Maps of Papillomavirus mRNA Transcripts

Carl Baker[a] **and Charles Calef**[b]

[a] *Laboratory of Tumor Virus Biology,National Cancer Institute, National Institutes of Health,Bethesda, Maryland 20892-5055*
[b] *MS K710, Los Alamos National Laboratory, Los Alamos, New Mexico 87545*

Explanation of Maps

In the 1995 compendium, we presented maps showing the open reading frames (ORFs) and mRNA transcripts of seven papillomaviruses (BPV-1R, HPV-5, HPV-8, HPV-11, HPV-16R, HPV-31, and HPV-47). These maps can be obtained via the World Wide Web at http://hpv-web.lanl.gov/HTML_FILES/HPVcompintro2.html#comp95. This year we present a new map for HPV-1a and updated maps for HPV-11, HPV-16R, and HPV-31. Facing each map is a brief description of the transcripts, and following the maps is a list of references from which the maps were compiled.

In each map the significant ORFs are shown in their proper reading frames as colored rectangles. At the upper left end of the rectangle are two numbers. The first corresponds to the nucleotide (nt) position of the ORF start, the first nucleotide following a stop codon. The second number records the nt position of the first ATG, which is also indicated by a dotted line within the rectangle. The position of the last nt in the stop codon of each ORF is printed at the lower right corner of the rectangles. The rectangle's colored fill begins at the ATG and extends to the end; thus it corresponds to the coding sequence, not the entire ORF. In the cases where no ATG exists in the ORF, only one number is present in the upper left corner and the rectangle is completely filled with color. Below the ORFs is a scale of the genome divided into thousands. On the scale are placed the positions of promoters (represented by arrows) and the poly(A) signals. The exact position of the poly(A) signal is indicated. Located below the genome scale are diagrams of mRNA species, most of which are spliced. The exons are illustrated by narrow black rectangles, while the introns are indicated by black hairlines between. The numbers printed below the lines indicate the 5' and 3' termini of the RNAs, and the 5' and 3' splice junction positions. The splice junction numbers give the position of the last nucleotide in the exon before the splice and the position of the first nucleotide of the exon following the splice. Splice junctions in parentheses were deduced from the genomic sequence and have not been confirmed by cDNA sequencing. Where 5' or 3' ends of the RNAs are uncertain, no nt position is given. Superimposed on the exons are colored rectangles representing the gene, or part thereof, coded by that portion of the exon. The coding potential of each transcript is also listed at the right and includes all proteins which could be translated from a given mRNA. In that list a ^ symbol between two gene names (e.g., E1^E4) indicates a fusion product. The * symbol indicates a truncated form of a protein. Alternative splicing can join the same upstream exon to different downstream exons. The resulting truncated proteins are identified by roman numerals, such as E6*I, E6*II, and E6*III as illustrated by HPV16R species B, C, and D.

Occasionally an intron occurs wholly within an ORF. In such cases it is possible that the normal beginning of the ORF may be spliced to a downstream, in-frame portion of the orf. This results in a truncated version of the complete protein, bearing a normal n-terminal portion (“head”) and normal c-terminal portion (“tail”), but missing the middle. The E6*I protein from species B in HPV16R is an example.

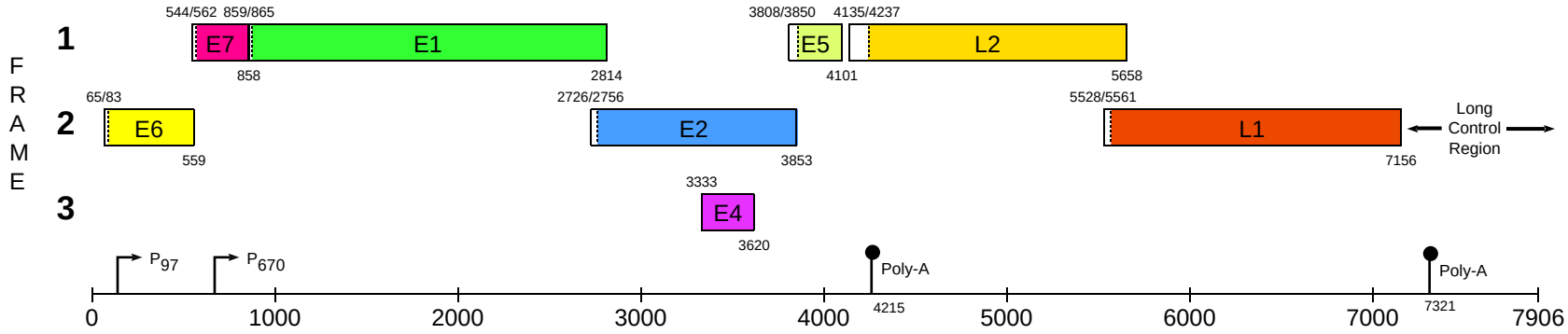
Truncation can also occur when an upstream exon is spliced to a downstream exon such that the downstream exon is translated in a reading frame that results in a stop codon soon being encountered. The result is a protein composed of a head, translated in one of the “normal” ORFs such as E6 or E1, followed by a short tail that is out of frame. We have chosen to draw this out-of-frame (OOF) tail as a narrower box whose color corresponds to the main body of the protein. An example of this situation is the E6*II protein from species F and J in HPV16R. Although the tail of this protein lies within the E6 region, it is not in the E6 ORF, but is translated in a different frame which results in a stop codon being quickly encountered.

