Analysis of the Sequences of the L1 and L2 Capsid Proteins of Papillomaviruses

Joeseph Icenogle

Centers for Disease Control
1600 Clifton Road NE, B6 109
Atlanta, Georgia 30333

The human papillomaviruses (HPV) compilation and analysis of nucleic acid and amino acid sequences published in September 1994 lists nucleic acid and predicted amino acid sequences corresponding to all or part of the open reading frames (ORFs) of the capsid proteins of 82 different genotypes of papillomavirus (PV). Some sequence analysis of the capsid protein ORFs, including phylogenetic trees, linear correlations, some synonymous/nonsynonymous frequencies and protein information content and density are included in section III of this publication. It was clear at the 14th International Papillomavirus Conference in Quebec City (1995) that a number of groups are currently applying considerable research effort to studying the properties of PV capsid proteins; therefore significantly more information about these proteins will be forthcoming. The primary objectives of the current summary/analysis are to provide a short review of the properties of PV capsid proteins, to provide a framework of sequence analysis upon which forthcoming information about the properties of capsid proteins might be mapped, and to interpret some of this information in terms of the variation in the capsid proteins.

Virions of each of the PVs contain at least two viral proteins, a major protein (L1), with an apparent MW based on SDS PAGE of approximately 55 kDa, and a minor protein (L2), apparent MW approximately 75 kDa. The sizes of the L1 protein and the L2 protein vary slightly between PV genotypes, but they are both about 500 amino acids in length. The reason(s) for the aberrant mobility of the L2 protein has not been determined. Virions have been reported to contain about 10 times as much L1 protein as L2 protein [3]. Although L2 protein is present in small quantities, there is evidence from antibody binding and assembly studies that the L2 protein is an integral structural component of the PV capsid [1, 2].

Because more than one potential initiation codon exists in the ORFs coding for the capsid proteins of some PV genotypes, there is uncertainty about the N-terminal sequences of some capsid proteins. However, because N-terminal sequences are partially conserved between genotypes, consensus start codons can be identified by comparison of groups of related PVs. In the carboxy-terminal portion of the L1 protein is a region which functions as a nuclear localization signal [4]. A similar region exists in the L2 protein. While the HPV type 16 (HPV 16) L1 protein is N-glycosylated in expression systems, clear evidence for glycosylation of the HPV 16 L2 protein is lacking [5, 6]. A role for glycosylation in virion assembly or function has not been established. In the vaccinia system, the L1 proteins which are incorporated into HPV 16 capsids are not glycosylated [5]. On the other hand, there is evidence that purified BPV virions contain minor amounts of glycosylated L1 [7]. Both the HPV 16 L1 and L2 proteins are phosphorylated when expressed in the baculovirus system, with L2 being more heavily and more stably phosphorylated [6].

Three-dimensional image reconstructions from electron micrographs of frozen-hydrated samples have revealed the basic architecture of the PV capsid [8, 9]. The capsid is a T=7 icosahedral lattice composed of 72 pentameric capsomers and has a maximum diameter of approximately 60 nm. The capsomers extend radially about 6 nm from an apparently continuous protein shell which is about 2 nm in width. Thus, the volume available for DNA is a sphere approximately 44 nm in diameter. Since the L2 protein is a minor component of the viral capsid, most, if not all, of the capsomers must contain 5 L1 molecules. Results from experiments on the location of the L2 protein in the icosahedral lattice have not yet produced a location for it. Thus, it is not yet known if the L2 protein is present in some or all capsomers. Disulfide

bonds exist between L1 molecules in PV capsids that are likely important in the assembly and stability of virions [10].

