to these known ORFs and transcribed sequences, five novel transcripts were found, covering the previously unannotated intergenic regions. Two transcripts, designated by size in kilobases as T<sub>2.5</sub> and T<sub>3.0</sub>, hybridized with regions G and H (nt 23308 to 23900). These transcripts are noteworthy since they are not induced by TPA treatment (class I or latent transcripts) and show no inhibition by PFA (Fig. 1B). A third transcript, T<sub>4.5</sub>, also hybridized with regions G and H and shows induction by TPA (class II transcript). The fourth larger transcript, T<sub>6.1</sub>, is inducible by TPA and hybridized with regions extending from H to BB (nt 23600 to 29700). Finally, a small transcript, T<sub>1.5</sub>, overlapping with T<sub>6.1</sub> in the L and M regions (nt 24802 to 25388) was similarly induced by TPA treatment. In contrast to the two class I and one class II transcripts found, T<sub>6.1</sub> and T<sub>1.5</sub> are expressed only after TPA treatment (class III transcripts) with T<sub>6.1</sub> sensitive to PFA treatment, consistent with late lytic transcription (Fig. 1B), and with T<sub>1.5</sub> resistant to PFA inhibition, consistent with early lytic transcription (Fig. 1C).

Directionality for these transcripts was examined using riboprobe hybridization (Fig. 2). Using H region (nt 23600 to 23900) riboprobes, T<sub>2.5</sub> and T<sub>3.0</sub> are transcribed in the minus orientation, but the directionality of T<sub>4.5</sub> and T<sub>6.1</sub> was not determined due to nonspecific hybridization (Fig. 2A). To determine transcript direction for T<sub>6.1</sub>, riboprobes from the T region (nt 27030 to 27320) were examined and showed that T<sub>6.1</sub> is transcribed in the plus orientation, the orientation opposite that of the overlapping *ORFK6* transcript (0.8 kb) in the minus orientation (Fig. 2B). The band detected at ~2.2 kb is nonspecific and can also be seen when the KSHV-negative B cell line BJAB is used.

**Transcript mapping.** To accurately determine the size and orientation of these transcripts, a TPA-treated BC-1 lambda phage library was screened with probes from the L (nt 24802 to

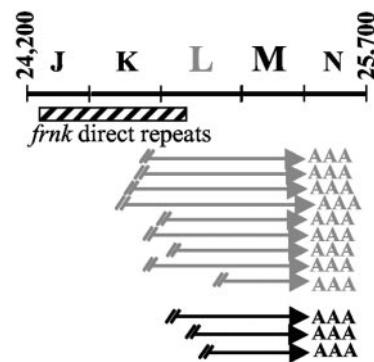


FIG. 3. Lambda phage screening with probes L and M for identification of T<sub>1.5</sub>. Drawing depicts nine distinct clones from phage screening with Probe L (gray) and three clones from screening with Probe M (black), all with poly(A) tails at nt 25440.

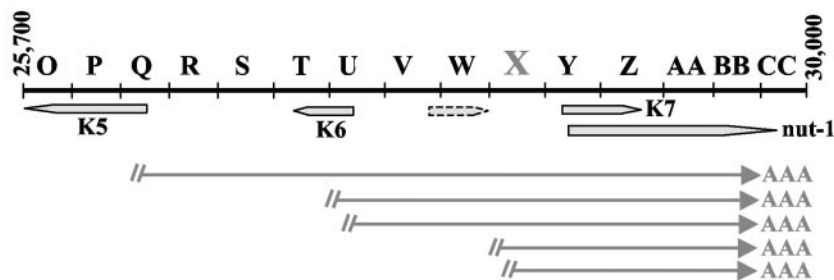


FIG. 4. Lambda phage screening with probe X for identification of  $T_{6.1}$ . Schematic aligns the five clones identified from phage screening with Probe X. All clones were coterminal with the *PAN* transcript, with poly(A) tails at nt 29741. Analysis using the MacVector program (Accelrys, Inc.) found one possible ORF (nt 27887 to 28192) on the plus strand (dashed ORF box) in addition to the already annotated *ORFK7* and  $T_{1.5}$  described in this paper.