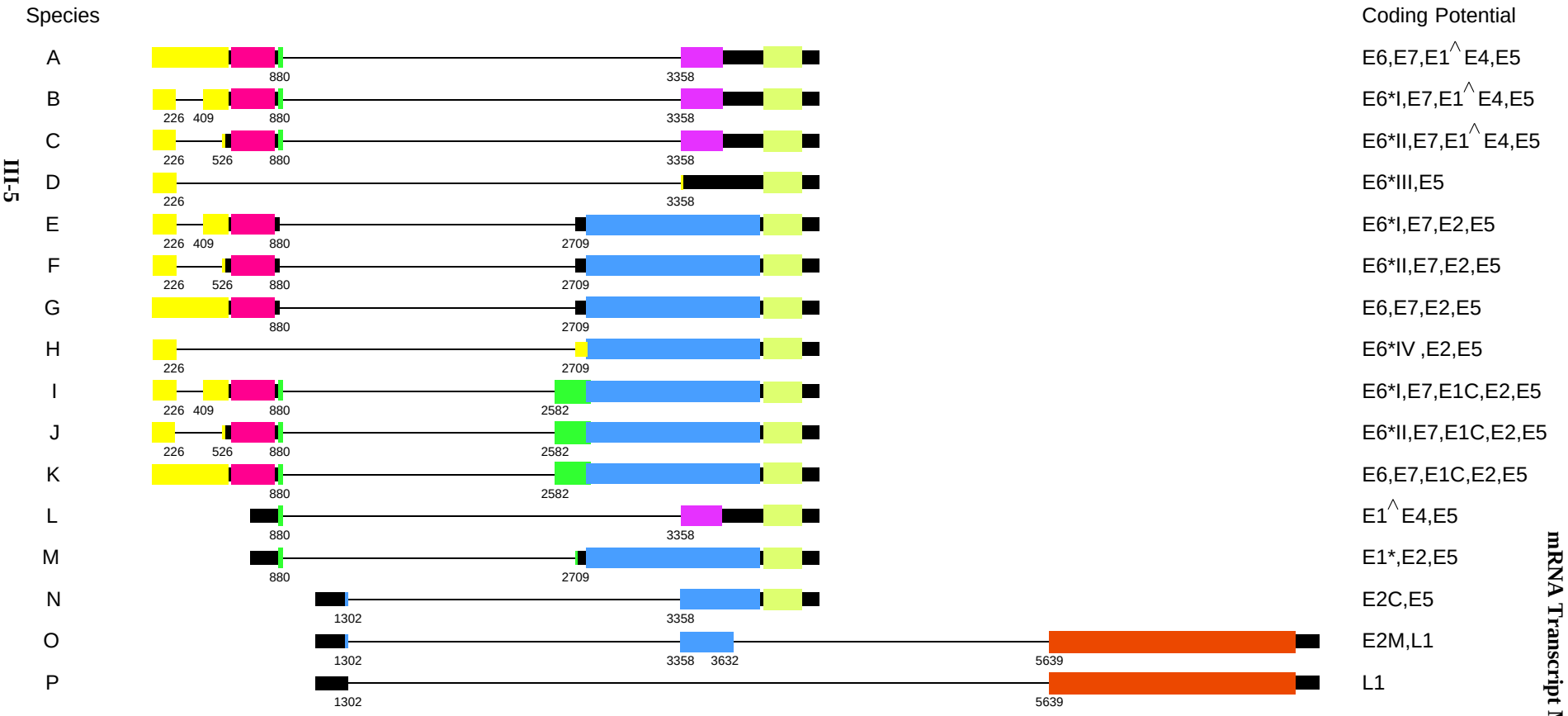
The next level of complexity is exemplified by the E1M fusion protein coded by species B in HPV1a. Here the first 139 amino acids are normal E1 followed by 20 amino acids translated from RNA in the E2/E4 overlapping region of the genome. However, these 20 amino acids are neither E2 or E4 because they happen to be translated in frame 2. In the same mRNA species (species B in HPV1a) the “converse” situation is also illustrated; the E2C protein bears an OOF head (beginning at a methionine in the E1 region) joined to a normal E2 tail. The narrow blue bar in exon 1 goes with the thick (normal) blue bar in exon 2. Ocassionally a short normal head is joined to a short OOF tail, such as in the E1* protein of species C in HPV1a. We illustrate these merely because they have the potential to be translated.

