

# The E4 Protein

John Doorbar<sup>a</sup> and Gerald Myers<sup>b</sup>

<sup>a</sup> *MRC Virology Unit, University of Glasgow, Glasgow, U.K.*

<sup>b</sup> *Theoretical Division, Los Alamos National Laboratory, Los Alamos, N.M., U.S.A.*

## Introduction

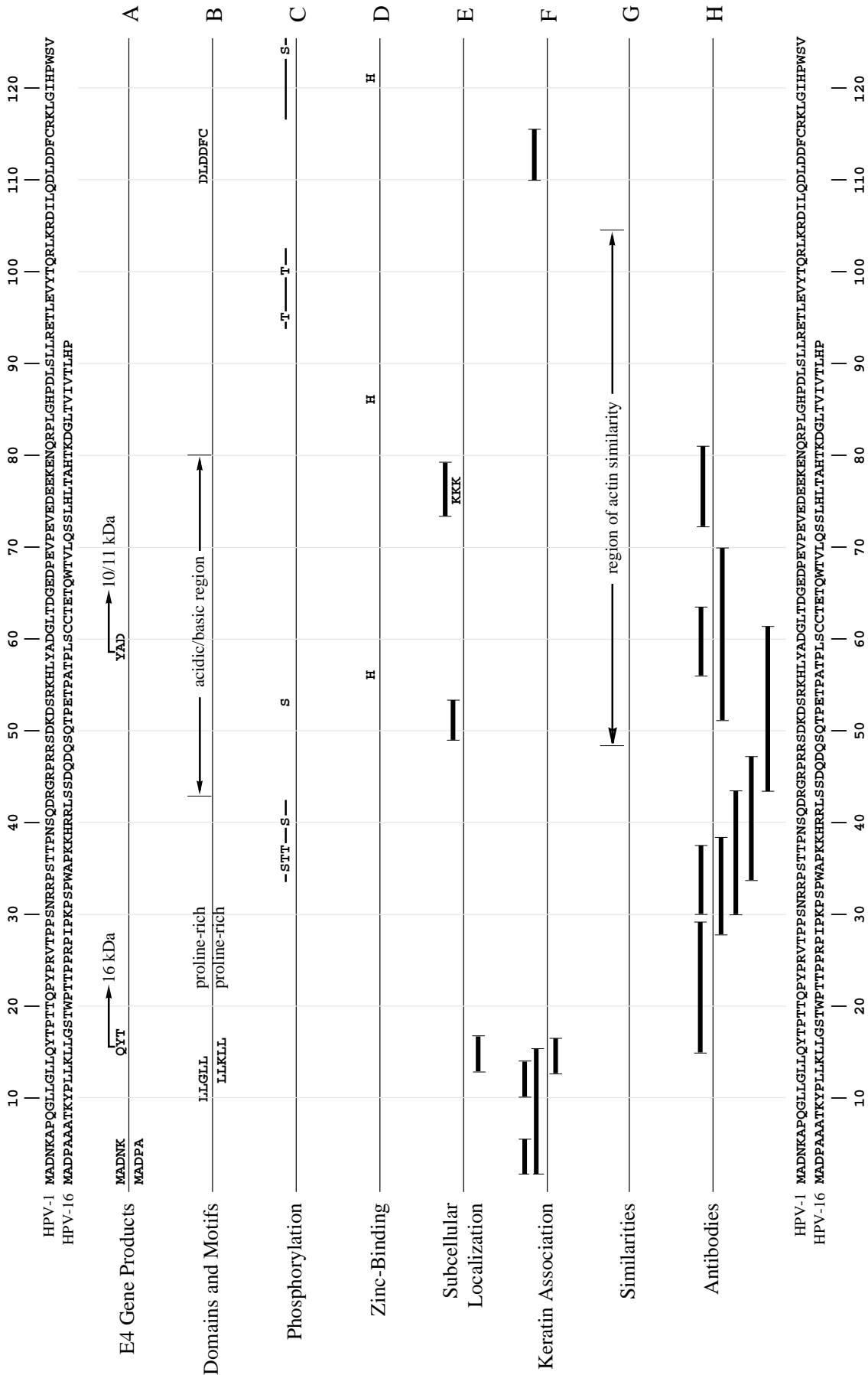
In plantar warts induced by HPV1, over 95% of viral transcripts initiate at a differentiation dependent promoter within E7 [1], and contain a single splice which joins the 5' end of E1 (16 nucleotides) onto the body of the E4 ORF. This mRNA encodes the primary E4 gene product (E1 $\wedge$ E4) which is an abundant, predominantly cytoplasmic protein of 17KDa [2, 3]. Progressive cleavage at the N-terminus gives rise to smaller species of 16K (-15 a.a.), 11K and 10K (-59 a.a.) [4, 5] while multimerisation, which requires the C-terminal half of the protein, produces dimers and higher order complexes. In HPV1-induced warts, E4 expression coincides with the onset of vegetative viral DNA replication and occurs immediately after cells have left the basal layer [2, 3, 6] (higher layers for mucosal types [7–10]). The smaller E4 species are derived from the full length protein during differentiation [3, 4]. E4 is phosphorylated *in vivo* [3, 11] and can bind zinc [12].

A characteristic feature of certain E4s is their ability to aggregate into cytoplasmic and nuclear inclusion granules [3,13]. Inclusions are most noticeable in cutaneous lesions, and have distinctive appearances depending on the infecting HPV type [14–16]. E4 is present in the inclusions of HPV1, 2 and 4 [13], but the exact composition of these structures has not been elucidated. Similar structures are produced following *in vitro* expression of the HPV1 E4 proteins in VX2R cells [17, 18]. Assembly of E4 into inclusions may be linked to E4 expression levels. E4 accounts for 20–30% of cell protein in HPV1-induced warts [2, 3]. Mucosal HPV types (HPV6 and 16), which have not yet been shown to contain prominent inclusions, express E4 at much lower levels [7–10, 19–22].

The majority of E4 in HPV1-induced tumours exists as soluble cytoplasmic complexes, which may be simple E4 multimers [3, 6]. Following expression in certain epithelial cell lines *in vitro*, both HPV 1 [23] and HPV 16 E4 [23, 24] can associate with the keratin intermediate filament network, and in the case of HPV16, can cause its collapse. The conserved N-terminus (MADXXA) and the leucine cluster (LLXLL) located nearby are important for this activity [5, 17]. Keratin association has also been seen in keratinised epithelium formed by grafting HPV16-containing keratinocytes [25]. Association may require additional cellular proteins or may be modulated by post-translational modification [5, 6, 17, 18].

The role of E4 in the virus life cycle is the subject of speculation, as functional analysis has been restricted by difficulties in mimicking productive infection *in vitro*. In warts, levels of E4 expression usually correlate with those of virion coat proteins [7, 9, 13, 20], and in the upper layers, E4 and L1 are expressed from the same bicistronic message [1, 26–28]. An important role in productive infection seems likely. Involvement in virus maturation [2] and/or vegetative viral DNA replication [3, 17, 29] (possibly by sequestering inhibitory factors) has been suggested. Alternatively, E4 may interfere with normal differentiation in order to create the conditions required for high level virus synthesis. E4 may interact with 'death proteins' [17] or may block effective translation of host protein [30]. Nuclear retention (parakeratosis) and thickening of the differentiated layers (acanthosis) are characteristics of virus infection. Association of E4 with keratins has suggested a role in virus release, *viz.* to increase the fragility of the infected cell [24]. Other possible functions of intermediate filaments, such as in RNA processing, signal transduction, protein degradation or cell differentiation may also be perturbed [30, 31].

The E4 gene is entirely contained within the central portion of the E2 ORF and is expressed from two abundant mRNA species during differentiation. The first encodes E1 $\wedge$ E4 and ends at the early polyA site: the second encodes both E1 $\wedge$ E4 and L1 and ends at the late polyA site. In cutaneous warts, E1 $\wedge$ E4 mRNA is present from the spinous layer upwards while E1 $\wedge$ E4-L1 is translated in higher



## E4 Protein

layers. E4 is highly divergent at the primary amino acid level between different viruses (Appendix A,B). Common characteristics are, however, apparent: all are proline rich (15% on the average, especially towards the N-terminus) and many have a central hydrophilic region containing a stretch of acidic and basic residues. Where E4 proteins have been examined *in vivo*, they have been found to share common physical properties [13]: acidity, occurrence as dimers, proteolytic processing. On the other hand, predicted sizes for E4 monomeric forms vary enormously—9 kDa in HPV-33 to 25 kDa in HPV-5—suggesting that homology is not all-encompassing (Appendix C). In the curatorial section that follows, the E4 proteins of HPV-1 and HPV-16 are compared (1E4 and 16E4), with annotation of HPV-1 above, and annotation of HPV-16 below, the horizontal lines. Information about other E4 sequences can be found in the appendices to this section.

There are few reviews which discuss E4 at any length, but see [30,31] and key papers [3, 5, 6, 13, 17, 18, 24].

### A. E4 Gene Products.

E4 proteins are usually found as multimers made up of various monomeric forms, the HPV-1 source for which is a 17 kDa protein made from a spliced E1/E4 mRNA; the five N-terminal amino acids, MADNK, are encoded from the 1E1 ORF [4,33]. The five N-terminal amino acids, MADPA, of the HPV-16 protein are similarly derived from 16E1 by splicing [34]. The coding potentials of HPV E4 ORFS, which overlap the coding region for the hinge sequence of E2 proteins, range from 73 amino acids (HPV-34) to 326 amino acids (HPV-21), and without evidence for intragenic duplication, one must be cautious about treating these ORFS as homologous. Among the larger E4s, for example HPV-21, an N-terminal elongated region is characteristically methionine-rich; splicing, however, will probably eliminate this methionine-rich stretch from many of the cutaneous types (for example, HPV-47 with a splice site at 3324).

The 1E4 protein is further processed by proteolytic cleavage to 16 kDa and to 11 and 10 kDa forms [3,4]. The N-termini of the cleavage events have been identified as Y16 and A59 when 1E4 was expressed in baculovirus [5]. Smaller E4 cleavage products, 10–11 kDa, have been detected in HPV-6 and HPV-11 [21] but not in HPV-16.

Early results suggested that the E4 protein of HPV-1 forms complexes [2,3] and this has recently been confirmed [6]. Multimerization appears to be mediated by sequences in the C-terminal domain of 1E4. While high molecular weight species have been observed also in HPV-2,-4,-6,-11 and -31 [8,13,19,21,22], they have not been reported for HPV-16: dimerization of the 1E4 gives 34, 32, 23, and 21 kDa species [3,4,5], and a probable hexamer/octomer of 105 kDa has been encountered [6]. HPV-2 and HPV-4 dimers of 33 and 40 kDa also exist [13], suggesting that post-translational processing is similar in at least the cutaneous viruses.

### B. Motifs and Domains.

An LLXLL motif is found at the N-terminus of supergroup A viral E4s (HPV-6, -11, -16, -18, -45, etc.), with minor exceptions being LLXLC (HPV-3, -28 and -19) and LLXLY (HPV-29). Other supergroup E4 proteins tend to possess leucines in the same region, but only some (e.g., HPV-1) have LLXLL. This motif is thought to be important for E4 association with cellular keratin and the formation of cytokeratin structures [5]. In certain PVs, splicing appears to eliminate any semblance of a leucine-rich motif (for example, HPV-47).

Downstream from the leucine-rich region in both mucosal and cutaneous viral E4s is a proline-rich region. The prolines apparently have no role in the formation of the filamentous networks [5], but may constitute a linker with the rest of the E4 polypeptide. The proline-rich region may extend through the entire central region of the E4 molecule, however in many types the central region is hydrophilic, and in a few types it is excessively hydrophilic: mutations in this charged region can, among other things, affect the distribution of E4 protein within the cell [5; E below].

The C-terminus, though poorly conserved, is thought to constitute another domain. Deletion of the DLDDFC sequence of 1E4 results in the collapse of the cytokeratin network [5]. In most cutaneous

papillomaviruses (i.e., supergroup B viruses), this motif is DLXDYW. No similar motif has been identified in mucosal viral sequences.

### C. Phosphorylation.

Several studies have indicated that 1E4 is phosphorylated and that phosphorylation increases as the infected cell migrates to the surface of the skin [3,11]. Some of the sites have been mapped [11]: four serines and threonines are present in the phosphorylated peptide 33–43; S53 appears likely to be phosphorylated; and peptides 94–102 and 117–125, which contain both threonines and serine, are also known to be modified. No modified tyrosines have been found. Grand and coworkers propose that gross overall charge changes, rather than mere conformational changes, are the likely result of the modification [11].