Information about the capsid proteins of PVs is important for a number of reasons. First, sequence analysis of capsid proteins continues to be an important means of identification of new papillomaviruses [11]. Second, the organization the PV capsid is of interest. While the structure of the simian virus 40 capsid has been determined crystallographically and capsids of PVs likely have a similar structure, similarities and differences in the organization of the capsids of the two viruses are of significant structural interest [12]. Third, either the L1 protein alone or the L1 protein and the L2 protein in combination can assemble into virus-like particles (VLPs) in various expression systems [2, 13, 14]. Immunization of animals with such VLP preparations or with inactivated virions from various animal papillomaviruses (e.g. CRPV, COPV) has been shown to protect animals from PV associated disease, and thus there is the potential of prophylactic PV vaccines based on PV capsids [15, 16]. Increased understanding of the immunologic properties of PV capsids (e.g. neutralizing antibody binding sites) would be of significant value in the development of such vaccines. Fourth, an understanding of the mechanism of assembly of HPVs is particularly important because no convenient propagation system yet exists for producing HPV virions, and thus VLPs assembled in various expression systems will likely continue to be an important source of HPV capsid antigens. A fifth reason to study PV capsid proteins is capsid-receptor interactions. Results to date indicate that PVs attach specifically to a protein present on a number of cell types, indicating that the strong species and tissue specificities exhibited by PVs occur after attachment [17, 18, 19]. A number of PV particles have been observed to attach to the same receptor (e.g. BPV virions and VLPs, HPV 16 VLPs, and HPV 11 VLPs). Experiments with HPV 33 VLPs find that this receptor was not sensitive to sialidase, N-glycosidase, or octyl-beta-D-glycopyranoside. The receptor recognizes capsids containing either the L1 protein or both the L1 protein and the L2 protein. Sixth, an understanding of the process of PV uncoating may reveal potential targets for antiviral drugs. Finally, HPV VLPs are being increasingly used as antigens in seroepidemiologic studies of papillomavirus associated diseases [20, 21].

Virion assembly, virion stability, neutralization of virions by antibodies, binding of virions to receptors, and the uncoating of virions, are determined by the primary, secondary, tertiary, and quaternary structures of capsid proteins and complexes and their interactions with other macromolecules. There is currently significant information about the structure of VLPs and virions at about 3nm resolution from image reconstructions, but there is very little information about the secondary and tertiary structure of the L1 proteins or the L2 proteins or about the details of protein-protein interactions which determine capsid structure. Thus, it is now possible to map with confidence the locations of various domains in the capsid proteins only to the primary sequences of the L1 and L2 protein, or to properties entirely determined by the primary sequence (e.g. hydrophilicity profiles).

Some sequences of importance to virion assembly and stability have been mapped. The L2 protein, but not L1, was found to bind HPV 16 DNA; basic amino acids in the first 12 amino acids of the L2 protein were found to be important for this DNA binding [22]. A region of basic amino acids near the carboxy-terminus of the HPV 16 L1 protein was found to be important for nuclear localization of L1 [4]. A similar region occurs near the carboxy-terminus of papillomavirus L2 proteins [4]. The amino acid at residue 202 of HPV 16 L1 determines if VLPs of this genotype will efficiently assemble or not [2].

In animal papillomavirus systems (e.g. BPV and CRPV), both the L1 and L2 fusion proteins have been used to induce neutralizing antibodies/protection of animals [23, 24]. The N-terminal one third of the L2 protein of BPV was found to produce neutralizing antibodies in rabbits [23]. Determination of antibody binding sites on fusion proteins has been the subject of a number of studies, and antibody binding sites at many locations in the L1 and L2 proteins have been mapped. An excellent summary of the reactivity of antibodies to fusion proteins containing the L1 protein or the L2 protein has been published, and such a summary will not be repeated here [25]. There is some evidence of a T-cell response to the L1 and L2 proteins of CRPV, which increases as lesions progress from papilloma to carcinoma [26].

Virus-like particles from different genotypes produce type specific antisera and type specific neutralizing responses [27, 28, 29]. Thus, current data suggests that virions from different PV genotypes are serologically distinct. It is known that many antibodies to PV virions bind to conformational epitopes on the PV surface [30]. The availability of VLPs has allowed the measurement of the ability of antibodies obtained with other immunogens to bind to VLPs, which are likely to be conformationally correct. Figure 1 summarizes the current published data concerning the location of epitopes on either virions or VLPs [25, 31, 32, 33, 34]. The information to date with these antigens is similar to the data with fusion proteins: antibody binding regions of L1 protein in virions/VLPs are distributed throughout the protein, but there are few antibody binding regions near the carboxy-terminus of the L2 protein in VLPs/virions.