HPV-16

HPV-16 mRNAs isolated from transfected cells and a variety of tumor cell lines and lesions containing both extrachromosomal and integrated HPV-16 genomes have been analyzed in multiple laboratories [6,7,14,17–19]. Viral/host chimeric mRNAs have been purposely omitted. All nucleotide positions correspond to the revised HPV-16 sequence published in Part I of the 1995 compendium. Most mRNA species were determined by RT-PCR, so the identity of the 5' and 3' ends are not known. However, at least two promoters have now been identified for HPV-16. One promoter (P_{97}) is active in cervical cancers and cervical cancer-derived cell lines as well as in monolayer and organotypic raft cultures containing extrachromosomal HPV-16 DNA [7,21]. Species A–K are most likely transcribed from this promoter. A differentiation inducible promoter (P_{670}) has been identified by analysis of RNA obtained from organotypic raft cultures containing extrachromosomal HPV-16 DNA [7]. This promoter has a major start site mapped by primer extension analysis to around nt 670 and minor start sites at nt 693, 706, 713, and 766. It is likely that the E1^E4 mRNA (transcript L) is transcribed from this promoter. In addition, in situ hybridization analyses have suggested that the most abundant transcripts in intraepithelial neoplasias and invasive cancers are transcribed from this promoter [1]. An additional differentiation inducible promoter may also exist with a start site around nt 480 [7]. The promoter responsible for the transcription of the late mRNAs (species O–P) is not known. The early and late polyadenylation signals are located at nt 4215 and nt 7321, respectively [9,14].