25068), M (nt 25075 to 25388), X (nt 28229 to 28527), and H (nt 23600 to 23900) regions. In attempts to identify  $T_{1.5}$ , 12 clones hybridizing to the L and M probes were isolated and sequenced (Fig. 3). The clones ranged in length from 360 to 830 bp, all possessing a plus-orientation polyadenylation signal at nt 25440. All the transcripts were intronless and prematurely terminated at their 5' end in the *fnk* direct repeat region. Attempts to reverse transcribe through this region using 5'RACE were unsuccessful due to its repetitive feature. Translation analysis of the lambda phage inserts revealed a potential open reading frame (which we tentatively designate *ORFK4.4*) possibly encoding a 75-amino-acid polypeptide with a start AUG at nt 25198 and a stop codon at nt 25425.

The  $T_{6.1}$  transcript was similarly examined by lambda phage library screening using the probe from the X region (nt 28229 to 28527). Five clones were isolated and sequenced; these inserts ranged from 1.1 to 3.3 kb in length and all were coterminal with the *PAN* transcript at nt 29741 (Fig. 4). A comparison to the 1.1-kb *PAN* transcript shows that  $T_{6.1}$  is clearly distinct from the previously annotated and more highly abundant *PAN* sequence. 5' RACE of cDNA generated from BC-1 mRNA using minus strand primers from the H region revealed

the  $T_{6.1}$  start site to be at nt 23596. This is consistent with  $T_{6.1}$  being a continuous 6,145-nt transcript which is alternatively expressed with *PAN*. We can only hypothesize that the entire transcript is identical to the KSHV genomic sequence based on size comparison. The longest sequenced phage,  $\phi X-1$ , which is 3.3 kb in length (Fig. S1 in the supplemental material, GenBank DQ097683) is identical to the KSHV genomic sequence. An analysis of the 6.145-kb region using the MacVector program (Accelrys, Inc.) found one possible ORF, of over 100 amino acids in length, from nt 27887 to 28192 (Fig. 4) in addition to the already annotated *ORFK7* (nt 28622 to 29002) and  $T_{1.5}$  described here.

For analysis of the transcripts revealed using probe H, we screened over 2 million phages using this probe H (nt 23600 to 23900) and identified eight identical clones that were isolated for plasmid excision and sequencing. cDNA inserts were all 1.3 kb in length and possessed a minus-orientation polyadenylation signal at nt 18608. When mapped, all eight H phage clones encoded the entire *K3* open reading frame (nt 18596 to 19617) along with 208 bp of *ORF70* (nt 20096 to 20304) and 70 bp of the H region (nt 23770 to 23840) (Fig. 5; Fig. S2 in the supplemental material, GenBank DQ097682). 5'RACE was per-

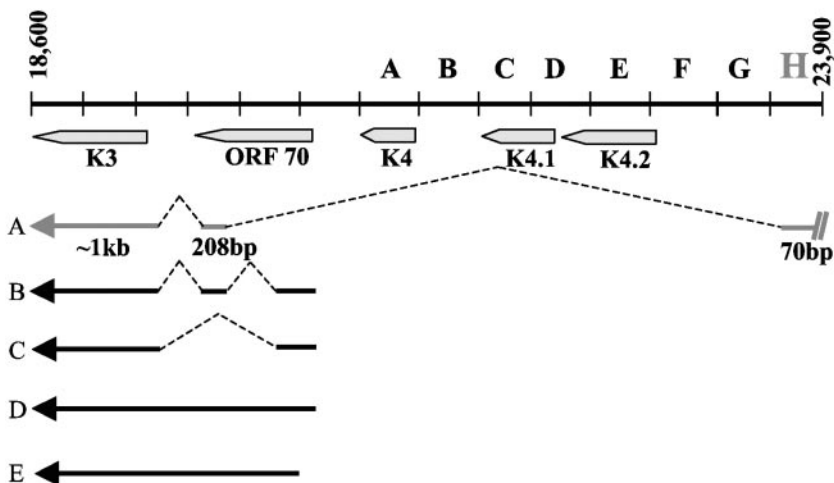


FIG. 5. Lambda phage screening with probe H for identification of  $T_{2.5}$ . (A) Schematic shows alignment to KSHV genome of eight identical clones retrieved from phage screening with Probe H. The previously unidentified transcript encodes all of *ORFK3* (nt 18596 to 19617), splices to include 208 bp of *ORF70* (nt 20096 to 20304), and then splices to a 70-bp fragment within region H (nt 23770 to 23840). (B to E) Rimessi et al. (24) identified four *K3* transcripts, a 1.5-kb immediate-early transcript (B) and three early transcripts of 1.3, 2.5, and 2.4 kb (C to E).