### D. Zinc-Binding.

HPV-1 E4 binds zinc and other divalent cations, although not to the extent that a true zinc-binding protein displays [12]. The one cysteine in 1E4 is not critical for zinc-binding, however histidines—H56, H86 and H121—are: Roberts et al. report that the amount of zinc bound is greatest with wild-type, next greatest with the single mutant H121A, and least (25% of the control level) in the triple mutant H56AH86AH121A [12]. The zinc-binding capability of E4 is not obviously related to dimerization or formation of cytokeratin structures.

### E. Subcellular Localization.

E4 is located in the cytoplasm for the most part, however intranuclear and perinuclear inclusions have been observed [14,29]. Mutations in the charged region of 1E4, residues 42–80 (see B above) result in nuclear localization and also “cytoplasmic spots” [5]; in particular, del49–53, del74–79, and D75KE76KE77K manifest this effect [5]. 16E4 del12–16 is exclusively intranuclear in about 80% of HPV-16 infected cells; similar deletions in 1E4 do not dramatically shift the localization [5].

### F. Keratin Association.

Deletion 2–15, which removes both the E1 contributed residues (A above) and the LLXLL motif (B above) abrogates the formation of filamentous networks [5]. Further deletion analysis by Roberts et al. reveals that the leucine-rich motif is crucial: del2–5, for instance, affects the morphology but not the formation of keratin associated networks. Del12–16 in 16E4, which eliminates the 16E4 LLXLL motif, also had a striking effect.

Association with keratin filaments in certain epithelial cells in culture was first identified in 1991 [24], and later confirmed by other work [5,23]. It is the only known function for E4. In some cell lines, however, E4 is either undetectable or it forms inclusion granules that don't appear to contain keratins [17,18]; the reason for this difference is the subject of speculation in the field. Some E4 can be found in association with keratin in vivo [25], but in naturally occurring cutaneous lesions much of the E4 is not obviously keratin associated [6]. There is no convincing evidence that E4 associates directly with keratin filaments or monomers [3,6,18]. Immunofluorescence staining in biopsies suggests that only a fraction of E4 is associated with keratin (2,3,8,9,13,18,20). The bulk of E4 in HPV1-induced lesions exists as soluble aggregates while the keratin network exists as an insoluble network [6,23]. Association may be modulated by post-translational modification of E4 (C, D above) and/or may require other cellular proteins.

Del110–115 (DLDDFC) in 1E4 leads to a collapse of the cytokeratin network, an effect routinely seen with wild-type 16E4 [5].

Inclusion granules, which are distinct from, and may appear in the absence of, keratin associated filaments, are a characteristic feature of cutaneous warts [14]. In HPV-1, -2, and -4 induced lesions, they have been shown to contain abundant E4 protein [3, 2, 13]. Inclusion granules have not been reported in mucosal lesions caused by HPV-6, -11, -16, and -31 [7,9,10,19,20,21,22,25]. High molecular weight

## E4 Protein

aggregates of E4 may be intermediates in the formation of inclusions [6], and E4 proteins in inclusions appear to have more post-translational modification than the soluble E4 fraction [6].

### G. Similarities.

The significances of E4 similarities are difficult to interpret because they rarely extend to other members of the E4 family. An interesting similarity reported for 1E4 is with actin [5]. Some similarities tend to be the consequence of merely low-complexity stretches of prolines or glutamates.

Early on, the E4 coding sequence of HPV-1 was said to have some similarity to Alu sequences [35]. It has a higher GC content than the rest of the viral genome.

### H. Neutralizing Antibodies.

E4 is an abundantly immunoreactive protein that has been studied in both healthy individuals and cervical carcinoma patients. Type-specific polyclonal responses are routinely observed (see reference 36, for example). Doorbar and coworkers raised four monoclonal antibodies to 1E4 protein which bound to distinct regions of the naturally occurring 16 kDa protein [4]. A major antigenic region of 16E4, residues 28 to 47, was mapped using bacterial fusion proteins and phage into which DNA fragments of HPV-16 were cloned [36]. Monoclonal antibodies to the same antigenic region of 16E4 have been raised [10]. A second active region of 16 E4 has been identified in cervical carcinoma patients as residues 51 to 70 [37]; the recognition of this region is thought to be the consequence of using human antibodies rather than antibodies raised to bacterial fusion proteins. A synthetic peptide corresponding to residues 43 to 62 of 16E4 identified a strong IgA response from sera of 30 patients with HPV-16 carrying cervical neoplasia [39].

### References

- [1] Palermo-Dilts DA, Broker TR, Chow LT. Human Papillomavirus type 1 produces redundant as well as polycistronic mRNAs in plantar warts. *J.Virol* **64**:3144–3149 (1990).
- [2] Doorbar J, Campbell D, Grand RJA, Gallimore PH. Identification of the human papillomavirus-1a E4 gene products. *EMBO J.* **5**:355–362 (1986).
- [3] Breitburd F, Croissant O, Orth G. Expression of Human Papillomavirus type-1 E4 gene products in warts. *Cancer Cells* **5**. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory press, 1987:115–122.
- [4] Doorbar J, Evans HS, Coneron I, Crawford LV, Gallimore PH. Analysis of HPV1 E4 gene expression using epitope-defined antibodies. *EMBO J.* **7**:825–833 (1988).
- [5] Roberts S, Ashmole I, Gibson L, Rookes SM, Barton GJ, Gallimore PH. Mutational analysis of human papillomavirus E4 proteins: Identification of structural features important in the formation of cytoplasmic E4/cytokeratin networks in epithelial cells. *J.Virol.* **68**:6432–6455 (1994).
- [6] Doorbar J, Medcalf M, Napthine S. Role of keratins in the formation of E4 complexes in vivo. *Virology* 1996;in press.
- [7] Crum CP, Barber S, Symbula M, Snyder K, Saleh AM, Roche JK. Coexpression of the human papillomavirus type 16 E4 and L1 open reading frames in early cervical neoplasia. *Virology* **178**:238–246 (1990).
- [8] Brown DR, Bryan J, Rodriguez M, Rose RC, Strike DG. Detection of Human Papillomavirus types 6 and 11 E4 gene products in condylomata acuminatum. *J.Med.Virol.* **34**:20–28 (1991).
- [9] Brown DR, Fan L, Jones J, Bryan J. Colocalization of human papillomavirus type 11 E1/E4 and L1 proteins in human foreskin implants grown in athymic mice. *Virology* **201**:46–54 (1994).
- [10] Doorbar J, Ely S, Coleman N, Hibma M, Davies DH, Crawford L. Epitope mapped monoclonal antibodies against the HPV16 E1/E4 protein. *Virology* **187**:353–359 (1992).
- [11] Grand RJA, Doorbar J, Smith KJ, Coneron I, Gallimore PH. Phosphorylation of the human papillomavirus type 1 E4 proteins in vivo and in vitro. *Virology* **170**:201–213 (1989).

- [12] Roberts S, Ashmole I, Sheehan TMT, Davies AH, Gallimore PH. Human papillomavirus type 1 E4 is a zinc binding protein. *Virology* **202**:865–874 (1994).
- [13] Doorbar J, Coneron I, Gallimore PH. Sequence Divergence yet conserved physical characteristics among the E4 proteins of cutaneous human papillomaviruses. *Virology* **172**:51–62 (1989).
- [14] Croissant O, Breitburd F, Orth G. Specificity of the cytopathic effect of cutaneous human papillomaviruses. *Clin. Dermatol.* **3**:43–55 (1985).
- [15] Egawa K. New types of human papillomaviruses and intracytoplasmic inclusion bodies: a classification of inclusion warts according to clinical features, histology and associated HPV types. *B. J. Dermatol.* **130**:158–166 (1994).
- [16] Egawa K, Delius H, Matsukura T, Kawashima M, Villiers de E-M. Two novel types of human papillomavirus, HPV63 and HPV65: Comparison of their clinical and histological features and DNA sequences to other HPV types. *Virology* **194**:51–62 (1993).
- [17] Rogel-Gaillard C, Breitburd F, Orth G. Human papillomavirus type 1 E4 proteins differing by their N-terminal ends have distinct cellular locations when transiently expressed in vitro. *J. Virol.* **66**:816–823 (1992).
- [18] Rogel-Gaillard C, Pehau-Arnaudet G, Breitburd F, Orth G. Cytopathic effect in human papillomavirus type 1-induced inclusion warts: In vitro analysis of the contribution of two forms of the viral E4 protein. *J. Invest. Dermatol.* **101**:843–851 (1993).
- [19] Brown DR, Chin MT, Strike DG. Identification of human papillomavirus type 11 E4 gene products in human tissue implants from athymic mice. *Virology* **165**:262–267 (1988).
- [20] Brown DR, Bryan JT, Rodriguez M, Katz BP. Factors associated with detection of human papillomavirus E4 and L1 proteins in condylomata acuminata. *J. Inf. Dis.* **166**:512–517 (1992).
- [21] Tomita Y, Fuse A, Sekine H, et al. Human Papillomavirus type 6 and 11 E4 gene products in condyloma acuminata. *J. Gen. Virol.* 1991;**72**:731–734.
- [22] Pray TR, Laimins LA. Differentiation-dependent expression of E1/E4 proteins in cell lines maintaining episomes of human papillomavirus type 31b. *Virology* **206**:679–685 (1995).
- [23] Roberts S, Ashmole I, Johnson GD, Kreider JW, Gallimore PH. Cutaneous and mucosal human papillomavirus E4 proteins form intermediate filament-like structures in epithelial cells. *Virology* **197**:176–187 (1993).
- [24] Doorbar J, Ely S, Sterling J, McLean C, Crawford L. Specific interaction between HPV16 E1/E4 and cytokeratins results in collapse of the epithelial cell intermediate filament network. *Nature* **352**:824–827 (1991).
- [25] Sterling JC, Skepper JN, Stanley MA. Immunoelectron Microscopical localization of the human papillomavirus type 16 L1 and E4 proteins in cervical keratinocytes cultured in vitro. *J. Invest. Dermatol.* **100**:154–158 (1993).
- [26] Nasser M, Hirochika R, Broker TR, Chow L. A human papillomavirus type 11 transcript encoding an E1/E4 protein. *Virology* **159**:433–439 (1987).
- [27] Chow LT, Nasser M, Wolinsky SM, Broker TR. Human papillomavirus types 6 and 11 mRNAs from genital condylomata acuminata. *J. Virol.* **61**:2581–2588 (1987).
- [28] Hummel M, Hudson JB, Laimins LA. Differentiation-induced and constitutive transcription of human papillomavirus type 31b in cell lines containing viral episomes. *J. Virol.* **66**:6070–6080 (1992).
- [29] Jareborg N, Burnett S. Immunofluorescent detection of bovine papillomavirus E4 antigen in the cytoplasm of cells permissive in vitro for viral DNA amplification. *J. Gen. Virol.* **72**:2269–2274 (1991).
- [30] Doorbar J. The E4 proteins and their role in the viral life cycle. In: Lacey C, ed. Papillomavirus Reviews. London: Royal Society of Medicine in press, 1996.
- [31] Doorbar J. An emerging function for E4. *Papillomavirus Report* **2**:145–148 (1991).

## E4 Protein

- [32] Fuchs, P.G., Iftner, T., Weniger, J. and Pfister, H. Epidermodysplasia verrucaformis-associated human papillomavirus 8: genomic sequence and comparative analysis. *J. Virol.* **58**:626–638 (1986).
- [33] Chow, L.T., Reilly, S.S., Broker, T.R., and Taichman, L.B. Identification and mapping of human papillomavirus type 1 RNA transcripts recovered from plantar warts and infected epithelial cell cultures. *J. Virol.* **61**:1913–1918 (1987).
- [34] Doorbar, J., Parton, A., Hartley, K., Banks, L., Crook, T., Stanley, M., and Crawford, L. Detection of novel splicing patterns in a HPV16-containing keratinocyte cell line. *Virology* **178**:254–262 (1990).
- [35] Danos, O., Katinka, M., and Yaniv, M. Human papillomavirus 1a complete DNA sequence: a novel type of genomic organization amongst papovaviridae. *EMBO J.* **1**:231–236 (1982).
- [36] Kochel, H.G., Monazahian, M., Sievert, K., Hohne, M., Thomssen, C., Teichmann, A., Arendt, P., and Thomssen, R. Occurrence of antibodies to L1, L2, E4 and E7 gene products of human papillomavirus types 6b, 16 and 18 among cervical cancer patients and controls. *Int. J. Cancer* **48**:682–688 (1991).
- [37] Muller, M., Gausepohl, H., de Martynoff, G., Frank, R., Brasseur, R., and Gissmann, L. Identification of seroreactive regions of the human papillomavirus type 16 proteins E4, E6, E7 and L1. *J. Gen. Virol.* **71**:2709–2717 (1990).
- [38] Suchankova, A., Krchnak, V., Vagner, J., Hamsikova, E., Krcmar, M., Ritterova, L., and Vonka, V. Epitope mapping of the human papillomavirus type 16 E4 protein by means of synthetic peptides. *J. Gen. Virol.* **73**:429–432 (1992).
- [39] Dillner, J. Mapping of linear epitopes of human papillomavirus type 16: the E1, E2, E4, E5, E6 and E7 open reading frames. *Int. J. Cancer* **46**:703–711 (1990).