A sequence in the L2 protein, (`TTPA(V/I)(L/I)(N/V)(V/I)`), is highly conserved in all HPV types detected in mucosal lesions, and might be involved in determining tissue tropism [35]. This and the other regions described above are also included in Figure 1.

The intertype variability of the L2 proteins is higher than for the L1 proteins (see Part III of the 1994 compendium). For example, comparing proteins from HPV 16, 31, 33, 35, 52, 58, and 67 gave an average information density for the L1 proteins of 0.768479, and 0.636568 for the L2 proteins. An analysis of information density versus position for the L1 proteins of various phylogenetic groups of PVs reveals similarities and differences in the pattern of conserved and variable regions between groups (Figure 2a). The position of some regions is consistent when multiple PV groups are compared (see arrows in Figure 2a). The consistent position of the variable regions in these comparisons of the L1 proteins likely reflects lack of constraints (relative to the conserved areas) in these regions of the L1 proteins of various PVs. Since the L1 protein is the major capsid protein and is largely responsible for essential functions including virion assembly, virion stability, and virion uncoating, a consistent pattern of conserved and variable regions is perhaps not surprising. The common structural elements found in other icosahedral virus structures suggests that the L1 protein has a β-sandwich core structure and that at least some of these variable regions likely occur between strands of this core. However, proof of this possibility will no doubt require a crystallographically determined structure of at least one PV capsid or virion.

At least one variable region in the L1 protein comparison, (the sixth variable region from the amino-terminus designated in Figure 2a) is a region of high variability occurring at a consistent position when some groups were compared (e.g. Group A9 and A10), but is not clearly identified as a highly variable region in other groups of viruses (e.g. Group A7). It would be interesting to determine why some regions of the L1 proteins are highly variable in some groups and not in others.

An information density profile of the L2 proteins reveals far less consistency in the positions of variable regions than is seen in the L1 proteins (Figure 2b). There is little overall conservation of the positions of the variable regions in the amino-terminal 2/3 of different groups of L2 proteins. However, the positions of 6 variable regions in the amino-terminal 2/3 of some groups of L2 proteins occur at nearly the same position in the alignments of different groups (Group A9 and Group A10). Even though comparisons of variable regions in the carboxy-terminal 1/3 of L2 proteins between groups is difficult because of low sequence homology between groups, all groups do have at least one variable region in the carboxy-terminal 1/3 of the L2 protein. Following an interpretation of these results similar to that for the information density profiles of L1 proteins, the carboxy-terminal one third of L2 may have been under similar selective pressures (lack of constraints) during the evolution of most PVs, while the remaining two-thirds of L2 may have been under variable pressures in the different phylogenetic groups of PVs.