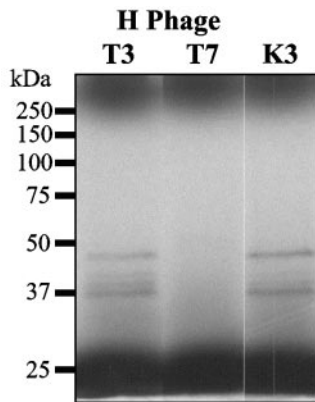


FIG. 6. In vitro translation of 1.3-kb H phage clone  $\phi$ H-3 expresses MIR1 protein. The excised clone from phage screening was in vitro translated in both directions using either T3 or T7 polymerase along with a plasmid containing *ORFK3* (P140) as a positive control. The resultant banding pattern matches between the T3 polymerase-translated H phage clone and *ORFK3* control plasmid.

formed with cDNA reverse transcribed from poly(A)-selected BC-1 mRNA using a *K3* primer for the un-nested PCR followed by a nested primer specific for the 70 bp in region H but could not identify the 5' end of this transcript.

To determine if the H phage clone ( $\phi$ H-3), which was shown to contain *ORFK3*, encodes MIR1, we performed in vitro translation and used a plasmid containing full-length *ORFK3* as the positive control. The results showed identical translation

products for both *ORFK3* and the H phage clone, suggesting that the H phage clone produces MIR1 at low constitutive levels during latency (Fig. 6). This was confirmed by Northern blot analysis using the different regions of the H phage clone as separate probes (Fig. 7). When probes from *K3* and *ORF70* regions are used, TPA-inducible transcripts that have been previously described (1.3 kb, 1.5 kb, and 2.4 to 2.5 kb) are seen (22). Use of a probe from the unique 70-bp exon of region H, however, reveals two constitutively expressed transcripts ( $T_{2.5}$  and  $T_{3.0}$ ) that are not induced by TPA treatment and one class II transcript,  $T_{4.5}$ . As seen in Fig. 1,  $T_{3.0}$  and  $T_{4.5}$  extend from the H region (nt 23600 to 23900) into the G region (nt 23300 to 23600). We were unable to further characterize these two transcripts. The  $T_{2.5}$  transcript is obscured when Northern blotting is performed using *K3* probes by more abundant TPA-inducible transcripts. Thus, while the *K3* gene encoding MIR1 can be classified as a class II gene (constitutive plus TPA induced), its expression is determined by a complex set of alternative transcripts that include a constitutive transcript (class I) containing the H exon, which is expressed during PEL cell latency, as well as more abundant transcripts (class III) induced by TPA treatment, which lack this exon.

DISCUSSION

The goal of this study was to analyze the gene-poor region of the KSHV genome between open reading frames *K4.2* (nt 22598 to 23146) and *K7* (nt 28622 to 29002). This region contains two genes, *K5* (nt 25713 to 26483) and *K6* (nt 27137 to