HPV16R





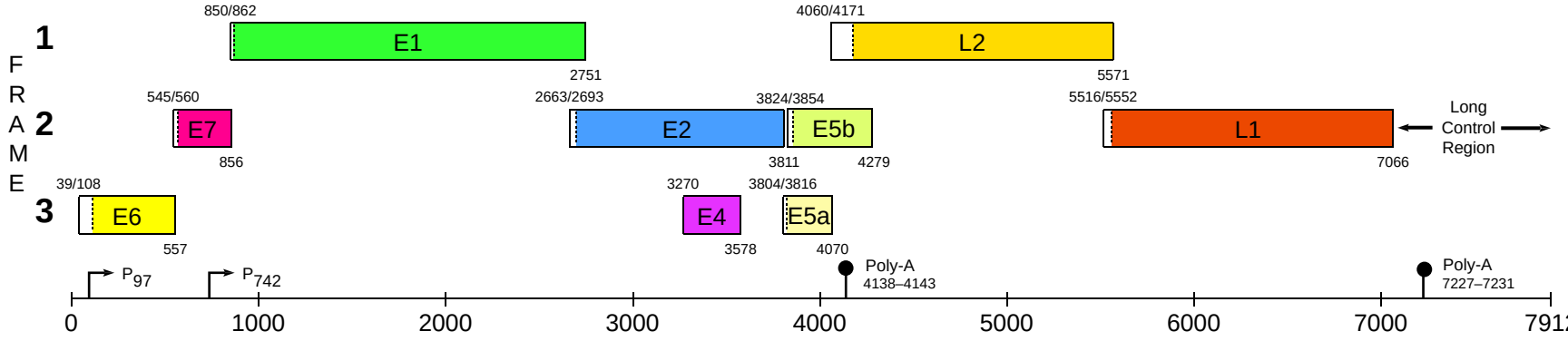
III-5
AUG 96

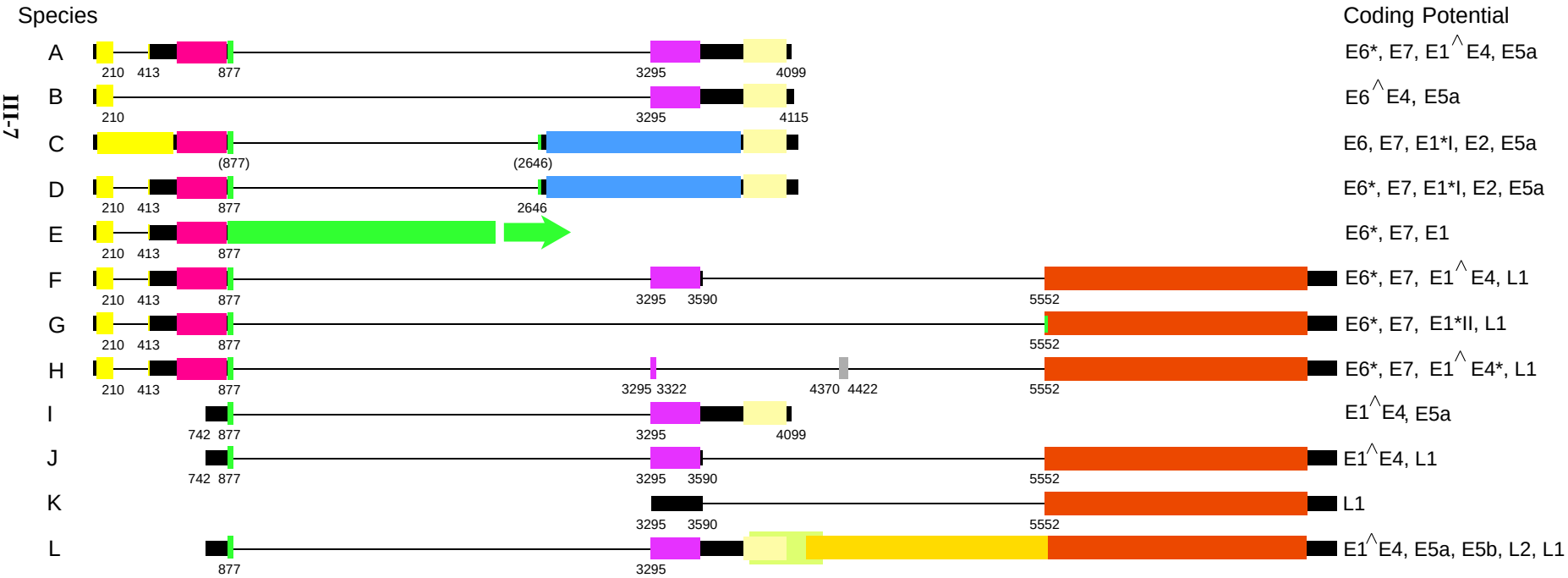
mRNA Transcript Maps

HPV-31

HPV-31 mRNAs were investigated in CIN612 cells grown in monolayer cultures and in organotypic raft cultures treated with TPA or C8 [8,12, Ozbun and Meyers, manuscript in preparation]. CIN612 cells contain extrachromosomal HPV-31. Two promoters were identified by primer extension and RNAse protection experiments. P_{97} is active in both monolayer and organotypic raft cultures; P_{742} is active only in differentiated raft cultures [8]. No promoter could be identified in the E6 ORF [8]. The mRNA structures identified are shown in the facing map. Species A–E are presumably transcribed from P_{97}, although this has not been verified. It is possible that one or more of these species is transcribed from a promoter in the LCR. The 5' ends of species F–H have been mapped to P_{97} by nuclease S1 and ExoVII analysis (Ozbun and Meyers, manuscript in preparation). Species I, J, and presumably L are transcribed from P_{742}. It is not know whether the 5' end of transcript K represents the 5' end of the mRNA or a splice junction. Additional L1 mRNAs appear to be transcribed from P_{97}, but the exact splice structure of these mRNAs is unknown. Early mRNAs were polyadenylated between nt 4099 and 4125 [8]. The exact sites of polyadenylation at the late poly(A) site were not determined in these studies, but a putative polyadenylation signal exists at nt 7227.