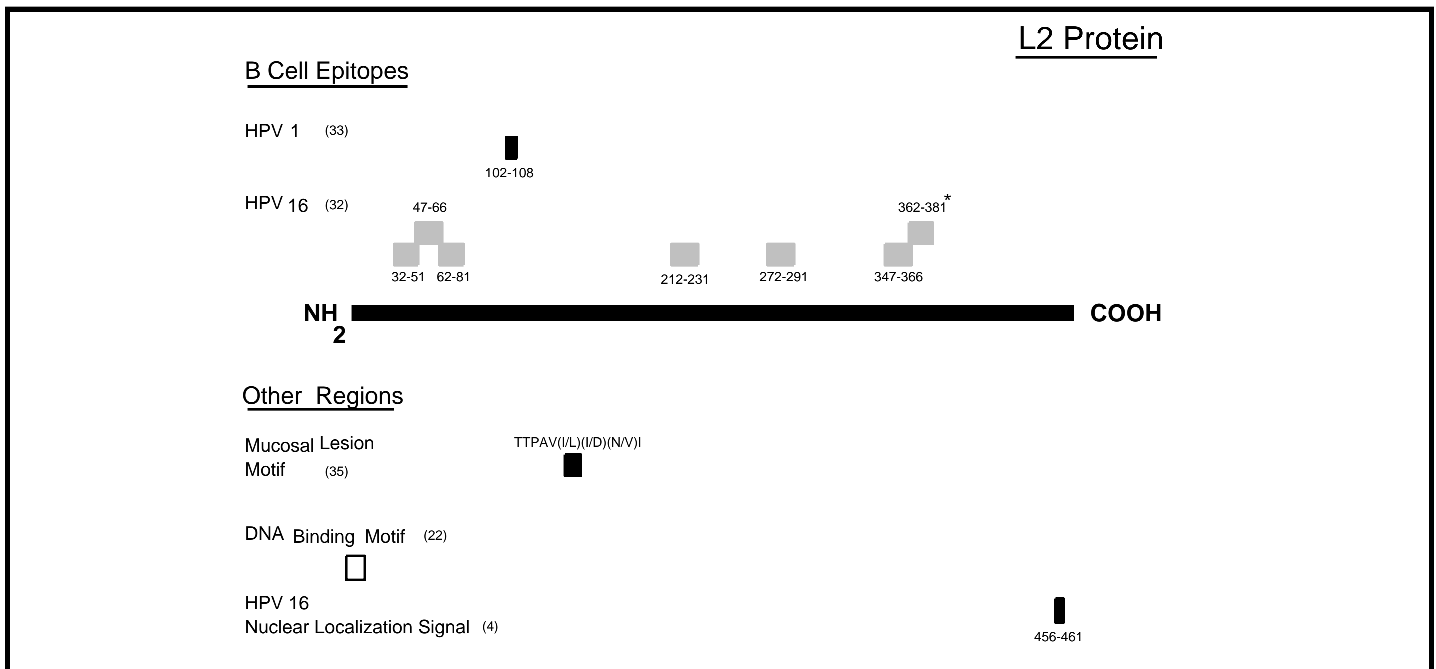
Calculations of the hydrophilicity of a group of L1 proteins and L2 proteins are presented in Figure 3. In each of the three groups of L1 proteins presented, there is high intragroup similarity in the pattern of hydrophilicity obtained. In some cases, regions with a distinctive pattern of hydrophilicity are conserved although the position of the pattern is different in different members of the group. For example, the pattern found from approximately amino acid 320 through 370 of HPV 16 is present in BPV 1, BPV 2, DPV, and EEPV, but in DPV

it occurs at a position closer to the amino terminus. This region is also clearly an example of a region with a hydrophilic profile which is conserved between groups. While the similarity in this profile is easily predicted for the A9 and A7 groups by the high sequence homology between the L1 proteins of these groups in this region, such predictions from sequence alignments alone are difficult for other PVs. For example, the BPV 1 L1 protein has only about 50% sequence homology with the HPV 16 L1 protein in this region. Some regions with distinctive profiles are conserved within a group but not between groups, such as the hydrophilic region around amino acid 55 in group A9. The hydrophilic profiles of the L2 proteins are less conserved both within groups and between groups than the profiles of the L1 proteins. There are however regions with conserved profiles. For example, the profile in the first 80 amino acids of the HPV 16 L2 protein is conserved in other PV L2 proteins.

The antigenic index profiles of the same proteins included in Figure 3 are shown in Figure 4. Although this index must be considered cautiously since it depends on predicted secondary structures of the proteins, more conservation was found in the respective antigenic index profiles of L1 proteins than L2 proteins .

Figure 1. Summary of the B-cell epitopes which are exposed in VLPs and virions. The four studies which were included are indicated [25, 31, 32, 33, 34]. For ease of comparison with original work, the HPV 16 L1 amino acid numbering for this figure only is from the first start codon in the ORF rather than the consensus start codon. The maximum reactivities for each of the two capsid proteins, observed in a study involving antisera to synthetic peptides, are indicated by an asterisk [32]. Other regions of importance or potential importance are included for comparison.