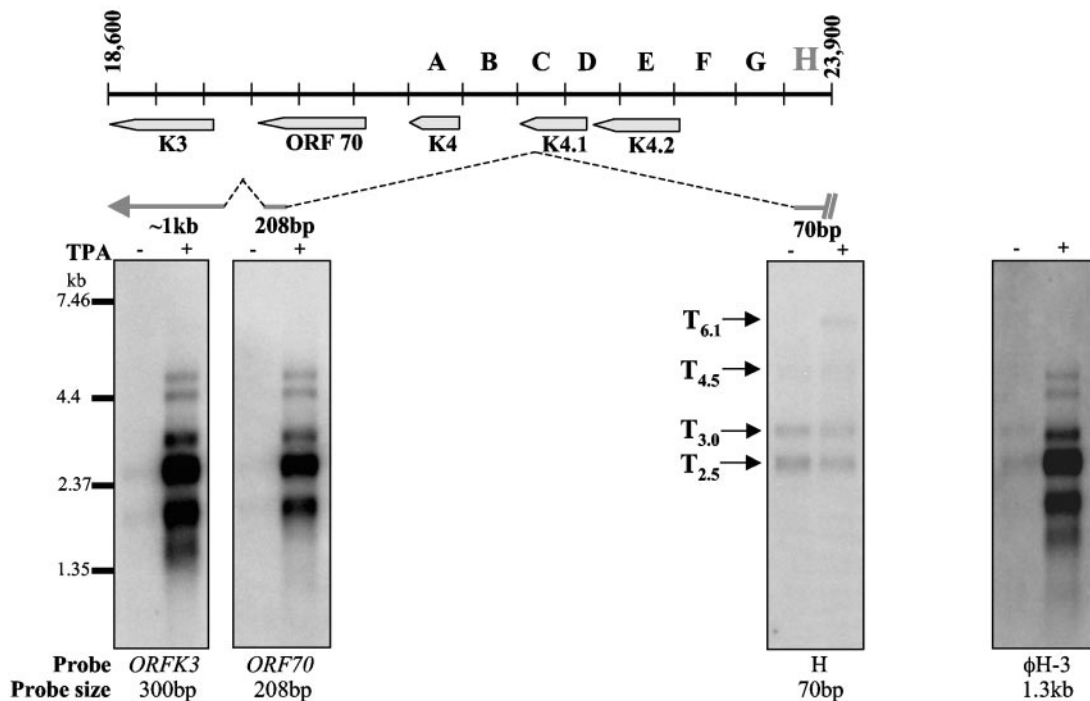


FIG. 7. Latent *ORFK3* transcripts are identified by Northern blotting with H probe. The schematic shows the splicing pattern of the excised H phage clone. mRNA from untreated or TPA-treated BC-1 cells was used for Northern blot analyses along with probes for each portion of the clone: 300 bp of *K3*, 208 bp of *ORF70*, or 70 bp of region H. The constitutive  $T_{2.5}$  transcript encoded by the H phage clone  $\phi$ H-3 is seen by Northern blotting with the H fragment but is obscured by induced transcripts when hybridized with probes from *ORFK3* and *ORF70* regions or with the entire  $\phi$ H-3 clone. Two additional constitutive bands ( $T_{3.0}$  and  $T_{4.4}$ ) hybridize to the H fragment but were not isolated.

27424), whose transcript sizes combined are ~1 kb of the sequence, leaving approximately 5.5 kb of sequence unannotated. Recently, a possible transcript from this area has been reported. AuCoin et al. (1) and Lin et al. (16) described two duplicated copies of the lytic DNA replication origin [designated *ori-Lyt* (L) and *ori-Lyt* (R)] located in the KSHV genome between *K4.2* and *K5* and between *K12* and *ORF71*, respectively. In addition, Wang et al. (32) discovered a polyadenylated RNA of 1.4 kb on the plus strand downstream of the *ori-Lyt* (L) domain. The transcript was absent in viral latent infection but induced upon TPA treatment of the PEL cell line, BCBL-1. It is likely that this transcript corresponds to the  $T_{1.5}$  transcript described in this study. The sequence of the RNA consists of the whole *fnk* direct repeat region (nt 24285 to 24902) and possibly encodes an ORF of 75 amino acids. From this study, we have found multiple transcripts previously unidentified in the KSHV genome in the area between 23 kb and 29.7 kb.