### Appendix A: E1^E4 Splice Sites and Products

The following summary of E1 donor and E4 acceptor splice sites is based on predictions for the most part, the majority of which were first compiled by Tom Broker. We have extended Broker's list to include the twenty or so PVs that have been subsequently sequenced. The predictions are made on the basis of 1) agreement with the experimentally-determined PV sites; 2) phase fit to the E1 and E4 reading frames; 3) agreement with sites in phylogenetically-related sequences; and 4) consistency with emerging consensus sequences for these PV donor and acceptor splice sites.

A recent summary of eukaryotic donor and acceptor splice site sequences by Senapathy et al. reports DNA consensus sequences of AG/GTAAGT and CAG/G respectively (In *Methods in Enzymology*, Vol. 183, ed. R.F. Doolittle, pp. 252–278, Academic Press, San Diego, 1990). Rice and colleagues have generalized these sites as (A/C)AG/GTAAGT and YTTYNCAG/R (In *Sequence Analysis Primer*, ed. Gribskov, M. and Devereaux, J., pp. 1–87, W.H. Freeman and Co., New York, 1992). Nonconforming sites are known, and the PV sites appear to be such. For supergroup A PVs, the consensus donor site appears to be (A/C)AG/GTACNR and the acceptor site is RYRYNYAG/YA; for supergroup B sequences, the respective consensus sites are AAG/GTA(G/C)NR and GTYACYAG/YT.

Inferred sites are summarized by groups of PV types. The contribution of the donor site in E1 is represented below as an amino acid sequence, but the E4 acceptor site is represented as a nucleotide sequence. The amino acid residues created by the E4 acceptor site are shown in the HMM-generated alignment that follows; alignment of amino acid sequences that would be encoded by the unspliced ORFs is presented in Part II.

#### Summary of Inferred PV Splice Sites

Group	PV Type	Coordinates	E1 AA	Acceptor DNA Sequence
A1	32	865^3325	MADDT	ATATCTAG/CA
	42	844^3307	MADDT	ATATCTAG/CA
A2	3	821^3341	MILQ	GTATCTAG/CA
	10	806^3392	MDDNNT	GTATCTAG/CA
	28	803^3332	MDDTS	GTATCTAG/CA
	29	818^3341	MADNS	GTATCTAG/CA
A3	61	826^3322	MADSE	GTATCTAG/TA
A4	2a	827^3296	MEDSE	GTGTCTAG/CA
	27	837^3306	MEDSE	GTGTCTAG/CA
	57	825^3294	MEDSE	GTGTCTAG/TA
A5	26	890^3338	MDCE	GTATCTAG/TA
	51	886^3319	MDCE	GTATCTAG/TA
A6	30	905^3329	MASPE	GTGTCTAG/TA
	53	907^3343	MASPE	GTGTCCAG/TA
	56	910^3349	MASPE	GTGTCTAG/TA
	66	910^3331	MASPE	GTGTCTAG/TA
A7	18	1000^3434	MADPE	ATGTGCAG/TA
	39	943^3418	MANRE	ATGTGCAG/TA
	45	929^3392	MADPE	ATGTGCAG/TA
	59	887^3353	MADSE	ATGTGCAG/TA
	70	943^3433	MANCE	ATGTACAG/TA
A8	7	883^3340	MADDS	TCTCCTAG/TA
	40	883^3343	MADSP	TCTCCTAG/TA



## Summary of Inferred PV Splice Sites (cont.)

Group	PV Type	Coordinates	E1 AA	Acceptor DNA Sequence
A9	16	880^3358	MADPA	GTGTTT TAG/CA
	31	877^3295	MADPA	GTATTT TAG/CA
	33	894^3351	MADPE	ATATCT TAG/CA
	35	883^3298	MADPA	GTATTT TAG/CA
	35h	883^3319	MADPA	GTATTT TAG/CA
	52	879^3345	MEDPE	GTATCT TAG/TA
	58	898^3355	MDDPE	ATACCT TAG/TG
A10	6B	847^3325	MADDS	GTATCT TAG/CA
	11	847^3325	MADDS	GTATCT TAG/CA
	13	858^3327	MAEDT	GTATCT TAG/TA
	PCPV	849^3324	MADNT	GTATCT TAG/TA
	44	847^3307	MADNT	ATATCT TAG/TA
	55	844^3307	MADNT	ATATCT TAG/TA
A11	34	863^3338	MADS	GTATTT TAG/CC
B1	5	983^3322	MTDPNSK	GTCACC AG/CT
	8	966^3303	MADHK	GTTACC AG/CT
	9	932^3275	MSDNK	GTTACT AG/CT
	12	976^3316	MADSK	GTCACC AG/CT
	14	746^3089	MADPK	GTCACC AG/CT
	15	911^3244	MSDDK	GTTACT AG/CT
	17	920^3272	MTDDNK	GTTACT AG/CT
	19	985^3340	MAESE	GTCACC AG/CT
	20	1007^3350	MADPK	GTCACC AG/CT
	21	1013^3350	MADPK	GTCACC AG/CT
	22	947^3299	MADDK	GTTACC AG/TT
	23	965^3314	MADDK	GTTACT AG/TT
	24	914^3263	MADNK	GTCACC AG/CT
	25	689^3325	MADPK	GTCACC AG/CT
	36	973^3310	MADPK	GTCACC AG/CT
	37	917^3269	MTDDTK	GTTACT AG/CT
	38	926^3266	MADDK	GTTACC AG/TT
47	981^3324	MADSK	GTCACT AG/CT	
49	926^3281	MADDK	GTTACT AG/CT	
B2	4	825^3156	MADK	ATTGTT TAG/CT
	48	800^3110	MADHK	GTTACT AG/CT
	50	800^3128	MAELK	GTCACT AG/CT
	60	822^3180	MADPNK	GTTACC AG/CT
Others	1	827^3200	MADNK	TCCACT TAG/CT
	41	1005^3354	MKTKRMK	ACCGAC AG/CT
	54	843^3279	MADNK	GTATCT TAG/TA
	63	834^3231	MTDR	GTCAAT AG/CT
	RhPV1	953^3380	MDPE	GTGTCT TAG/CG
	BPV1	864^3225	MANDK	GATTTT TAG/AG
	BPV4	1014^3390	MDPK	GTTGGG TAG/TA
	DPV	809^3191	MDKENA	CACTCT TAG/AG
	EEPV	798^3180	MAETA	CACTCT TAG/AA
	ROPV	incomplete seq.		GTTACC AG/CT
	CRPV	1371^3714	MAE	GTCACC AG/CT
	COPV	831^3144	MAARK	ACCACC TAG/TG
	PCPV1	849^3324	MADNT	GTATCT TAG/TA
	BPV2	864^3222?	MANDK	GATTTT TAG/AG

most-likely	AP.....TQ.YPLLKLLP.....P...TPTPPRPPP.....	23
HPV54	V-KRH.....C-....A-....N...-..D.Q-I-	22
HPV32	--QPP.....K-R---D--NWY...K-P...-Y-T-HLH-Q.	31
HPV42	--HPP.....-R---D--SW...YNKC.A-QTHCT-QR.	31
HPV3	-H.....ER---D-C.....-V--KR--	20
HPV28	-H.....EK----D-C.....A-VT--RR-	21
HPV10	-H.....EK----D-C.....A-VT--QR--	21
HPV29	-H.....KK---D-Y.....T...P--T--AR--	22
HPV61	V-RTP.....CEK-----D.....T...CG-T-H----	27
HPV2a	--RPAP.....RT-NH---E--Y.....QSQ-QSQ-QQNQQE.	34
HPV27	--RPPP.....RT-NH---D-.....Y-QSQ-QHQHQQ.	31
HPV57	V-RPPP.....RT-TH---D--R.....-..QSQ-Q-Q-QQ.	30
HPV26	V-AAA.....-K-----S.....Q...YQ-----K.	26
HPV51	V-A.....ATR---Q--N.....N...YQ--Q--I-	24
HPV30	V.....SDPT-----S.....NT...--T-IK--	26
HPV53	V.....LDPT-----S.....-N-T-I--	23
HPV56	V-V.....D-T-----S.....-N-T-T--	23
HPV66	V-V.....D-T---R--T.....N...-T-TG-P--	25
HPV18	V-V.....-TR---S--N.....S...YS--HRI-	24
HPV45	V-V.....-TR---R--D.....S...YN--R--	24
HPV39	V-V.....-DR---N--.....N...YQ-----I-	24
HPV70	V-V.....T---S--Q.....N...YN--I-	24
HPV59	V-V.....-SK---D--S.....N...YH--QR--	24
HPV7	VL.....SKG---R-.....-SDI-...-T	21
HPV40	VL SRN.....D---R--.....-D--L-	22
HPV16	-A.....-K-----G.....ST...W--T---I-	24
HPV35h	-A.....-N-----H.....SY...-T---I-	24
HPV31	-V.....-K---G--Q.....SYQ.Q--T--HRI-	25
HPV52	VT.....K-----S.....-YA-K.P-R-	20
HPV33	-T.....K-----T.....YRQTITD.	19
HPV58	VI.....K-----Q.....-QR--	17
RhPV1	-LPTA.....-N-----A.....D...CN-STHH--	26
HPV6b	-LH.....KK--F-N--.....H...-HR--	20
HPV11	-L.....YEK---N--.....H...-H--	21
HPV44	VL.....CKT---G--.....H...-P--L	23
HPV55	VL.....CKT---G--.....H...-P--L	23
HPV13	VL.....YKK---G--.....H...-P--HR--	23
PCPV1	VL.....CKK---G--.....H...-LHR--	23
HPV34	-RV.....K-----D-CTQATA.	19
HPV19	-QHPPTHQE...DKE--.T-PPRPP.....-PL.-R-DS-Q-PENSHNKPTPKDG.	48
HPV25	--PPPSHQE...DKQ--.T-PPRPP.....L...-R-DS--QAPENSHNKAPRDRGGTNGG.	51
HPV20	--PPPTHQE...DKQ--.T-PPRPP.....-PL.-R-DS--ENSHNKPTPKGE.	45
HPV21	--PHPTHQE...DKQ--.T-PPRPP.....-PL.-R-DS--ENSHNKPTPKEE.	46
HPV14d	--PLPSHQE...DKQ--.T-PPRPP.....-PL.-R-DS--QIPENSHNKPTPKE.	47
HPV5	--RLQGRQE...DKQ--.T-PPRPP.....-P-Q--LT-R-DSSPHQNSHNKPKPEEESTDGGPPA.	57
HPV36	--PLQGRQE...DKQ--.T-PPRPP.....-P-Q--LT-R-SSSPHQNSHNKPKPEEESTDGGPPA.	57
HPV47	--HHQGHQE...DKQ--.T-PPRPP.....-P-Q--LT-R-DANPSINSHNKPKNEEESTDGGHQAE.	58
HPV12	--HHBHGQ...DKE--.T-FARPP.....-P-PT--D-VTNSHNKPTPGGKDTDDHLAEQ.	53
HPV8	--PPPDHQ...DKQ--.T-PPRPP.....-PL.-R-DSSG-LQNSHNKPKPKDEGTGDRPAE.	56
HPV24	--HHPT.....P...EGPENYPD.....-PL.-R-Q.AQ-NSHNKPVVTKQPSGRGTGEGSQA.	40
HPV15	-LRRQLE...KGQPPST-HPNRR.....TD...SFL---C--ENGHHHEPKPDA.	54
HPV17	-LRRQLEKQ...PTRPPST-HPGRHQ.....QG...DSL---C--EPHNGHHHG.	43
HPV37	-LRRQLEKQ...QTGQPPST-YPGRRS.....HG...DSL---C--ENGHHHGD.	44
HPV9	-LRHQLET...GERPPST-FPGRGR.....QQHRDSL---C--EDPGHHPDDTNEKPLA.	51
HPV22	VLRRQL...ESPPRT-HPNRH.....-P...-HNGHHKS-TDTAEK.	38
HPV23	VLRRQL...ETPPTT-FP-H.....-PC.P-HNGHH--TADTAE.	37
HPV38	VLRRQL...ETPP.T-HPGRH.....S...PSL---C--NGHHPKQHG.	36
HPV49	--PHHPRGYENPPTPAPFTT-STRHP.....APQ.Q-P--SA--QPQPQEHLSHHPKVPVPGKEGTGEK.	65
HPV4	-LH.....-P.P-STLRNN.....S...Y-G--AT-KL.	25
HPV65	-Q.....Q-P.P-STLRNN.....N...Y-G-Q.H--T.	23
HPV48	--R.....EI-.I--AG.....S...GSR--AARRA.	24
HPV50	--.....GVR-G--T.....N...S-GL-H-TTR.	22
HPV60	--QNN.....PPTTT-P-PGS.....NP...-SS-H-HRRL.	28
BPV1	EIA.....Q-ESGSHP-D-K.....E...-LQEK.K-SQ.	25
BPV2	EIA.....Q-E.F-HQ-D-K.....E...-LQEK.K-SQ.	24
EEPV	ET.....-E.TASKDSG.....A...-GSVEAREG.	22
DPV	ERP.....SKDCG.T-GAVKEA.....D...P--R-TA-CF.	27
BPV4	V.....ED.NAGP-PG.....-TPEDVADR-	21
HPV41	.....A--GRE	7
COPV	V.....PE...PPDRNS.....-V...-PRG-V-V-	22
CRPV	--PSRW.....SVPLKT-SP-RP.....-VQC--LKRK.	29
ROPV	-Q.....PP.-G.PS--A.....T...PL-T--R--	22
HPV1a	--.....G-G--Q.....Y...--TQ-Y-RV.	21
HPV63	--.....H.-G--G--Q.....T...PTQ--KDN--	22
MnPV	-QVSI.....SPhL-SPAK-RRVEEEEE...AKA--R-ASQ.	33