L1 & L2 Capsid Proteins



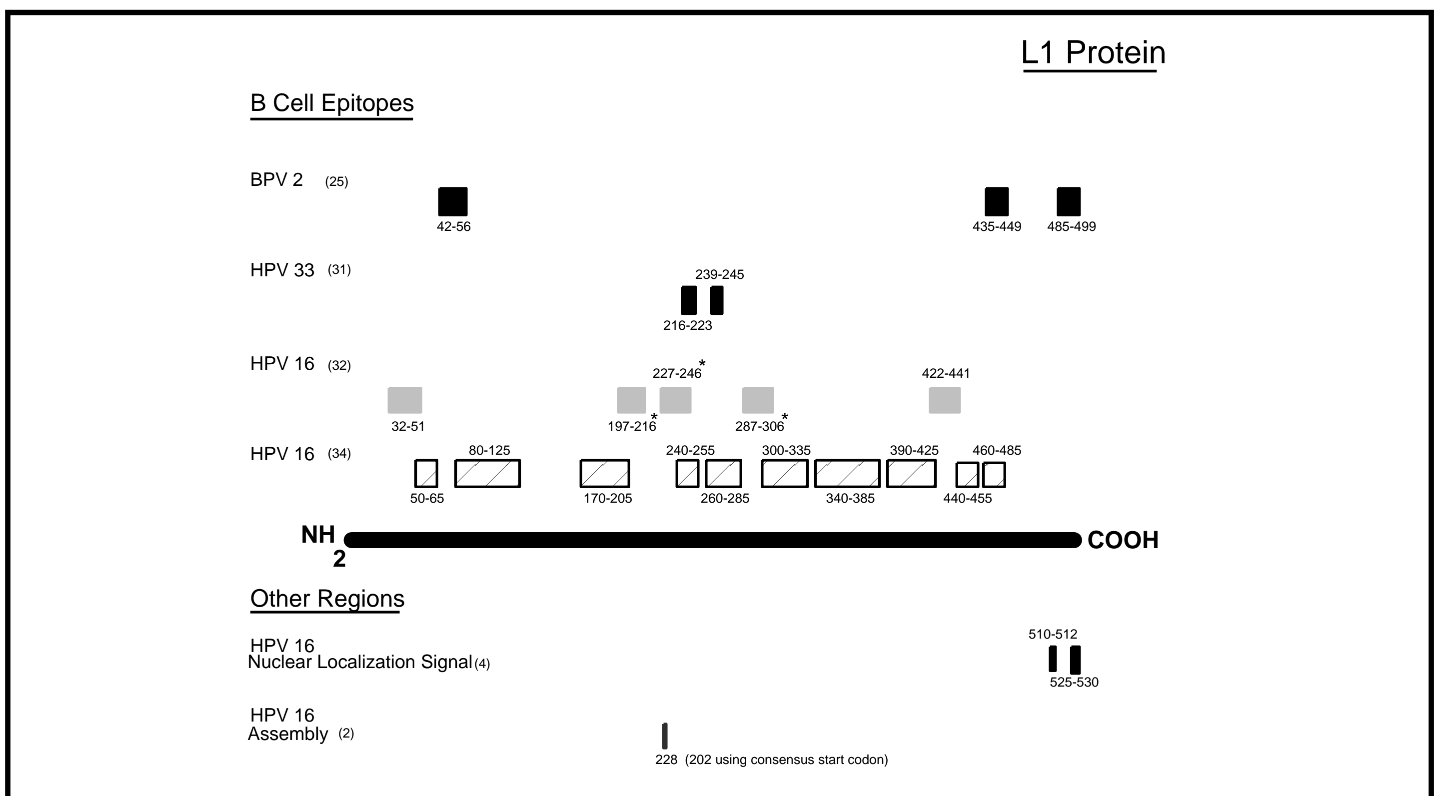


Figure 2. Comparison of L1 and L2 protein conservation in phylogenetically defined groups. Plots represent the information density, averaged over windows of 10 amino acid residues [36, 37]. A value of 0 indicates little or no similarity among the types within a group while a value of 1 indicates identity. Gaps in the plots represent gaps inserted into all types in the relevant group to maintain alignment between groups. The arrows in 2a indicate the position of the most variable regions when multiple HPV groups (Groups A5-A7, A9, A10, and B1) were compared. The 6 amino-terminal arrows in 2b indicate the position of 6 variable regions which occur at fairly consistent locations in the A9 group comparison and the A10 group comparison. The other arrow in 2b is a reference position only (see text). The group designation is based on the phylogenetic analysis of a segment of the L1 protein of [11].