The  $T_{1.5}$  transcript was detected by Northern blot analysis and partially identified by lambda phage screening. All 12 clones excised from the phage screening had polyadenylation signals in the plus orientation at nt 25440, corresponding to the 3' end of the 1.4 kb transcript recently described. Our attempts to identify the 5' end of this transcript were unsuccessful. Wang et al. (32) performed 5'RACE and discovered that the transcript started at or near nt 24243. This transcript includes the entire *fnk* direct repeat sequence that may explain the difficulty of performing 5'RACE. The 5' end of the transcript starts at or near nt 24243 according to Wang et al. (32), and the *fnk* direct repeat region starts at nt 24285, leaving approximately 40 bp of the 5' end of the transcript outside this direct repeat region. From sequencing our lambda phage clones, and in conjunction with data reported by Wang et al. (32), we hypothesize that the 3' end of this transcript (nt 25198 to 25425) possibly encodes an open reading frame of 75 amino acids.

In the region spanning KSHV nt 24802 to 29719, we identified a large class III early transcript of approximately 6.1 kb ( $T_{6.1}$ ). This transcript was shown to be present in 17 consecutive probes by Northern blot analysis of TPA-stimulated PEL cell mRNA. RACE identified the 5' end of the transcript at nt 23596 and the five excised clones from lambda phage screening identified the 3' end of the transcript at nt 29741 of KSHV, making the transcript coterminal with *PAN*. The longest sequenced phage (3.3 kb in length) is identical to the KSHV genomic sequence. Size limitation of phage clones in our cDNA library prevents us from confirming that the entire transcript also has identity to the genomic sequence. However, contingent upon identity between genomic and the 6.145-kb cDNA sequences, analyses revealed one possible ORF in addition to *ORFK7* and  $T_{1.5}$  described in this paper. *ORFK7* is present within  $T_{6.1}$  yet has no identifiable canonical or noncanonical polyadenylation site, whereas  $T_{1.5}$  does possess a polyadenylation site. It is conceivable that there is a small splice site in the region of the  $T_{1.5}$  polyadenylation signal or that read-through occurs which produces  $T_{6.1}$ . Briata et al. (2) showed read-through of polyadenylation of HLA-DQB transcripts. In addition, read-through has been shown in HSV between ORFs UL29 and UL30 and between ORFs UL30 and UL31 of the genome due to weak polyadenylation signals (23).

The presence of large transcripts is not unique to KSHV, as long primary transcripts have been reported for other herpesvirus genomes. In herpes simplex virus, a minor 8.5-kb LAT transcript (*mLAT*) accumulates during latency but is less precisely defined than other herpes simplex virus LAT transcripts and has been reported to be an unstable transcript (23). The 8.5-kb large latency transcript (*LLT*) of pseudorabies virus is the only transcript made during latent viral infection (7). Smuda et al. (29) found large overlapping early lytic transcripts of 6 kb, 8 kb, 10 kb, and 14 kb within the human cytomegalovirus genome. Wirth et al. (33) identified a 6-kb immediate-early transcript and six late lytic transcripts, ranging from 4.5 kb to >8 kb, in bovine herpesvirus 1. It is unclear why these viruses produce such large transcripts, although it has been postulated that their size may lead to RNA stability by the formation of pseudoknots. Aside from functional significance for the virus, the presence of a large transcript spanning and overlapping smaller transcripts has implications in gene expression analysis using DNA microarrays.

Three transcripts,  $T_{2.5}$ ,  $T_{3.0}$ , and  $T_{4.5}$ , were identified using probe H. Phage screening identified  $T_{2.5}$  as a ~1.3-kb alternative transcript of *ORFK3*. Through in vitro translation, this transcript was shown to produce banding patterns identical to those of MIR1. Through Northern blot analysis, the alternative *K3* transcript was identified, and riboprobes confirmed its directionality on the minus strand. Rimessi et al. (24) identified four transcripts which encode *K3*, one immediate-early and three early transcripts (Fig. 5B to E). The 1.5-kb immediate-early transcript (B) and the 1.3-kb early transcript (C) contain the entire *ORFK3* coding region but have different splicing patterns that include parts of the *ORF70* (thymidylate synthase) sequence. The other two early transcripts (D and E) are intronless, and the 2.5-kb transcript (D) is capable of producing both *ORF70* and *K3*. We confirmed these transcripts with Northern blotting and 5' RACE.  $T_{2.5}$  is a novel class I *K3* transcript, however, encoding *ORFK3* in its entirety and splicing to the 3' end of *ORF70* and then splicing a second time to 70 bp within region H.