E4 Appendix A

most-likely	..... P P P P P . R P . W A G P K K P . . . . . T R G P P R R R R L . . . . . E . . S D S . . . . .	51
HPV54	..... H H . . . . . V . . . . . T T . - Q - Q . . . . . S R A . . . . . - - - - - . . . . . N E L . . . . .	43
HPV32	..... - S Q G V . T A . . . . . T Q T A Q T E Y . . . . . Y - K T - - - P P - R . . . . . - . . . . . N - T . . . . .	61
HPV42	..... - L . . . . . T . T T . . . . . T Q T V Q T E Q . . . . . H T - C . - S K P H - H . . . . . - . . . . . N - T . . . . .	59
HPV3	..... . . . . . K - R W A . - - . . . . . K D R S - S D . . . . . S D S . . . . . - S T . . . . . G S . - S . . . . .	46
HPV28	..... . . . . . K - R W A . - - . . . . . K D R S - N D . . . . . S D S . . . . . - H S T . . . . . G . - S . . . . .	46
HPV10	..... . . . . . K - R W A . - - . . . . . R D R N - S D . . . . . S D S . . . . . - S T . . . . . D . - T . . . . .	46
HPV29	..... . . . . . K - R W G . L R . . . . . R D R N G N D A . . . . . G L K Q S G L G H S S . . . . . S . - S . . . . .	52
HPV61	..... . . . . . - - - - - . A . - - - - - P - R H . . . . . P R C . . . . . - - - - - . . . . . I . - - - - -	49
HPV2a	..... . . . . . Q E E Q L R - - . . . . . K R . . . . . C - P - R R Q . . . . . R V . . . . . - P S . . . . . A . - V . . . . .	61
HPV27	..... . . . . . E Q E Q L R - Q . T C . . . . . C - P - R R H . . . . . R V . . . . . - P S . . . . . A . - G . . . . .	58
HPV57	..... . . . . . Q S . . . . . - - . . . . . H S R T P P R R . . . . . H R V . . . . . - H P S . . . . . A . - G . . . . .	53
HPV26	..... . . . . . - T . . . . . C . - - - - - R - - - - - . . . . . - H T Q . . . . . - - - - - . . . . . D . . . . .	45
HPV51	..... . . . . . L . . . . . P A . - - - - - - - - - - . . . . . H N S . . . . . - - - - - . . . . . N - X . . . . .	42
HPV30	..... . . . . . - - - - - . - - - - - . - - - - - T - - - - - . . . . . R P . . . . . P H G R . . . . . - - - - - . . . . . N V L . . . . .	46
HPV53	..... . . . . . - - - - - . - - - - - . - - - - - T - - - - - H . . . . . H P C . . . . . G - E . . . . . N . V P E . . . . .	44
HPV56	..... . . . . . - - - - - . - - - - - . - - - - - T - T . . . . . Q Y . . . . . P T . . . . . D . Q E N . . . . .	42
HPV66	..... . . . . . - - - - - . L . - - - - - T . . . . . R Y . . . . . P T . . . . . D . Q E N . . . . .	42
HPV18	..... . . . . . A . . . . . C . - - - - - Q R . . . . . - A . - - - - - . . . . . L . H - L . . . . .	44
HPV45	..... . . . . . K . . . . . H . - - - - - Q N . . . . . - S . - - - - - . . . . . L . - - L . . . . .	44
HPV39	..... . . . . . - Q . . . . . Q . . . . . H . - - - - - Q . . . . . S . - - - - - . . . . . - - - L . . . . .	43
HPV70	..... . . . . . - Q . . . . . Q . . . . . H . - - - - - L . . . . . S . - - - - - . . . . . A . - V E . . . . .	43
HPV59	..... . . . . . K . . . . . - T . . . . . - - - - - R G . . . . . - V . . . . . - - - - - . . . . . - - - Q . . . . .	44
HPV7	..... . . . . . - - T - - . C . T T . - P T . . . . . C . - - - - - P P K . . . . . Y . T T T . . . . .	45
HPV40	..... . . . . . - T - T . P . . . . . P Q R - P - R . . . . . S . A - P - H - P . . . . . - - - - - . . . . . E . . . . .	48
HPV16	..... . . . . . K . . . . . S . - - - - - H . . . . . - - - - - . . . . . S . - - - - - . . . . . Q . . . . .	41
HPV35h	..... . . . . . K . . . . . A . . . . . - - - - - Q . . . . . - - - - - . . . . . Q I T . . . . . N . D F E . . . . .	43
HPV31	..... . . . . . K . . . . . A . . . . . - - - - - V - V . . . . . C . G G - - - - - . . . . . L . - - Q . . . . .	46
HPV52	..... . . . . . - Q . . . . . C . - - - - - V . - - - - - T H . . . . . T Y N . . . . . H H . . . . . N . D . . . . .	40
HPV33	..... . . . . . . . . . . H H - Q . . . . . - - - - - . . . . . P N . . . . . D . D - L . . . . .	30
HPV58	..... . . . . . . . . . . - T T K . . . . . V . . . . . H - G . . . . . Q . . . . . - - - - - . . . . . D . . . . .	29
RhPV1	..... . . . . . - T . . . . . . . . . . P - - - - - R - T . . . . . C . G H - - - - - . . . . . Q . - E C . . . . .	43
HPV6b	..... . . . . . - L . . . . . C . . . . . Q - - - - - R - T . . . . . Q C . . . . . K - - - - - . . . . . G . N E H . . . . .	40
HPV11	..... . . . . . L Q . . . . . C . . . . . P - - - - - R - T . . . . . A C . . . . . - - - - - . . . . . G . - E H . . . . .	41
HPV44	..... . . . . . H R - H - H C . . . . . L - P - R T A . . . . . W - - - - - H V . . . . . N . D P E . . . . .	48
HPV55	..... . . . . . H R - H L H C . . . . . P - P - R N A . . . . . W - - - - - H V . . . . . N . D P E . . . . .	48
HPV13	..... . . . . . - Q . . . . . C . . . . . A . - - - - - R - N . . . . . V C . . . . . K - - - - - . . . . . V . N - N . . . . .	43
PCPV1	..... . . . . . A Q . . . . . C H . . . . . P S . . . . . Q - I . . . . . V C . . . . . K - - - - - P I . . . . . N . D F E . . . . .	44
HPV34	..... . . . . . . . . . . - T H R T R . . . . . V C . . . . . Q H G N . . . . . G . I - - - - - . . . . .	35
HPV19	..... . . . . . G T R D D L P A G P D D K . . . . . K . - A . R N . D Q G - N . . . . . P S P - R G - G - G . . . . . F . R L T G D H D P N	93
HPV25	..... . . . . . P P A G H D D S K P K R A R G D Q G - S - G - G . . . . . G P . S P A - V . . . . . S D - - R G - G - G . . . . . N L . - R L S G D Q D P	105
HPV20	..... . . . . . G T D G D L P V G Q G E Q - K . - A . R G D G P Q S . . . . . P S P S - G - G - G R . . . . . G . T G L G L G L G L	91
HPV21	..... . . . . . G T D G D R P V G P G E R - K . - I . K G - D R G . . . . . S P - R G - G - G R . . . . . G . - - P D P G P D P	90
HPV14d	..... . . . . . E G T D A D R P V G P G E R - K . - G . R G - D R G . . . . . S P - R G - G - G . . . . . G . - - L D P G R N R	92
HPV5	..... . . . . . S Q G D R K R S K G D Q G P D T G P G L G P G R G - S - K . - T . . . . . L G P - P G - G . . . . . P R - S - R L G P L Q . . . . . A . D R D P E E . . . . .	115
HPV36	..... . . . . . G Q G D R K R S K G D Q G P D T D P L G P D R G - S - G . - T - Q . P L - L P P . . . . . G L - - R - S P - - . . . . . G . - S G Y Q P D . . . . .	114
HPV47	..... . . . . . Q G D R K R T K G D P D P D P G R G P V L K P T L - - - - - . P . . . . . P T - - G L R R S T R L V L V - G Q G P P P . . . . . D . L P A . . . . .	116
HPV12	..... . . . . . G D R K R S K G D Q G R D T A P S L T P G R A - S - K . - G . . . . . L - P - P Y . . . . . G P P - - R - S H - - . . . . . G . T G G R D . . . . .	107
HPV8	..... . . . . . Q D R K K S R G D Q G R D T A P G L A P G R S - G L . G . - . . . . . L - P - P Y . . . . . G P - - R - S P - Q . . . . . F . G P G P D R D P E . . . . .	111
HPV24	..... . . . . . P L T P D A D D D P R P G K R S K G D E H G - A - G . - A . A - P L - L D L . . . . . D P P Q - G - D Q P P G . . . . . A . T G G . . . . .	102
HPV15	..... . . . . . T E K N L A L Q P - - - - - G G . - K D . K D K D - - - - - T Q . . . . . Q G D Q - - - - - Q G G D K . . . . . K . - - - - - P G . . . . .	86
HPV17	..... . . . . . D T G G K R L A L Q - - - - - P . . . . . G T K D - T S . . . . . D D Q - - - - - H G G D K . . . . . Q . - - - - - P G . . . . .	79
HPV37	..... . . . . . E E K H L A L Q - - - - - G K . K D K E - T - Q . . . . . Q G D Q - - - - - P G G N K . . . . . Q . P P G . . . . .	82
HPV9	..... . . . . . P P P G R K D R D K E K E K E K E K E K . K . . . . . T T - D - G . . . . . D P . R V E Q K P K . . . . . G . E G . . . . .	93
HPV22	..... . . . . . L V L Q S P - S G G K K G E . R D K D - - - - - Q . . . . . Q G E E K - D Q G P . . . . . - - - - - . . . . . A P . . . . .	74
HPV23	..... . . . . . K H L A L Q - - - - - G . G K . K D K E - - - - - S . . . . . P G E E K - D Q G P . . . . . G . A E . . . . .	71
HPV38	..... . . . . . D T G E K H L A L Q - - - - - A G . K G K D - E K . . . . . P Q A - K G E E K . . . . . A . D Q G . . . . .	71
HPV49	..... . . . . . T L V L Q Q P P T P G K R S R D D D P G L E P G - A D G K . - A . P Q - - - - - . . . . . A V P D - D P D P L P . . . . . - - - - - . . . . . D P E G P . . . . .	120
HPV4	..... . . . . . - S R R A . L L . . . . . E G - N R G N . . . . . P - - - - - P - P L K P R E Y D . . . . . Y - E . . . . .	57
HPV65	..... . . . . . - S L - R . - A L V V G - N R G N L . . . . . N R P P Q R - P K P - G Y . . . . . - - - - - . . . . . Y - E . . . . .	57
HPV48	..... . . . . . L E . G D . - A . S Q K T P T . . . . . S . - - - - - P - H P . . . . . D . Y E . . . . .	49
HPV50	..... . . . . . A N R K D . L E . . . . . A V . N Q - - - - - Y . . . . . R - P N H - P - H Q . . . . . Q . Y - F . . . . .	50
HPV60	..... . . . . . L . . . . . T . . . . . E D R - H - R . . . . . E S L A L - - - - - V . . . . . F D Y D A E . . . . .	55
BPV1	..... . . . . . - S L S L . L C . S . - - - - - P P A . . . . . V . . . . . P S . . . . . - - - - - . . . . . Q A . . . . .	45
BPV2	..... . . . . . - S L S L . L C . S . - - - - - P P A . . . . . Y . . . . . P S . . . . . - - - - - . . . . . Q A . . . . .	44
EEPV	..... . . . . . - T Q - T . E . . . . . C L T L L L D . . . . . N . . . . . P P F . . . . . V . A P . . . . .	44
DPV	..... . . . . . T L . . . . . L L . . . . . E - T - F T V . . . . . P S . . . . . E . . . . . A . K T G . . . . .	46
BPV4	..... . . . . . - D L - E . T . . . . . G . - - - - - S R G R . . . . . S - L R D - D H G H . . . . . D H . D R L . . . . .	50
HPV41	..... . . . . . - Q R Y Y . D R . . . . . R G R D D A E T . . . . . R K R G S R S P Q P L S . . . . . - - - - - . . . . . D E E . . . . .	38
COPV	..... . . . . . L - G K . G R . . . . . H G - L D G G . . . . . - - - - - . . . . . R G S P E G Q . . . . . - - - - - . . . . . D E E . . . . .	48
CRPV	..... . . . . . Q G - K . - - - - - V H . - - - - - D E G Q . . . . . G . . . . . H Q G . . . . . C . N E G . . . . .	51
ROPV	..... . . . . . L Q Y - Q . A . . . . . R T . I R - - - - - . . . . . R S S R Y - G - F . . . . . V . T - G G D P D P Q . . . . .	55
HPV1a	..... . . . . . T - S N . - R . . . . . P S T T P N S . . . . . Q D - R - - - - - S D K . . . . . D . - R K . . . . .	50
HPV63	..... . . . . . K L - E K - Q - R . . . . . R G R D T T R . . . . . N . . . . . - - - - - L F . . . . . A . - - - - - G . . . . .	46
MnPV	..... . . . . . I . . . . . - V . . . . . S L Q D - T T . . . . . G G N Q Q R - - - - - G . . . . . - - - - - . . . . . R G A . . . . .	59