HPV31





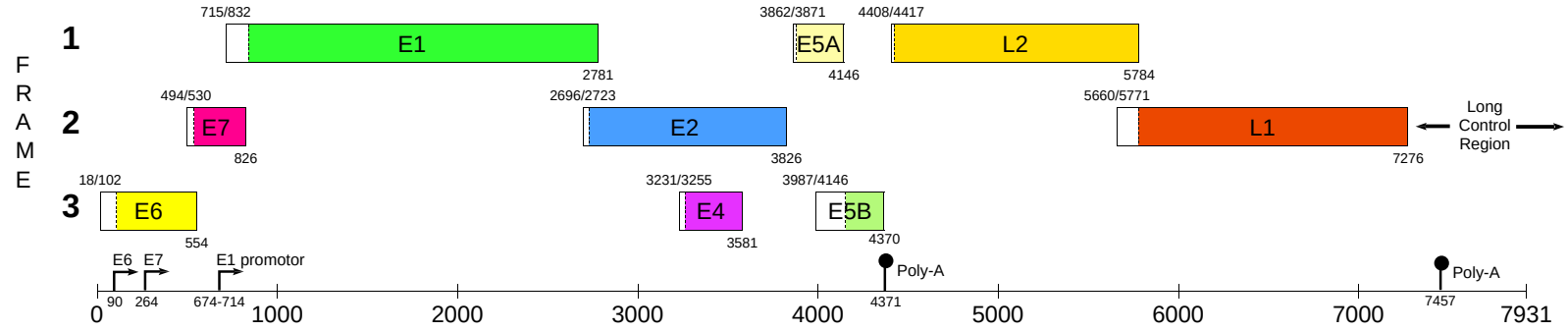
III-7
AUG 96

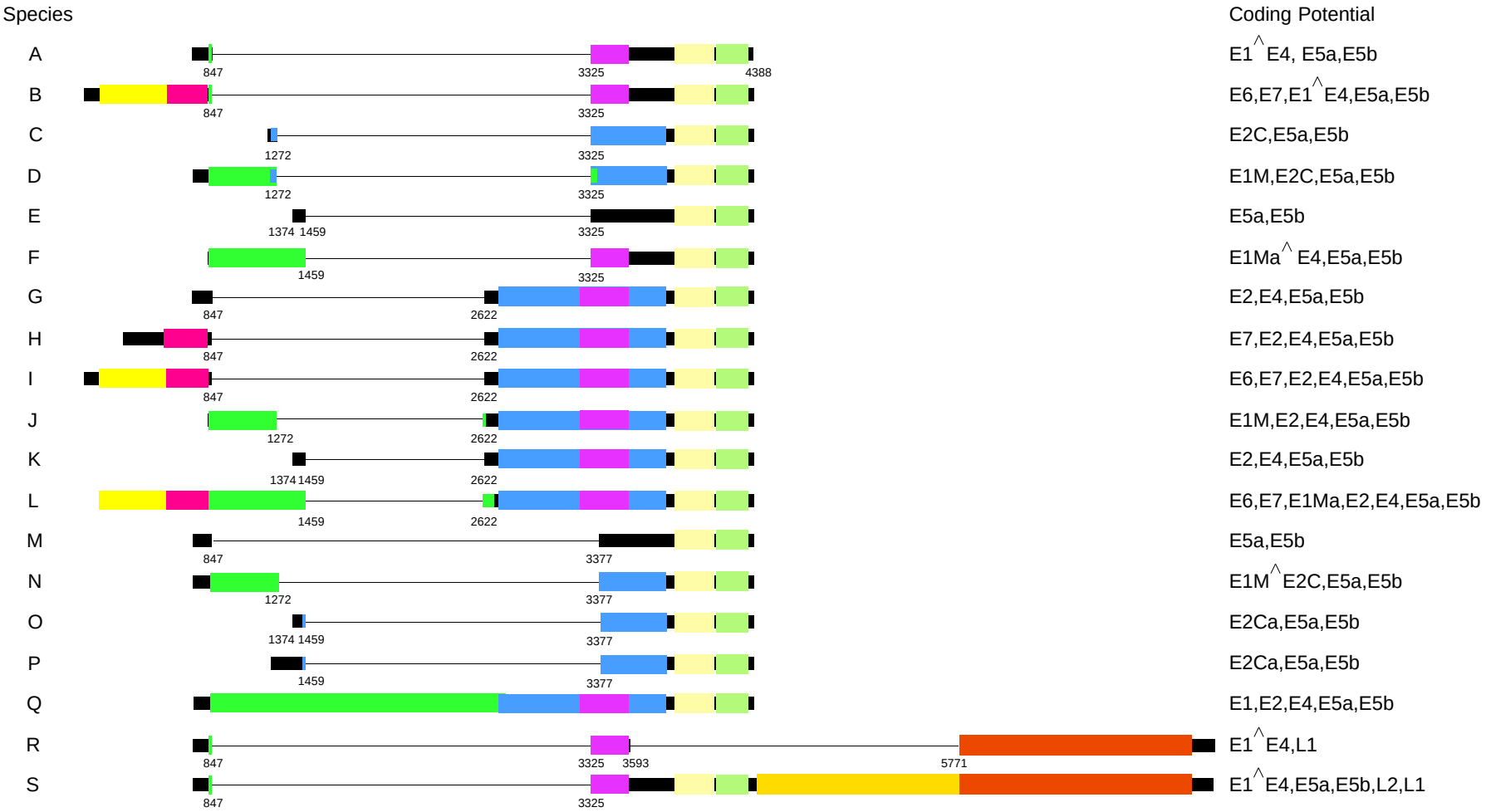
mRNA Transcript Maps

HPV-11

HPV-11 mRNAs, isolated from genital condyloma acuminata and experimental condylomatous cysts implanted under the renal capsules of nude mice, have been analyzed by several laboratories [2,3,10,15,16]. HPV-11 mRNAs expressed in cultured cells derived from laryngeal papillomas and expressed from transiently replicating HPV-11 genomes in SCC-4 cells have also been analyzed [5,13]. The types of analyses include cDNA cloning (species A), electron microscopy analysis of R-loops (species A–D, G, I, Q–S), and RT-PCR analysis (species C–G, I–P, R). The deduced mRNA structures are presented in the HPV-11 map. The 5' ends of most mRNAs have not been mapped at the nucleotide level. However, three promoters have been mapped by Smotkin et al. [20] using nuclease S1 analysis: the E6 promoter initiates at nt 90, the E7 promoter at nt 264, and the E1 promoter at nts 674–714. Dilorenzo and Steinberg [5] used an RNAse protection assay to confirm the locations of the E6 and E7 promoters, but mapped a slightly broader distribution of 5' ends from the E1 promoter to nt 677–726. In addition, they showed that the E1 promoter is differentiation specific. Renaud and Cowsert [13] used a modified RACE procedure to map a 5' end slightly further downstream to nt 743. The 5' end of the cDNA representing species A is located at nt 716 [10], so this mRNA is presumably transcribed from the E1 promoter. Renaud and Cowsert [13] also used 5' RACE to map additional 5' ends to the vicinity of nt 1374 (species E, K, O), suggesting that there may be an additional promoter in this region. The early poly(A) site has been identified from analysis of species A cDNAs; poly(A) addition sites were found at nt 4388, 4390, and 4392 [10]. The late poly(A) site has not been identified experimentally, but the L1 and L2 mRNAs are presumably polyadenylated utilizing the poly(A) signal at nt 7457.