Studies have identified several KSHV lytic proteins that play a role in immune regulation during lytic replication. MIR1 (encoded by *ORFK3*) and MIR2 (encoded by *ORFK5*) downregulate immunoregulatory proteins, including MHC-I and ICAM-1 from the cell surface to limit immune recognition (8, 9, 12, 13, 14, 26). Tomescu et al. (31) showed that MHC-I, PE-CAM, and ICAM-1 are downregulated during de novo infection with KSHV of both immortalized and primary human endothelial cells.  $T_{2.5}$  is a likely candidate responsible for expression and MHC-I downregulation during latency.

In this study, we have identified four novel KSHV transcripts— $T_{2.5}$ ,  $T_{3.0}$ ,  $T_{4.5}$ , and  $T_{6.1}$ —and have confirmed the existence of  $T_{1.5}$ , the 1.4-kb transcript described by Wang et al. (32). All four of these transcripts were identified with Northern blot analyses using Probe H (nt 23300 to 23600).  $T_{2.5}$  and  $T_{3.0}$  were classified as class I, latent transcripts, and  $T_{4.5}$  was classified as a class II transcript; all of these are transcribed in the minus orientation. The large  $T_{6.1}$  transcript was confirmed as a class III late lytic transcript with a positive orientation. These results emphasize the complexity of KSHV transcripts and the fact that, despite having performed extensive whole genome

analyses, researchers will likely continue to find novel KSHV transcripts with important biologic functions.

#### ACKNOWLEDGMENTS

We thank Arlee Fafalios for her help with lambda phage screening and sequencing of clones.

This work was supported by Public Health Service grants CA67391-06, CA83485-01, and CA87661-01 from NIH/NCI.