most-likely	.....DSDSGEVEG.....	PTPTTPPAPTG.....	72
HPV54	E-TA.QTSN.....	H-APQT.....	57
HPV32	--LC.SHQQ.....	S-CS-T.-SQ-Y.....	80
HPV42	--VDSRHHS.....	TCS-QT--S-ASP.....	81
HPV3	N-S-NSNSN.....	NI-KP--RK-LN.....	67
HPV28	S.....	--KP--RK-LN.....	60
HPV10	S--K.....	.G-KI--RR-RN.....	61
HPV29	S.T-SSSSNR.....	-R--P--RK-VH.....	73
HPV61	.....TET-SS.....	S--QHKKTT-S.....	68
HPV2a	S----SI.....	-G--LRERSER.....	79
HPV27	S----SI.....	SG--LRERSER.....	76
HPV57	S----S.....	NS--LRGRSEK.....	71
HPV26	.....VDLTPP.....	S-QS-LS-QLPH.....	65
HPV30	EPQ-PT-Q.....	--PDS-L-ESPT.....	66
HPV53	P.Q-PT-L.....	TP-HS-LPQ-ES.....	63
HPV56	-P-Y.GNQN.....	L--PES-TQSVS.....	62
HPV66	.....PEQVQNQ.....	L--PES-THTVS.....	62
HPV18	.....TVD.....	SRRSS.....	53
HPV45	--VD.....	SQSST.....-D.....	55
HPV39	.....VQ.....	SQS-LS--E.....	56
HPV70	S.PD.....	PQKQT.....	51
HPV59	--VD.THS.....	--LSL.....	56
HPV7	ATHRP-S--.....	E-E-C-SVQW-DV.....	67
HPV40	E.TD.....	-CPS-LLWANHS.....	63
HPV16	.....Q-QTP.....	--A--LSCC-E.....	59
HPV35h	.....GVP.....	S-----SECDS.....	58
HPV31	E.Q-QST.....	-----TSCCEA.....	64
HPV52	Q.T-QTP.....	--S--TTFCD.....	58
HPV33	QTPQ.....	--PS-LQSCSV.....	45
HPV58	--IYQTP.....	T-S--QSIQ-A.....	49
RhPV1	V.GQTQ--.....	IQCG.....	55
HPV6b	E.E.....N.....	SPLA--CVW--L.....	56
HPV11	V.-R.....	PL---CVW--S.....	55
HPV44	.....PPQ.....	-----ET-SVS.....	63
HPV55	.....PPQ.....	-----GT-SVS.....	63
HPV13	EDLHVPL.....	--R-HK-LCVS.....	62
PCPV1	.....PPTVL.....	NSK--LTLCPV.....	62
HPV34	.....V.TQ.....	.....RG.....	41
HPV19	.....PEERPPPLE.....	HP-PPVTN--GH.....	119
HPV25	DPEEKP.....QPPEGEVQGHQPQPPVT-PQ.....	HL-PP-LP--N-HNDRD.....	149
HPV20	NRRAGGLGTDHD.....PDPEGESAPLPPPPQPPD-Q---	HP-PP--P-HN-RD..S.....	144
HPV21	GPiPGPLNRLTSRNTDSDPEGKCPSSLPPPPPPQPTTPE-QG--	HP-PP--P-NGHGD.....	152
HPV14d	LSGGLGTD.....QDPDPDKKCPESQPPPE-----	HP-PP-NGHNGH.....	135
HPV5	.....GQPPEAE-----	HPGG.DQGHP-PP---HN-H.....	147
HPV36	.....HDPEAPLE-----	GGH..GHHP-PP--P--N.....	144
HPV47	.....PPVE-----	HPQKDRDHP-P--QNGHGK.....	145
HPV12	.....RNPEE-G---	HP--P-LSGGDP.....	129
HPV8	DGLQP.....PLGEGQVEGHPG-GDQPQ.....	HP-P--SNGHK.....	147
HPV24	.....V.GETPP.....	GNEESQP--GE.....	120
HPV15	.....EGT-ADGD.....	D-EK--S--P-E.....	106
HPV17	.....EGSDASGDEN.....	A--E--QD---E.....	103
HPV37	.....EGTDADGDEN.....	A--E--V---E.....	106
HPV9	.....G-E.....	-P-Q--LP---E.....	112
HPV22	.....SGEG-PPDD.....	S-EN-QN--G.....	95
HPV23	.....NGGG-KPKD.....	P-EE-QN--G.....	92
HPV38	.....P.EAPTG--GTP..	GD-P-ED-QS--P-EG..E.....	99
HPV49	.....E.-L-QPP-I.....	-A-RE--GAE.....	140
HPV4	.....-EK-NQ.....	-GQEK---KEEE.....	77
HPV65	.....-DK-NQ.....	-QGER---KEEE.....	77
HPV48	.....-DENR-N.....	LE-P--H-ED.....	67
HPV50	.....EDD-K-N.....	TI--DTESHQON.....	70
HPV60	.....-PT-.NK-N.....	YP-ESR-V-KDA.....	75
BPV1	.....V.GY-T-LA.....	R---IF.LQAR.....	64
BPV2	.....V.GY-T-LA.....	R---IF.LQAR.....	63
EEPv	.....E.LAKTGV-P.....	F-ARL-T-HHHP.....	65
DPV	.....VGPL.....	-ARL-T-HHSP.....	61
BPV4	.....RRGRTP-DE.....	TRGYRV-GD-RE.....	71
HPV41	.....LT-A..D.....	P-RR-N-G-RRRL..F.....	57
COPV	.....--EE-A-N.....	YP-SRS.R-RR-R.....	69
CRPV	.....R.Q..N-N.....	RP-R-K.....	63
ROPV	E.....LDSTQQ-PEDK-NIP.....	--S-PT-S--P.....	83
HPV1a	.....H.LYADGL.....	TDGED-EV-EVE.....	69
HPV63	.....PT-E..-.....	.G-EV-EI--SD.....	62
MnPV	.....RTP-P-TTA.....	QR-KR-RRAC-RK.....	81

E4 Appendix A

most-likely	.EE.....GEVEGG.....PPPWTV.....	86
HPV54	.....-A.....	61
HPV32	.....-P.....	84
HPV42	.AH.....-L.....	87
HPV3	..-H.....VNH--L.....	75
HPV28	..-Q.....VNH--L.....	68
HPV10	..-R.....VDK--.....	69
HPV29	..-R.....VDQ--.....	81
HPV61	.....G--.....	72
HPV2a	.....K-S.....	83
HPV27	.....K-S.....	80
HPV57	.....R-S.....	75
HPV26	SPD.....SQ--I.....	73
HPV30	.QT.....VSTQ-T.....	74
HPV53	.PT.....QS-SQ-T.....QTTQT-T.....	79
HPV56	.QH.....THSA.....	69
HPV66	.QQ.....TQTS.....	69
HPV18	.....IVDL.....	57
HPV45	.....VSTP-C.....	61
HPV39	.....C--I.....	61
HPV70	..-.....CS--L.....	57
HPV59	.....ACQ--T.....	62
HPV7	V--.....NT--L.....	75
HPV40	..-.....ST--L.....	70
HPV16	.....TQ--.....	64
HPV35h	.....V--.....	63
HPV31	.....T--.....	69
HPV52	.....NN--.....	64
HPV33	.QT.....--I.....	52
HPV58	.....--DH.....	55
RhPV1	.....--DH.....	58
HPV6b	.....D--.....	61
HPV11	.....D--.....	60
HPV44	..-T.....AT--.....	71
HPV55	.DT.....AT--.....	71
HPV13	.QT.....TT--.....	70
PCPV1	.RT.....VE--.....	70
HPV34	.....--L.....	45
HPV19	.PQPLPPQPPC.....DQEGAAGS-AAG--AAGSGAVGG..EGNDPESH--P-PNGHE.....	170
HPV25	PGQ.....-AAG-AAAGAGDD.....DTDCLN--PSDHEE.....	180
HPV20	CG-.....-A-G-ADKEQG.....EGDHH--PPPQNGH.....	173
HPV21	H--GP.....LE-A-G--DGH.....PP--APPNGHE.....	180
HPV14d	..-.....-A-G-AGGDG.....HPP--PPNGHD.....	160
HPV5	SGHEPKVQQPEGPEGREGHEEGAVGGEG-DE--H.....--PPPPTNG.....	191
HPV36	.GHECGPKPQGPE.....GREGVEGPGV-A-G--DED.....DH--PPPPTNG.....	187
HPV47	..-T.....QGA--GDKGEQGAVGGESSDGEGDHSQPPLTPPN.....	182
HPV12	.DHPPPTPENGH.....NGEKEDGEK-A-G-SDET.....DH--PPPQNDP.....	170
HPV8	..-G.....D--E--AVGGDG.....NDHPP--PPNVQ.....	176
HPV24	.G.....-A--H.....--PPPPPEPKPH.....NG.....	143
HPV15	.G.....-PQQGPN.....RSPSHD-D-DPSRD.....	134
HPV17	.G.....-A-G-PQHGP.....SRG-S-DPDRGHD.....	128
HPV37	.G-.....-PR.....HD-NQGSPSHDPGR.....G.....	130
HPV9	.G-.....-PR.....PG-S-VP-PAPTPG.....R.....	137
HPV22	.....-A.....-S-GPAQGRD.....	112
HPV23	.....-S-APDQDP.....	108
HPV38	G--.....-TA--GRS.....PARDQDP SH.....	120
HPV49	.G.....-H.....--PPPVNGKEGAAGQGGES.....	168
HPV4	..-.....E-E-ER.....N.-DL.....	90
HPV65	..-.....-E-ER.....D.-SL.....	90
HPV48	.....E-QR.....GN.-GP.....	76
HPV50	.....-EP.....	73
HPV60	..-Y.....TKS-D.....	83
BPV1	.....ALFS.....AL-PPR.....	74
BPV2	.....ALFS.....AL-PPR.....	73
EEPV	.RG.....-PWA.....-I-PPR.....	77
DPV	.RG.....-AWA.....-I-PPRCRAR.....	77
BPV4	.....-D--A.....--NGND.....	82
HPV41	L--.....TEDRL.....	65
COPV	RRL.....HNK-DQNIN.....	81
CRPV	.....RIL.....	66
ROPV	.....-T-P-SR.....	90
HPV1a	..-D.....-K.....NQRPL.....	79
HPV63	..-.....K-NR.....-E-LP-V.....	75
MnPV	E-T.....PSSG--GGRRRRGL..AGRPEAVPRTT-RP.....	113