HPV11





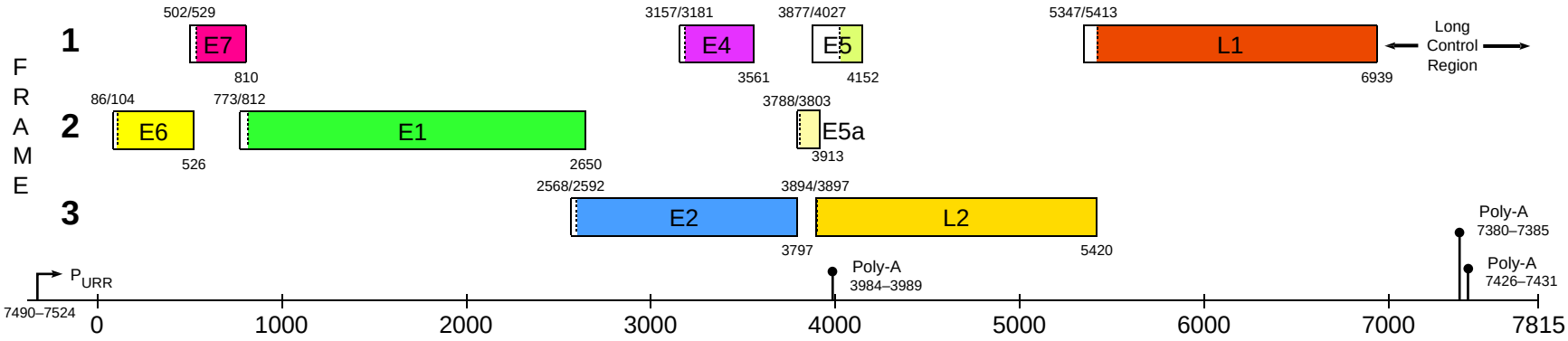
III-9
AUG 96

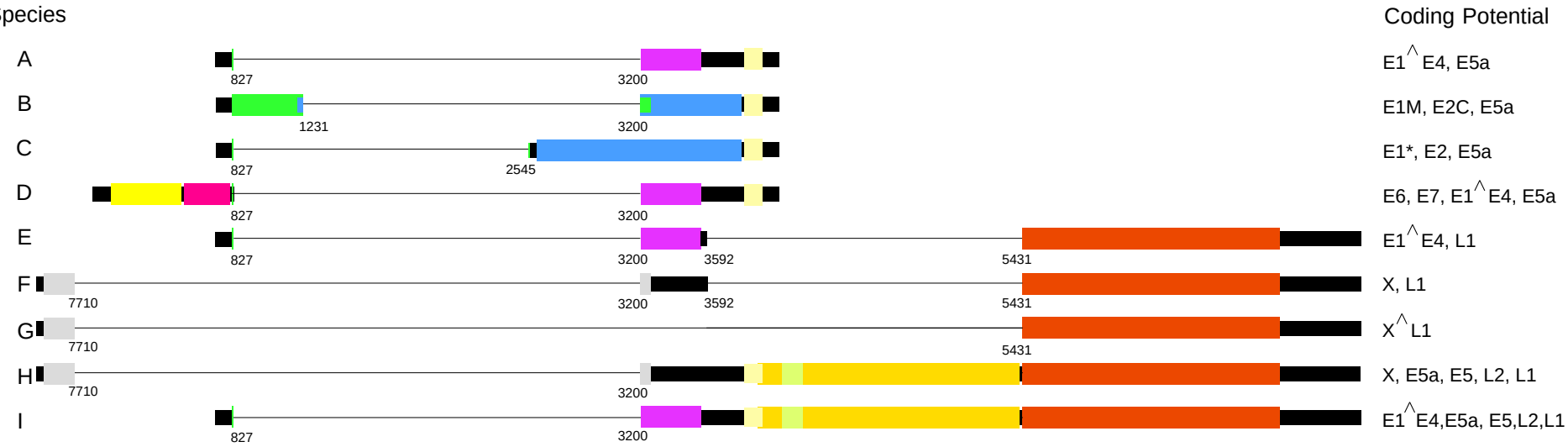
mRNA Transcript Maps

HPV-1a

HPV-1a mRNAs from plantar warts and cultured keratinocytes infected with HPV-1 were analyzed by electron microscopy of R-loops and by RT-PCR [4,11]. The deduced mRNA structures are shown in the facing map. There are three putative promoters, only one of which has been accurately mapped. The major promoter is located in the E7 ORF and presumably transcribes species A–C, E, and I. The most abundant mRNA in the plantar wart is species A which encodes the E1^E4 mRNA [4]. A minor promoter precedes the E6 ORF and presumably transcribes at least species D which can encode the E6 and E7 proteins. A very minor wart specific promoter was mapped to the URR by primer extension analysis and has start sites from nt 7490–7524 and major start sites at nt 7509, 7510, and 7511 [11]. The late mRNAs F–H are transcribed from this promoter. Species F, G, and H contain two weak translation initiation codons in exon 1 which, in species G, are in frame with the L1 ORF and could therefore encode longer L1 proteins. However, the most abundant L1 mRNA (E) is transcribed from the promoter in the E7 ORF. The potential peptides starting in exon 1 of species F, G, and H, here called "X," are shown as gray boxes because they correspond to none of the recognized ORFs of HPV1a. The early and late poly(A) sites have not been determined experimentally. However, polyadenylation signals (AAUAAA) are present at nt 3984–3989, 7380–7385, and 7426–7431.

HPV1a





III-11
AUG 96

mRNA Transcript Maps

Acknowledgements

We would like to thank Thomas Broker and Louise Chow for providing materials for the HPV-1a and HPV-11 maps and Michelle Ozbun and Craig Meyers for providing HPV-31 transcript data prior to publication.

References

[1] Bohm, S., S. P. Wilczynski, H. Pfister, and T. Iftner. 1993. The predominant mRNA class in HPV16-infected genital neoplasias does not encode the E6 or the E7 protein. *Int. J. Cancer* **55**:791–798.

[2] Chiang, C.-M., T. R. Broker, and L. T. Chow. 1991. An E1M^E2C fusion protein encoded by human papillomavirus type 11 is a sequence-specific transcription repressor. *J. Virol*. **65**:3317–3329.

[3] Chow, L. T., M. Nasseri, S. M. Wolinsky, and T. R. Broker. 1987. Human papillomavirus types 6 and 11 mRNAs from genital condylomata acuminata. *J. Virol*. **61**:2581–2588.