#### REFERENCES

- AuCoin, D. P., K. S. Colletti, Y. Xu, S. A. Cei, and G. S. Pari. 2002. Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) contains two functional lytic origins of DNA replication. *J. Virol.* **76**:7890–7896.
- Briata, P., S. F. Radka, S. Sartoris, and J. S. Lee. 1989. Alternative splicing of HLA-DBQ transcripts and secretion of HLA-DQ  $\beta$ -chain proteins: allelic polymorphism in splicing and polyadenylation sites. *Proc. Natl. Acad. Sci. USA* **86**:103–1007.
- Cai, X., S. Lu, Z. Zhang, C. M. Gonzalez, B. Damania, and B. R. Cullen. 2005. Kaposi's sarcoma-associated herpesvirus expresses an array of viral microRNAs in latently infected cells. *Proc. Natl. Acad. Sci. USA* **102**:5570–5575.
- Cesarman, E., P. S. Moore, P. H. Rao, G. Inghirami, D. M. Knowles, and Y. Chang. 1995. In vitro establishment and characterization of two AIDS-related lymphoma cell lines (BC-1 and BC-2) containing Kaposi's sarcoma-associated herpesvirus-like (KSHV) DNA sequences. *Blood* **86**:2708–2714.
- Chandran, B., C. Bloomer, S. R. Chan, L. Zhu, E. Goldstein, and R. Horvat. 1998. Human herpesvirus-8 ORF K8.1 gene encodes immunogenic glycoproteins generated by spliced transcripts. *Virology* **249**:140–149.
- Chatterjee, M., J. Osborne, G. Bestetti, Y. Chang, and P. S. Moore. 2002. Viral IL-6-induced cell proliferation and immune evasion of interferon activity. *Science* **298**:1432–1435.
- Cheung, A. K., and T. A. Smith. 1999. Analysis of the latency-associated transcript/UL1-3.5 gene cluster promoter complex of pseudorabies virus. *Arch. Virol.* **144**:381–391.
- Coscoy, L., and D. Ganem. 2000. Kaposi's sarcoma-associated herpesvirus encodes two proteins that block cell surface display of MHC class I chains by enhancing their endocytosis. *Proc. Natl. Acad. Sci. USA* **97**:8051–8056.
- Coscoy, L., and D. Ganem. 2001. A viral protein that selectively downregulates ICAM-1 and B7-2 and modulates T cell costimulation. *J. Clin. Investig.* **107**:1599–1606.
- Gao, S. J., L. Kingsley, M. Li, W. Zheng, C. Parravicini, J. Ziegler, R. Newton, C. R. Rinaldo, A. Saah, J. Phair, R. Detels, Y. Chang, and P. S. Moore. 1996. KSHV antibodies among Americans, Italians and Ugandans with and without Kaposi's sarcoma. *Nat. Med.* **2**:925–928.
- Glenn, M., L. Rainbow, F. Aurade, A. Davison, and T. F. Schulz. 1999. Identification of a spliced gene from Kaposi's sarcoma-associated herpesvirus encoding a protein with similarities to latent membrane proteins 1 and 2A of Epstein-Barr virus. *J. Virol.* **73**:6953–6963.
- Haque, M., K. Ueda, K. Nakano, Y. Hirata, C. Parravicini, M. Corbellino, and K. Yaminishi. 2001. Major histocompatibility complex class I molecules are down-regulated at the cell surface by the K5 protein encoded by Kaposi's sarcoma-associated herpesvirus/human herpesvirus-8. *J. Gen. Virol.* **82**:1175–1180.
- Ishido, S., J. K. Choi, B. S. Lee, C. Wang, M. DeMaria, R. P. Johnson, G. B. Cohen, and J. U. Jung. 2000. Inhibition of natural killer cell-mediated cytotoxicity by Kaposi's sarcoma-associated herpesvirus K5 protein. *Immunity* **13**:365–374.
- Ishido, S., C. Wang, B. S. Lee, G. B. Cohen, and J. U. Jung. 2000b. Down-regulation of major histocompatibility complex class I molecules by Kaposi's sarcoma-associated herpesvirus K3 and K5 proteins. *J. Virol.* **74**:5300–5309.
- Jenner, R. G., M. M. Alba, C. Boshoff, and P. Kellam. 2001. Kaposi's sarcoma-associated herpesvirus latent and lytic gene expression as revealed by DNA arrays. *J. Virol.* **75**:891–902.
- Lin, C. L., H. Li, Y. Wang, F. X. Zhu, S. Kudchodkar, and Y. Yuan. 2003. Kaposi's sarcoma-associated herpesvirus *ori-Lyt*-dependent DNA replication: identification of the *ori-Lyt* and association of K8 bZIP protein with the origin. *J. Virol.* **77**:5578–5588.
- Neipel, F., J. C. Albrecht, A. Ensser, Y. Q. Huang, J. J. Li, A. E. Friedman-Kien, and B. Fleckenstein. 1997. Human herpesvirus 8 encodes a homolog of interleukin-6. *J. Virol.* **71**:839–842.
- Neipel, F., J. C. Albrecht, and B. Fleckenstein. 1997. Cell-homologous genes in the Kaposi's sarcoma-associated rhadinovirus human herpesvirus 8: determinants of its pathogenicity? *J. Virol.* **71**:4187–4192.