most-likely	...QTLSSVT..L..TWE.TFDGLVV...IVQDLEDY....WKKLHL	119
HPV54	...T-TGT---I..-TR.-K--TQ...V-.T-.....--	86
HPV32	...DRKG-HL-.I..-I..-S--TR-E...-R-.....RQ	108
HPV42	...DCVG-EL-.V..KTV.-S--TT...E-R-.....R-	112
HPV3	...-GPGT--...RVH.-PS-IE...TL-TV.....C-	99
HPV28	...QGPAT--...HVV.-PS-TQ...TL-TV.....--	92
HPV10	...-GPGC--...EVH.-PS-TQ...TL-TV.....--	93
HPV29	...-GPGT--...QVK.-PT-TQ...-L-TV.....--	105
HPV61	...L-SG-T--...V..-A..QKQ-TT...T-.TV.....--	96
HPV2a	...T-SGA---...-AQ.-PG-AT...TL.T-.....C-	108
HPV27	...T-KGA---...-AQ.LPG-TT...TL.T-.....C-	105
HPV57	...K-TGA---...-AQ.-PG-AT...TL.T-.....C-	100
HPV26	...--TTYT-Q..V..EAI.-RE-TR...VTK-.....C-	98
HPV30	...--TD-ALVE...HVT.-QKST--...-K-.....--	99
HPV53	...PENT-L-E...RVT.-PKST--...-R-.....--	103
HPV56	...TQTIPT-E..V..EVS.-TTTTL...-R-.....R-	93
HPV66	...TTPT-E..V..HVS.-HKAT--...-K-.....R-	91
HPV18	...STHF-Q..-HLQAT.-K-NS...V-.T-.....R-	83
HPV45	...-TR-C-Q..V..QVT.-KE-KC...V-.T-.....R-	85
HPV39	...L-TH-T--...V..QAT.-Q--TS...V-.T-.....R-	86
HPV70	...LQ..V..KAA.-N--TS...V-.T-.....--	76
HPV59	...V-TQ--C..I..QAT.-R--TSL..A-.T-.....R-	87
HPV7	...E-EHARLI...KAT.-KS-T--...E-.I-.....--	100
HPV40	...-EHARL...KAT.-GT-T--...E-L-.....--	95
HPV16	...-Q--LH...-AH.-K--T--...-T-.....-P	87
HPV35h	...L-EG-TLH...-AQ.-KT-V--...V--.....--	88
HPV31	...S-VGL--Q...-HAQ.-KQ--S...VL-.....--	94
HPV52	...LHGD--LQ...-SAQ.-K--HI..QLV-.....--	89
HPV33	...EQHVLO...-AQ.-SS--C...VL.T-.....--	75
HPV58	...EEEDYT-Q...-VH.-KG-TC...VLK.F.....-SC..I	83
RhPV1	...KAGQ-F-D...HTT.-LQ-VP...T-.TI.....R-	83
HPV6b	...E-TT-L-.I..-TS.-K--TT...T--.....R-	86
HPV11	...-STT--L-.I..-TS.-KE-TT...T--.....R-	85
HPV44	...--TT--L-.V..-TV.-K--TTI...--.....R-	96
HPV55	...--TT-TL-.V..-TV.-K--TTI...F--.....R-	96
HPV13	...--TT-TL-.I..-TI.-K--TT...T--.....--	95
PCPV1	...KTT-TI-.I..-TTT-SN-TT...T-V.V.....--	95
HPV34	...Y...I...-Y..SPQOThRP..S-LYI.....M-R...Q	66
HPV19	...DS-G--G-MAS.-LTQ--HQ--TR--EE...QE-----R-LIP...Q	212
HPV25	...S-G--G-MAS.-LTQ--HQ--TQ--QD...QE-----R-LIP...Q	221
HPV20	EGS-S--GNMAS.-LL---QQ-TQ--QD...QE-----M--SIP...Q	216
HPV21	...ES-S--GN-AS.-LT---SL-NQ--QE...QV-----T--SIP...Q	222
HPV14d	...ES-S--G--AS.-LV---SH-NQ--QE...QE---G---T--SIP...Q	202
HPV5	...HEGG----AS.-LVK--GH--Q--QS...QD-----ATP...Q	233
HPV36	...QES---GA-AC.-LTK--SH--L--QN...QD-----SIP...Q	229
HPV47	ESDGS--NT-AC.-LAR--SN--Q--QN...QG---G---R--GTP...Q	225
HPV12	EQGLG-AGN-AC.-LSK--DL-NL--QN...ED--G---STP...Q	213
HPV8	...EGS--GC-GS.-LLK--DQ-NL--QN...QG-----T--STP...Q	217
HPV24	DATHG--GT-AS.-LG---ES-RQ--EE...QE--D---RR-SIP...Q	186
HPV15	...HEG--HG-AS.RLQ---AQ--H--EN...LG--K---R--GTP...Q	175
HPV17	RDPEG--PG-AL.RLSK--NQ--Q--ET..L-G--Q---GIP...Q	171
HPV37	RDPEG--PG-AL.RLTQ--SQ-EQ--ET...-D--K---T--GIP...Q	173
HPV9	GPEEG--PGLAS.RLMK--HE--Q--QD...-TG--H---LR-KTP...H	180
HPV22	PVHES--TG-AS.RLTK--QH--Q--DS...-G--RN---TQ-KTP...Q	155
HPV23	...DH-S--QG-AL.HLVK--RH--Q--DT..V-E--RN---MQ-KTP...Q	150
HPV38	...EP--QG-AY.RLTK--RQ--Q--DK..V-E--RG---QT-QTP...Q	160
HPV49	LFLEG-A-.....RLTR-DQEYKQ--DD...-LD---G---RR-AIL...Q	207
HPV4	...HH.....-LQK-GADI-K-KDK..VCR--DS...KQ--GIR...L	123
HPV65	...RH.....-LGK--SDIEQ-KDK..VCR--DN...KL--GIH...P	123
HPV48	...--HQ.....-LRK-DQDLQR-QDT..VTH--D---R--GIR...H	111
HPV50	...--RQ.....-LKK--EDLTQ-QRM..VC---D---KR--GIHQ..S	109
HPV60	...RVPWGLPQ...-LKR--ADI-LFLEA..VY---Q-F...KE--GILQ..SC	124
BPV1	...CRAR..YRW--HQGRKKRSSRP.TP.....RKN...Q	101
BPV2	...CRAR..YRW--HQGRKKKINRPTQ-RR.....N-	101
EEPv	...ARAR..Y..R..FCYQDHQ...-QRRR.....RRT-QT	102
DPV	...YRR-PGAYLY..P..-V..LDE-RRI...TRR.R.....QINT..T	106
BPV4	...A-EHRLRQ.-LTK--DDLQR-RDK..LRL--.....S-	113
HPV41	...TS--E-L-KDIESDI-HFERK-R-L..LQ-KD.....TI	97
COPV	...YEPPAA...PEDD--D-CKKLT...-P..F.....LF	106
CRPV	...-PGTSD..RLQRT.LDEE-RR...LEEH-PGGI...DGFAS-	100
ROPV	...PP-DHLLQ.R..LE..EIRQ-QES..LQE--EEFGNLYLR-GIR...Q	131
HPV1a	...GHDDL-LL..RETLEVY-QRLKRD...-L--D-F...CR--GIHPWSV	120
HPV63	...ENGWH-FL..RETLEH.QLGR-QRE..VN--F--L...YRR-GIH...P	113
MnPV	...SGGLGLP..R..QSSYPARPPI...S-R.-YPV...TGGRGEQP..S	148

**Appendix B: Secondary Structure Prediction from E4 Sequences**

Protein sequences such as the E4 sequences that display less than 30% similarity might nevertheless be shown to have similar structures. In general, we tend to learn more about structure from dissimilar (but homologous) proteins than from highly similar proteins.

This appendix summarizes the secondary structure predictions over the E4 HMM-predicted sequence as determined by several different algorithms, Gibrat, Levin, DPM, and SOPMA. Two consensus structures are also reported, one based on the four different algorithms, the other (at the top of the print-out) based on individual E4 sequences as analyzed by the SOPMA method. The derivation of an HMM model sequence ('most likely sequence') is discussed in appendix A and elsewhere in Part III (Farmer and Myers). The various methods for secondary structure prediction are also discussed elsewhere in Part III of this compendium.

The structural code encompasses lower and upper case letters for alpha-helix (h, H), beta-sheet (e, E), turns (t, T), and random coil (c, C). States that are predictable with greater confidence are shown in upper case. The criteria for designating a state as upper or lower case are spelled out in the general discussion of this approach in Part III: states that are predicted in upper case letters have i) scores that are equal to or greater than the median average for scores assigned to that state over all positions and ii) scores that are in the upper quartile of difference from the second highest predicted state. Hence the absolute and the relative scores must meet stringent requirements to warrant upper case prediction.

hvp_e4_splice.SOPMA	ccc.....cc.cccceccc.....c...cCCCcCCCCC.....	24
Gibrat_E4_splice	--.....-H.-HHH-E--.....	23
Levin_E4_splice	-S.....S.-..EE--.....S-S.....	23
DPM_E4_splice	.....-C--.....	23
SOPMA_E4_splice	--.....TT.-EEE-EE.....	23
Consensus_E4_splice	--.....-EE--.....	23
HPV54	E--TC.....-EE-EE.....	22
HPV32	---CC.....-C--HE-EHHC...C-C.....	31
HPV42	---CC.....-C--EHHC...C-CC.....	31
HPV3	.....-C-EEE-EE.....E----	20
HPV28	HH.....T.-EEE-EE.....	21
HPV10	HH.....-..-EE-EE.....	21
HPV29	-T.....-C-EEE-EE.....E.....	22
HPV61	E--CC.....-C--HE-EE.....E...T-----	27
HPV2a	---CCC.....CC-C-EE-EE.....-.....CCCC.....	34
HPV27	---CCC.....CC-C-EE-E.....-.....CCC.....	31
HPV57	---CCC.....CC-C-EE-E.....-.....	30
HPV26	---HC.....-..-EE-EEE.....E...E-----	26
HPV51	--T.....-C--HHHH.....-.....	24
HPV30	--.....CC--EEE-E--.....-C.....	26
HPV53	--.....CC--EE-EE.....EE-----	23
HPV56	--.....C--EE-EE.....	23
HPV66	--T.....C--EE-EEE.....	25
HPV18	--T.....-C--HE-EH.....-.....	24
HPV45	---T.....-C-TEE-EEE.....-.....	24
HPV39	--T.....-C--E--.....T.....	24
HPV70	---.....C--E-EHE.....-.....	24
HPV59	---.....-C--EEE.....-T.....	24
HPV7	EE.....-C-EE-EE.....E-----	21
HPV40	EE-TT.....-..-EE-EE.....	22
HPV16	HT.....TT.-EE-EE.....-C.....	24
HPV35h	T.....-C-EE-EE.....-C.....	24
HPV31	H.....-H.-EEE-EE.....-CC.....	25
HPV52	E.....-..-EE-EE.....E...E-----	20
HPV33	E.....T.-EE-EE.....E...E-----	19
HPV58	EE.....T.-EE-EE.....E-----	17
RhPV1	---CC.....-EE-EE.....	26
HPV6b	H.....-C--E-EE.....E...T.T---	20
HPV11	HH.....H-C-EE-EE.....E...T-----	21
HPV44	EE.....-C-EE-EE.....E.....	23
HPV55	E.....-C-EE-EE.....E.....	23
HPV13	EE.....-C-EE-EE.....E.....	23
PCPV1	E.....-T--E-EE.....-.....	23
HPV34	-T.....-..-EE-E.....-HHH.....	19
HPV19	T-CCCCC...CCC--C--C--CC-----TCCCCCCCCC.....	48
HPV25	---CCCCC...CCC--C--C--.....CCCCCCCCCCCCCCCC	51
HPV20	---CCCCC...CCC--C--C--CC-----CCCCCCCCC.....	45
HPV21	---CCCCC...CCC--C--C--CC-----CCCCCCCCC.....	46
HPV14d	---CCCCC...CCC--C--C--CC-----CTCCCCCCCCC.....	47
HPV5	---CCCCC...CCC--C--C--.....CCCCCCCCCCCCCCCC	50
HPV36	---CCCCC...CCC--C--C--.....CCCCCCCCCCCCCCCC	50
HPV47	---CCCCC...CCC--C--C--.....CCCCCCCCCCCCCCCC	50
HPV12	---CCCCC...CCC--C--C--.....CCCCCCCCCCCCCCTT	48
HPV8	---CCCCC...CCC--C--C--CC---T-T---CCCCCTCCCCCCTCC	53
HPV24	---CCC.....-C--C--C--CC-----CCCCEEEECCCCCCCC	46
HPV15	HHHHHT...TCC-C--C--C--EC-----CCCCCCCCC.....	44
HPV17	EE-CCCCC...CCC-C--C--C--.....C-----CCCCCCCC.....	43
HPV37	EE-CCCCC...CCCC-C--C--C--.....C-----CCCCCCCC.....	44
HPV9	EHHCHCTT...CCC-C--C--CTT-----CCC-----CCTCCCCCCCCCCCC.....	51
HPV22	---CEC.....CC-C--C--C--C-----CCCCCE.....	38
HPV23	EE-CCT.....CC-C--C--C--CC---T-----CCCHC.....	37
HPV38	E-ECEE.....CC--C--C--.....-.....CCCCCCC.....	36
HPV49	---CCCTCCCCCCCC-C--C--C--CC-----CCCCCEHHCCCCCCCC	58
HPV4	EE.....-T-E-E.....-.....C.....	25
HPV65	TT.....T--..-E-.....-.....	23
HPV48	-TT.....EEE.EEEEH-T.....-HHHHHH.....	24
HPV50	-T.....T.-EE-EEE.....-H.....	22
HPV60	---CC.....-EEEE-C--.....C-----EEE.....	28
BPV1	EE.....C-CT--T--H...HEE--T--.....	25
BPV2	E.....C--C--C--.....-E-----	24
EEPV	-H.....H-----CTT-----EEE-----	22
DPV	---.....CCC--CE-.....-E.....	27
BPV4	---.....-T-C--.....-H-T--.....	21
HPV41	---.....-T--.....	7
COPV	--.....-C--.....C-----E-----	22
CRPV	---CCC.....EEEEC--C--.....-EE-----	29
ROPV	---.....-EE-.....E.....	22
HPV1a	-T.....T.-EE-EEE.....E-----EE.....	21
HPV63	EE.....E.EEEEC-----	22
MnPV	-EEEE.....EEC--CEEEHHHC-----T--HH.....	33