[4] Chow, L. T., S. S. Reilly, T. R. Broker, and L. B. Taichman. 1987. Identification and mapping of human papillomavirus type 1 RNA transcripts recovered from plantar warts and infected epithelial cell cultures. *J. Virol*. **61**:1913–1918.

[5] Dilorenzo, T. P. and B. M. Steinberg. 1995. Differential regulation of human papillomavirus type 6 and 11 early promoters in cultured cells derived from laryngeal papillomas. *J. Virol*. **69**:6865–6872.

[6] Doorbar, J., A. Parton, K. Hartley, L. Banks, T. Crook, M. Stanley, and L. Crawford. 1990. Detection of novel splicing patterns in a HPV16-containing keratinocyte cell line. *Virology* **178**:254–262.

[7] Grassmann, K., B. Rapp, H. Maschek, K. U. Petry, and T. Iftner. 1996. Identification of a differentiation-inducible promoter in the E7 open reading frame of human papillomavirus type 16 (HPV-16) in raft cultures of a new cell line containing high copy numbers of episomal HPV-16 DNA. *J. Virol*. **70**:2339–2349.

[8] Hummel, M., J. B. Hudson, and L. A. Laimins. 1992. Differentiation-induced and constitutive transcription of human papillomavirus type 31b in cell lines containing viral episomes. *J. Virol*. **66**:6070–6080.

[9] Kennedy, I. M., J. K. Haddow, and J. B. Clements. 1990. Analysis of human papillomavirus type 16 late mRNA 3' processing signals in vitro and in vivo. *J. Virol*. **64**:1825–1829.

[10] Nasseri, M., R. Hirochika, T. R. Broker, and L. T. Chow. 1987. A human papilloma virus type 11 transcript encoding an E1--E4 protein. *Virology* **159**:433–439.

[11] Palermo-Dilts, D. A., T. R. Broker, and L. T. Chow. 1990. Human papillomavirus type 1 produces redundant as well as polycistronic mRNAs in plantar warts. *J. Virol*. **64**:3144–3149.

[12] Pray, T. R. and L. A. Laimins. 1995. Differentiation-dependent expression of E1^E4 proteins in cell lines maintaining episomes of human papillomavirus type 31b. *Virology* **206**:679–685.

[13] Renaud, K. J. and L. M. Cowsert. 1996. Characterization of human papillomavirus-11 mRNAs expressed in the context of autonomously replicating viral genomes. *Virology* **220**:177–185.

[14] Rohlfs, M., S. Winkenbach, S. Meyer, T. Rupp, and M. Durst. 1991. Viral transcription in human keratinocyte cell lines immortalized by human papillomavirus type-16. *Virology* **183**:331–342.

[15] Rotenberg, M. O., C.-M. Chiang, M. L. Ho, T. R. Broker, and L. T. Chow. 1989. Characterization of cDNAs of spliced HPV-11 E2 mRNA and other HPV mRNAs recovered via retrovirus-mediated gene transfer. *Virology* **172**:468–477.

[16] Rotenberg, M. O., L. T. Chow, and T. R. Broker. 1989. Characterization of rare human papillomavirus type 11 mRNAs coding for regulatory and structural proteins, using the polymerase chain reaction. *Virology* **172**:489–497.

[17] Sherman, L. and N. Alloul. 1992. Human papillomavirus type 16 expresses a variety of alternatively spliced mRNAs putatively encoding the E2 protein. *Virology* **191**:953–959.

[18] Sherman, L., N. Alloul, I. Golan, M. Durst, and A. Baram. 1992. Expression and splicing patterns of human papillomavirus type-16 mRNAs in pre-cancerous lesions and carcinomas of the cervix, in human keratinocytes immortalized by HPV 16, and in cell lines established from cervical cancers. *Int. J. Cancer* **50**:356–364.

[19] Smits, H. L., M. T. E. Cornelissen, M. F. Jebbink, J. G. Van den Tweel, A. P. H. B. Struyk, M. Briet, and J. ter Schegget. 1991. Human papillomavirus type 16 transcripts expressed from viral-cellular junctions and full-length viral copies in CaSki cells and in a cervical carcinoma. *Virology* **182**:870–873.

[20] Smotkin, D., H. Prokoph, and F. O. Wettstein. 1989. Oncogenic and nononcogenic human genital papillomaviruses generate the E7 mRNA by different mechanisms. *J. Virol*. **63**:1441–1447.

[21] Smotkin, D. and F. O. Wettstein. 1986. Transcription of human papillomavirus type 16 early genes in a cervical cancer and a cancer-derived cell line and identification of the E7 protein. *Proc. Natl. Acad. Sci. USA* **83**:4680–4684.