- Nicholas, J., V. Ruvolo, J. Zong, D. Ciufu, H. G. Guo, M. S. Reitz, and G. S. Hayward. 1997. A single 13-kilobase divergent locus in the Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) genome contains nine open reading frames that are homologous to or related to cellular proteins. *J. Virol.* **71**:1963–1974.
- Paulose-Murphy, M., N.-K. Ha, C. Xiang, Y. Chen, L. Gillim, R. Yarchoan, P. Meltzer, M. Bittner, J. Trent, and S. Zeichner. 2001. Transcription program of human herpesvirus 8 (Kaposi's sarcoma-associated herpesvirus). *J. Virol.* **75**:4843–4853.
- Pfeffer, S., A. Sewer, M. Lagos-Quintana, R. Sheridan, C. Sander, F. A. Grasser, L. F. van Dyk, C. K. Ho, S. Shuman, M. Chien, J. J. Russo, J. Ju, G. Randall, B. D. Lindenbach, C. M. Rice, V. Simon, D. D. Ho, M. Zavolan, and T. Tuschli. 2005. Identification of microRNAs of the herpesvirus family. *Nat. Methods.* **2**:269–276.
- Raab, M. S., J. C. Albrecht, A. Birkmann, S. Yaguboglu, D. Lang, B. Fleckenstein, and F. Neipel. 1998. The immunogenic glycoprotein gp35-37 of human herpesvirus 8 is encoded by open reading frame K8.1. *J. Virol.* **72**:6725–6731.
- Rajcani, J., A. Vojvodova, and R. Ingeborg. 2004. Peculiarities of herpes simplex virus (HSV) transcription: an overview. *Virus Genes* **28**:293–310.
- Rimessi, P., A. Bonaccorsi, M. Sturzl, M. Fabris, E. Brocca-Cofano, A. Caputo, G. Mellucci-Vigo, M. Falchi, A. Cafaro, E. Cassai, B. Ensoli, and P. Monini. 2001. Transcription pattern of human herpesvirus 8 open reading frame K3 in primary effusion lymphoma and Kaposi's sarcoma. *J. Virol.* **75**:7161–7174.
- Russo, J. J., R. A. Bohenzky, M. C. Chien, J. Chen, M. Yan, D. Maddalena, J. P. Parry, D. Peruzzi, I. S. Edelman, Y. Chang, and P. S. Moore. 1996. Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus (HHV8). *Proc. Natl. Acad. Sci. USA* **93**:14862–14867.
- Sanchez, D. J., L. Coscoy, and D. Ganem. 2002. Functional organization of MIR2, a novel viral regulator of selective endocytosis. *J. Biol. Chem.* **277**:6124–6130.
- Sarid, R., O. Flore, R. A. Bohenzky, Y. Chang, and P. S. Moore. 1998. Transcription mapping of the Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) genome in a body cavity-based lymphoma cell line (BC-1). *J. Virol.* **72**:1005–1012.
- Sarid, R., J. S. Wieszorek, P. S. Moore, and Y. Chang. 1999. Characterization and cell cycle regulation of the major Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) latent genes and their promoter. *J. Virol.* **73**:1438–1446.
- Smuda, C., E. Bogner, and K. Radsak. 1997. The human cytomegalovirus glycoprotein B gene (ORF UL55) is expressed early in the infectious cycle. *J. Gen. Virol.* **78**:1981–1992.
- Sun, R., S. F. Lin, K. Staskus, L. Gradoville, E. Grogan, A. Haase, and G. Miller. 1999. Kinetics of Kaposi's sarcoma-associated herpesvirus gene expression. *J. Virol.* **73**:2232–2242.
- Tomescu, C., W. K. Law, and D. H. Kedes. 2003. Surface downregulation of major histocompatibility complex class I, PE-CAM, and ICAM-1 following de novo infection of endothelial cells with Kaposi's sarcoma-associated herpesvirus. *J. Virol.* **77**:9669–9684.
- Wang, Y., H. Li, M. Y. Chan, F. X. Zhu, D. M. Lukac, and Y. Yuan. 2004. Kaposi's sarcoma-associated herpesvirus *ori-Lyt*-dependent DNA replication: *cis*-acting requirements for replication and *ori-Lyt*-associated RNA transcription. *J. Virol.* **78**:8615–8629.
- Wirth, U. V., K. Gunkel, M. Engels, and M. Schwyzer. 1989. Spatial and temporal distribution of bovine herpesvirus 1 transcripts. *J. Virol.* **63**:4882–4889.
- Zhang, Y. J., J. H. Deng, C. Rabkin, and S. J. Gao. 2000. Hot-spot variations of Kaposi's sarcoma-associated herpesvirus latent nuclear antigen and application in genotyping by PCR-RFLP. *J. Gen. Virol.* **81**:2049–2058.
- Zong, J. C., D. M. Ciufu, D. J., Alcantor, X. Wan, J. Nicholas, P. J. Browning, P. L. Rady, S. K. Tying, J. M. Orenstein, C. S. Rabkin, I. J. Su, K. F. Powell, M. Croxson, K. E. Foreman, B. J. Nickoloff, S. Alkan, and G. S. Hayward. 1999. High-level variability in the ORF-K1 membrane protein gene at the left end of the Kaposi's sarcoma-associated herpesvirus genome defines four major virus subtypes and multiple variants or clades in different human population. *J. Virol.* **73**:4156–4170.