E4 Appendix B

hvp_E4_splice.SOPMA	.....CCCCc.Cc..cCcccCC.....cc...cccc	45
Gibrat_E4_splice	.....-CC-EEEE	47
Levin_E4_splice	--S-----S-----	47
DPM_E4_splice	.....-CC-----	47
SOPMA_E4_splice	.....-CC-HHHE	47
Consensus_E4_splice	.....-CC-EE	47
HPV54	.....C-H...HHHH	39
HPV32	.....EE-CC-EE	57
HPV42	.....C..CC--CC	55
HPV3	.....T-...-EE	41
HPV28	.....T-...-EE	42
HPV10	.....C-...-EE	42
HPV29	.....E-...-H...HEEEEEEE-E-	48
HPV61	.....C--...-EEE	45
HPV2a	.....CHC.....EE...-H	57
HPV27	.....CCT--E.E-...-TT-...-E...-EE	54
HPV57	.....TC...CEE	49
HPV26	.....TC...-EE	41
HPV51	.....TC...-EE	38
HPV30	.....TC...-EE	42
HPV53	.....C..C-...-EE	40
HPV56	.....C..C-...-EE	38
HPV66	.....E..E-...-EE	38
HPV18	.....T-...HHHH	40
HPV45	.....T-...HHHE	40
HPV39	.....T-...THHH	39
HPV70	.....T-...T-EE	39
HPV59	.....T-...-E...EEEE	40
HPV7	.....C-...-EE	41
HPV40	.....C-...-EE	44
HPV16	.....C-...-EE	37
HPV35h	.....C-...-EE	39
HPV31	.....-EEE...H..TT-EE	42
HPV52	.....C-...-EE	36
HPV33	.....C-...-EE	26
HPV58	.....E..E-...-EE	25
RhPV1	.....-HHH	39
HPV6b	.....-HHH	36
HPV11	.....-H...HH...HHHH	37
HPV44	.....C-...-EE...E-..	44
HPV55	.....C-...-EE...E-..	44
HPV13	.....EH...HHHE	39
PCPV1	.....E-...-EE	40
HPV34	.....H--E...EE...-EE	31
HPV19	.....CCCCCCCCCCCC	83
HPV25	.....CCCCCCCCCCCCCCCC	94
HPV20	.....CCCCCCCCCCCC	81
HPV21	.....CCCCCCCCCT	80
HPV14d	.....CCCCCCCCCT	82
HPV5	CCTCCCC.CCCCCCCCCCTCCCCCCCCCCCCCCCC	108
HPV36	CCCCCCC.CCCCCCCCCCCCCCCCCCCCCCCCC	106
HPV47	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCEEC	112
HPV12	HHHHT...TCTCCCTCTCCCCCEEECCCC	101
HPV8	CTH...CCCCCCCCCCCCCTTCTCCCTT--EE.EE.E-EEEE..CC-CC-EE	101
HPV24	CCCC...CCCCCCCCCCCCCCCCCCCC	98
HPV15	.....CHHHHHEECC--T--C-...-CC-CC-EE	82
HPV17	.....CCCCCEEE--T-...-C-CC-EE	75
HPV37	.....CCCEEE--T-...-C-CC-EE	78
HPV9	.....CCCCCCCCCTHHHHHTT--T-...-CC-CC-EE	89
HPV22	.....EEEECC--C-...-T-...-C-CC-EE	70
HPV23	.....CEEE--T-...-C-CC-EE	67
HPV38	.....CCTCHHEEE--T-...-TT-...-CCCT-EE	67
HPV49	CCCCTTC.EEEEECCCCCCCCCTTCTCCCC--T-...-T-...-C-CC-EE	114
HPV4	.....-HE.EE..TT-...-C-CC-EE	49
HPV65	.....-E.EEEEE--T-C..CC-CC-EE	52
HPV48	.....HH.H-H-...T-...-CC-EE	45
HPV50	.....H-HH.HH..HH-...C..C-CC-EE	46
HPV60	.....EE...HHHHCC-EE	49
BPV1	.....-EE.EE-...-EE	41
BPV2	.....EE.E-...-EE	40
EEPV	.....EEEEET-...-EE	40
DPV	.....EE...EE.E-...-EE	42
BPV4	.....-T-T-...T-...HHHHC-EE	45
HPV41	.....-EE-...-TTT-C..CC-CC-EE	34
COPV	.....-CC-CC-EE	44
CRPV	.....-T-...EE.EE-...-T-...-T-EE	47
ROPV	.....-CTC-EE	45
HPV1a	.....TTTCCC-EE	46
HPV63	.....HHHHHT-...-HH-H...E...EE-EE	42
MnPV	.....-E.EEET-...-C-CC-EE	55

hvp_E4_splice.SOPMA	....c.ccc.....cccccccc.	58
Gibrat_E4_splice	...E.----.-----EEE-	60
Levin_E4_splice	...-.....-T-----	60
DPM_E4_splice	...-.....-T-----	60
SOPMA_E4_splice	...-.....-T-----	60
Consensus_E4_splice	...-.....-T-----	60
HPV54	...H.HHH.....H----	51
HPV32	...-.....-T-----	69
HPV42	...T-----E-----	68
HPV3	...EC--H.....--HH----	55
HPV28	...-.....-T-----	49
HPV10	...-.....-T-----	50
HPV29	...-..-HH......H--H--C	61
HPV61	...E.E-----T-----	57
HPV2a	...H.HH-----T-----	67
HPV27	...-.....-HE-----	64
HPV57	...-.....-T-----	59
HPV26	...-.....-EEE-----	53
HPV51	...-.....-T-----	41
HPV30	...-.....-EE-----	54
HPV53	...-.....-EEE-----	51
HPV56	...-.....-T-----	50
HPV66	...-.....-T-----	50
HPV18	...H.HEE.....HH-----	48
HPV45	...E.EEE.....EH-----	48
HPV39	...H.--H.....HH-----	47
HPV70	...E.EE-----T-----	46
HPV59	...-.....-T-----	51
HPV7	...-EE-----T-----	54
HPV40	...-TT-----T-----	51
HPV16	...-.....-T-----	48
HPV35h	...-.....-T-----	47
HPV31	...E-----T-----	53
HPV52	...-.....-T-----	47
HPV33	...-.....-T-----	34
HPV58	...-..-TT.....-EEE-ET-	37
RhPV1	...H.-H-----E-----E	50
HPV6b	...-.....-T-----	44
HPV11	...T-----T-----	44
HPV44	...-..-T-----T-----	52
HPV55	...-.....-T-----	52
HPV13	...E.-TT.....-E-----	51
PCCV1	...-.....-T-----	51
HPV34	...-.....-E-----	38
HPV19	...E.EEECCCC.....CCCC--E----	107
HPV25	...EE.E--CCCCCCCC.....CCCCCCCC--E----	132
HPV20	...E.EEEEECCCCCCCCCCCC.....CCCCCCCCCCCCCCCC--EE--	129
HPV21	...-..-CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC--E----	138
HPV14d	...-..-E-CCCCCCCCCCCC.....CCCCCCCCCCCC--E----	123
HPV5	...-..-CCC.....CCC-----C	128
HPV36	...-..-CCCT.....CCCC--T--EE--C	128
HPV47	...-.....-T-----C	126
HPV12	...-..-CC.....C-----T-	117
HPV8	...-..-CCCCCTCTCCC.....CCCCCCCC--T-----T-	135
HPV24	...-.....-T-----	109
HPV15	...-.....-T-----	94
HPV17	...-..-T-----T-----C	89
HPV37	...-.....-T-----C	92
HPV9	...-.....-T-----	99
HPV22	...-.....-T-----	83
HPV23	...-.....-T-----	80
HPV38	...-..-T-----T-----C	80
HPV49	...-..-CC.....E-----	129
HPV4	CCCE-.....-HT-T--	65
HPV65	C.....-T-----	65
HPV48	...-.....-HHH-HHH.	57
HPV50	...E.EEE.....-T-----	58
HPV60	...EEEE--.....-E.T--	63
BPV1	...T.--E.....E.--TEEEE.	53
BPV2	...-.....-EEE-----	52
EEPV	...-.....-H.H-THH--C	53
DPV	...-..-T-----T-----	50
BPV4	...TC.-T-----T-----	59
HPV41	...-..-HH.....H-----	43
COPV	...-..-HHH.....HHHHHHHT.	57
CRPV	...-.....-T-----TT.	57
ROPV	...E.EE-CCCC.....CCCC--T-----	71
HPV1a	...T.TTH.....H.EE-TT--	57
HPV63	...-.....-T-----	51
MnPV	...-.....-T-----	68

E4 Appendix B

hpv_E4_splice.SOPMA	.....cCccCCcCCccc.....cc.....	CEECCc	72
Gibrat_E4_splice	.....	CEECCc	80
Levin_E4_splice	.....	SEEECC	80
DPM_E4_splice	.....	CCCCC	80
SOPMA_E4_splice	.....-THH.....HH.....	CCCCC	80
Consensus_E4_splice	.....	CCCCC	80
HPV54	.....		57
HPV32	.....		80
HPV42	.....C.....		83
HPV3	.....-E.....		69
HPV28	.....-E.....		62
HPV10	.....		63
HPV29	.....-E.....		75
HPV61	.....-EEEE.....		68
HPV2a	.....T--EE--TH--TT.....		79
HPV27	.....T--E--TTTT.....		76
HPV57	.....-EEE--TTT.....		71
HPV26	.....-C.....C.....		68
HPV30	.....-C.....-E.....		68
HPV53	.....E.....-T.....	CEECCc	72
HPV56	.....		64
HPV66	.....-EEE.....		64
HPV18	.....-TT--.....		53
HPV45	.....--T.....-E.....		55
HPV39	.....		56
HPV70	.....		52
HPV59	.....EEE--.....		56
HPV7	.....-EEEEEE.....EHH.....		70
HPV40	.....-EEEE--T.....T.....		65
HPV16	.....		59
HPV35h	.....		58
HPV31	.....-H--.....		64
HPV52	.....		58
HPV33	.....-E.....E.....		47
HPV58	.....-EE.....		49
RhPV1	.....EE--.....		55
HPV6b	.....-EE--T.....		56
HPV11	.....-E--T.....		55
HPV44	.....-TE.....EE.....		65
HPV55	.....		65
HPV13	.....-EEEE.....-T.....		64
PCPV1	.....-EEE.....		64
HPV34	.....		41
HPV19	.....-T--.....-CCCCCCCC.....CCTCCHCHHHHHCCCHETCEEE		150
HPV25	.....-CCCCC--.....CCCCC.....CCTCCHCCCCC		165
HPV20	.....CC..CC.....CETCCCCCCCC		158
HPV21	.....CC..CTTC.....EEEECCCCC		168
HPV14d	.....HH.....HHCCCCC		147
HPV5	CCC.CCC--C.....C--CCTCCCCCCCCCCCCCCCCCTTCCCCCTCC.....		181
HPV36	CC..CC.....CCCCCCCC.....CCCTCCCCCEEECCCCC		175
HPV47	CCCCCTC.....CCCCCCCCCTT		160
HPV12	.....-T.....-CCCCCCCC.....CCCCCTTCTCECCCCC		158
HPV8	.....-TT.....-C.....TTHHEEETCC		162
HPV24	.....	CEECC	128
HPV15	.....C.....	CCCCCCCCC	120
HPV17	.....C.....C.....	CCCCCCCCC	115
HPV37	.....C.....C.....	CCCCCCCC	116
HPV9	.....-TTC.....	CCCCCCCC	122
HPV22	.....	CCECC	102
HPV23	.....-T--TT--.....E.....	EECTC	99
HPV38	CC..CC--CC..CC.....	CHHTCCCC	111
HPV49	.....-T--TT.....	HHHTCC	148
HPV4	.....-HHHH.....HH.....	HHHHH	85
HPV65	.....-HTH.....HT.....	TTHHC	85
HPV48	.....	HHCH	71
HPV50	.....		70
HPV60	.....-T--EE.....		77
BPV1	.....E--EE.EEHT.....	CEEE	68
BPV2	.....-EE.EEHT.....	CEEE	67
EEPv	.....HHHH.....	CCCC	71
DPV	.....	CEEE	67
BPV4	.....-HE-E--T--H.....	CCTTC	76
HPV41	.....-EE.EHHH.....		60
COPV	.....-T--E-C.....CEE.....		72
CRPV	.....		63
ROPV	.....		83
HPV1a	.....ETT--T--H--.....HH.....	HHH	74
HPV63	.....	CCCC	68
MnPV	.....-EE-C.....T--.....	CCCCCCCCCTTE	97

hvp_e4_splice.SOPMA	.....cccccc	.....ccccceEE	.E..ecc.ccTccee	...eeH	100
Gibrat_E4_splice	.....EE	.....EEEE	---EH.HHCHH	---H-	108
Levin_E4_splice	.....E	.....EEEEHHHC	..HH-HHC-H	..HH-	108
DPM_E4_splice	.....	.....CCC	.C.C-	---C-	108
SOPMA_E4_splice	.....EE	.....EE-CC	---EE.E-T	---H-	108
Consensus_E4_splice	.....E	.....EE-CC	---E-H-C	---H-	108
HPV54	.....EE	.....EET	---	---	82
HPV32	.....	.....-T-C	---E.E	---C-E	105
HPV42	.....-EEE	.....EEE	---EE	---C	109
HPV3	.....EEEEEE	.....-T	---EE	---	95
HPV28	.....E-EEE	.....-T	---E	---	88
HPV10	.....TTEEEE	.....-T	---EE	---T-C	89
HPV29	.....ETEEEE	.....-TT	---EE	---E	101
HPV61	.....TEEE	.....EETT	---	---T	92
HPV2a	.....-EEE	.....EETT	---	---EE	104
HPV27	.....-EEE	.....E-TT	---E	---T-C	101
HPV57	.....-EEE	.....EETT	---	---E	96
HPV26	.....-TEE	.....TTEEE	---HHH	---T-T	95
HPV30	.....EE	.....-H-E	---E	---T	96
HPV53	.....C-E	.....-T	---E	---C-C	100
HPV56	.....	.....	---C	---C	90
HPV66	.....-E	.....E	---H	---H-C-E	88
HPV18	.....EEEE	.....	---EEC	---T-C	79
HPV45	.....EE	.....-E	---EE	---T	81
HPV39	.....-EE	.....EEETE	---	---	82
HPV70	.....TEEE	.....	---C	---T	72
HPV59	.....-EE	.....E-E-C	---	---	83
HPV7	.....HHHHH	.....HHHHH	---HHH	---T	96
HPV40	.....-TEE	.....EHHHH	---E	---T-C	92
HPV16	.....T-EEE	.....E-HH	---C	---TE	83
HPV35h	.....-EEE	.....EEE-C	.C.	---C-E	85
HPV31	.....-EEE	.....EEEE	---H	---E	91
HPV52	.....-EE	.....EET-CCC	.H.	CH-	86
HPV33	.....	.....	---C	---	71
HPV58	.....-EEEEC	.....T-HTE	---	---EE-T	77
RhPV1	.....-EE	.....ET-E	---	---C	79
HPV6b	.....-EE	.....E-E	---	---C	83
HPV11	.....-EE	.....EEEE	---	---C	82
HPV44	.....-EEEE	.....EETT	---	---EE-ET	93
HPV55	.....-EE	.....E-T-E	---	---E	93
HPV13	.....-EE	.....EETTE	---	---EE.E	92
PCPV1	.....	.....	---E-C	---C	92
HPV34	.....-EE	.....E	---	---E-C-TCCC	60
HPV19	TT..CCCCCTCC	-----CCCC	.....CCTHHHHHHHH	HHHHHHHHHHHHHHHH	199
HPV25	.....CCCCC	-----CHH	.....HHHHHHHHHH	HHHH-HCHHHHHHHHH	208
HPV20	.....CCCCC	-----CCCC	.....CCC-HEHTHHHH	-EEC-CHHHHHHHHH	203
HPV21	.....CC	-----CCCC	.....HHHHHHHHHH	HHHH-HHHHHHHHHHH	209
HPV14d	.....CCC	-----CCCC	.....CHHHHHHHHHH	-EE-C-HH-HHH	189
HPV5	.....	-----CCCC	.....CHHHHHHHHHH	HHHH-CHHHHHHHHH	220
HPV36	.....CC	-----CCCC	.....CHHHHHHHHHH	HHHH-HHHHHHHHHHH	216
HPV47	CCCTCCCCCCCC	-T-----CCCC	.....CCC-HHH-HHHE	-HHHHHHHHHHHHHHHH	212
HPV12	.....CC	-----CCCC	.....THHHHHTHHHH	HHHHHHHHHHHHHHHH	200
HPV8	.....CCCC	-----CCH	.....THHHHE--HC	HHHHHTHHHHEE-ST	204
HPV24	.....	-----CCCCCCC	.....CCCCCHHHHHHHHH	CHHHHHHHHHHHHHHH	173
HPV15	.....CCCCC	-----HH	.....HHHHHHHHHH	HHHHHHHHHHHHHHHH	162
HPV17	.....TCC	-----CCCC	.....CCC-EEE	-E-HHC-T-HHHHHHH	158
HPV37	.....CC	-----CCCCC	.....CCCC	---TC-E-HHH-HHHHHHHHH	160
HPV9	.....CC	---EE-CCCCC	.....CCCC	---CHHH.HHH-HHHHHHHHH	167
HPV22	.....	-----CCCC	.....CCHHHHHHHHHH	HHHHHHHHHHHHHHHH	142
HPV23	.....	-----CCT	.....TC-HHHHHHHH	-TEC-C-HHHHHHH	137
HPV38	.....CC	-----C	.....HHHHHHHH	HCH-HHHHHHHHH	147
HPV49	.....	-----CCCCCHHTHTHHEEHHHHHT	.....TCCC	-HHHHHHHH	194
HPV4	.....HH.HHH	.....	.....HH	.....HHHH-H-HHHHHHH	110
HPV65	.....-H.HHH	.....	.....HH	.....HHHH-T-HHHHHHH	110
HPV48	.....	.....	.....-HHH	.....HHHHHHHHHHHHHH	98
HPV50	.....	.....	.....-H	.....HHHH	95
HPV60	.....TT	.....	.....	-----CHH.HHHHHHHHHHHHHHH	109
BPV1	.....E	.....	.....-CHC	.CCE-E-C-TCTTCCC.CCC	97
BPV2	.....E	.....	.....-CCC	.CEE-E-C-C-CCCCTCCC	97
EEPv	.....	.....	.....-CCC	.C.CE.EECT-CC	94
DPV	.....	-----CCCC	.....CC	-----C.HH.HHNT-CC	100
BPV4	.....-TT	---	.....	.....HHHHHHHH.HHHHHHHHHHHHHHH	108
HPV41	.....HHHHH	.....	.....	.....HHHHHHHHHHHHHHHHHHHHHH	93
COV	.....E	---TCC	.....C	-----CTCHH.HHHHHHC	103
CRPV	.....EEE	.....	.....-CC	.SEHHH.HHHHHHH	88
ROPV	.....-C	.....	.....-EE	-H.H.HHH.HHHHHHHHH	114
HPV1a	.....	.....	.....-HHH	.....HHHHHHHHHH--HH	104
HPV63	.....-EETH	.....	.....HHHHHHHH	.....HHHHHH.HHHHHHH	100
MnPV	EE..TTCTCCTT	-----	.....-E-CC	.C.C-C-C-C-C	135

E4 Appendix B

hpv_E4_splice.SOPMA	hh.....ecc...c	106
Gibrat_E4_splice	--HHH...HHHHHH	119
Levin_E4_splice	--HHH...HHHHTH	119
DPM_E4_splice	CCCCC...CCCC-	119
SOPMA_E4_splice	--HTC...HHHH-E	119
Consensus_E4_splice	--HHC...HHHHH	118
HPV54	EE.....--	86
HPV32	.E.....--	108
HPV42	.E.....-E	112
HPV3	EE.....-E	99
HPV28	EE.....-E	92
HPV10	EE.....-E	93
HPV29	EE.....-E	105
HPV61	EE.....-E	96
HPV2a	EE.....-E	108
HPV27	EE.....-E	105
HPV57	EE.....-E	100
HPV26	.E.....HH	98
HPV30	.E.....CH	99
HPV53	.E.....--	103
HPV56	.E.....CH	93
HPV66	.E.....-E	91
HPV18	EE.....-E	83
HPV45	EE.....-E	85
HPV39	EE.....-E	86
HPV70	EE.....--	76
HPV59	EE.....-E	87
HPV7	EE.....-E	100
HPV40	.E.....HH	95
HPV16	EE.....C-	87
HPV35h	.E.....C-	88
HPV31	.E.....-E	94
HPV52	.E.....--	89
HPV33	EE.....-E	75
HPV58	.E.....-E-C..-	83
RhPV1	EE.....HH	83
HPV6b	.E.....--	86
HPV11	.E.....--	85
HPV44	.E.....-E	96
HPV55	.E.....HE	96
HPV13	.E.....-E	95
PCPV1	.E.....-E	95
HPV34	EE.....-HH...T	66
HPV19	--HHH...TCHE--...-	212
HPV25	--HHH...HHHE--...T	221
HPV20	--HHH...HHECC--	215
HPV21	CCCCC...CCCC--...-	222
HPV14d	--HTH...CHCC--...-	202
HPV5	--HHH...CHCCH--...-	233
HPV36	--HHH...HHHHT--...-	229
HPV47	CCCTE...EECTT--...-	225
HPV12	--TCH...HHHC--...-	213
HPV8	--HTH...HHHET--...-	217
HPV24	--HHH...HHHC--...-	186
HPV15	--HHH...HHHHT--...T	175
HPV17	C-HTH...CCCTT--...H	171
HPV37	--HTH...CEETT--...-	173
HPV9	T-HHH...HHHC--...T	180
HPV22	--HHH...HHHCT--...-	155
HPV23	--HHH...HHHCCT...T	150
HPV38	--HHH...HHHHT--...T	160
HPV49	--HCH...HHHHHH...H	207
HPV4	--HHH...HHHHTHH...H	123
HPV65	--HHH...HHHHTT...T	123
HPV48	--HHH...HHHHHH...-	111
HPV50	--HHH...HHHTT-C...-	109
HPV60	--HHH...TTHCCEEC.H-	124
BPV1	.....C-T...-	101
BPV2	CC.....C-	101
EEPv	T.....CCEE-E	102
DPV	.C.....C--C..T	106
BPV4	.....HHH	113
HPV41	CC.....HH	97
COPV	.C.....-E	106
CRPV	CCCCC...CCCC-	100
ROPV	--HHHHHHHHEHHC--...T	131
HPV1a	--HHH...HHHTT--CCEE	120
HPV63	--HHH...HHHHT	110
MnPV	.EECE...CCCC--T..T	148

### Appendix C. Phenogram Based on E4 Amino Acid Sequences

Phenetic analysis is a form of cluster analysis that attempts to capture the relatedness of sequences irrespective of evolutionary pathways—that is to say the simple similarity of sequences. In the following phenogram, E4 amino acid sequences beginning with the acceptor splice region are compared using the PIMA program as described by Korber et al. (*J. Virol.* **68**:6730–6744, 1994); this approach employs a hierarchical scoring scheme that allows for conserved substitutions in addition to identities. The abscissa records the similarity scores, whereas the ordinate merely records the number of sequences being compared. Sequences 14, 20 and 21 are clearly related; in contrast, the supergroup A (genital-mucosal) and supergroup B (cutaneous) sequences are no more similar than random sequences of approximately the same length, since the root connecting both is virtually at the zero score. The inference that these sequences are homologous must come from their equivalent positions in the PV genomes, not from their similarity scores. Viral types that are closely related in other coding regions—6 and 11, 18 and 45, 13 and PCPV—also have related E4 amino acid sequences, however in general the E4 similarities are weaker. Note that 5E4 and 36E4 are identical proteins, although the viruses 5 and 36 are neither phylogenetically nor phenetically identical overall. On the other hand, 58 and 33, which are generally close in phylogenetic analysis based on L1 or E6, have unexpectedly dissimilar E4 proteins. Thus the phenetic analysis reveals unexpected similarities and unexpected dissimilarities, although in general the degree of relatedness is what might be expected from phylogenetic analyses.