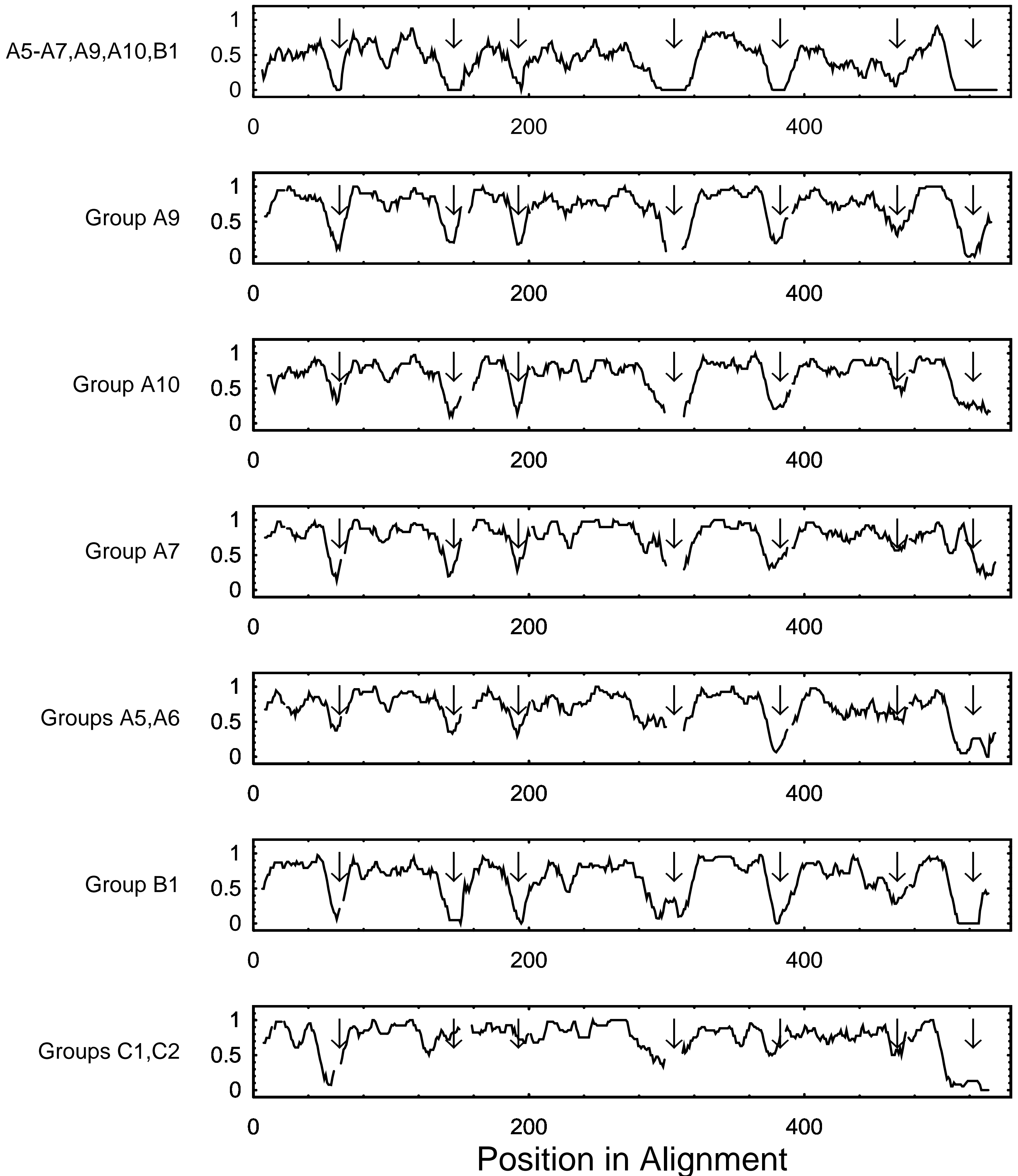
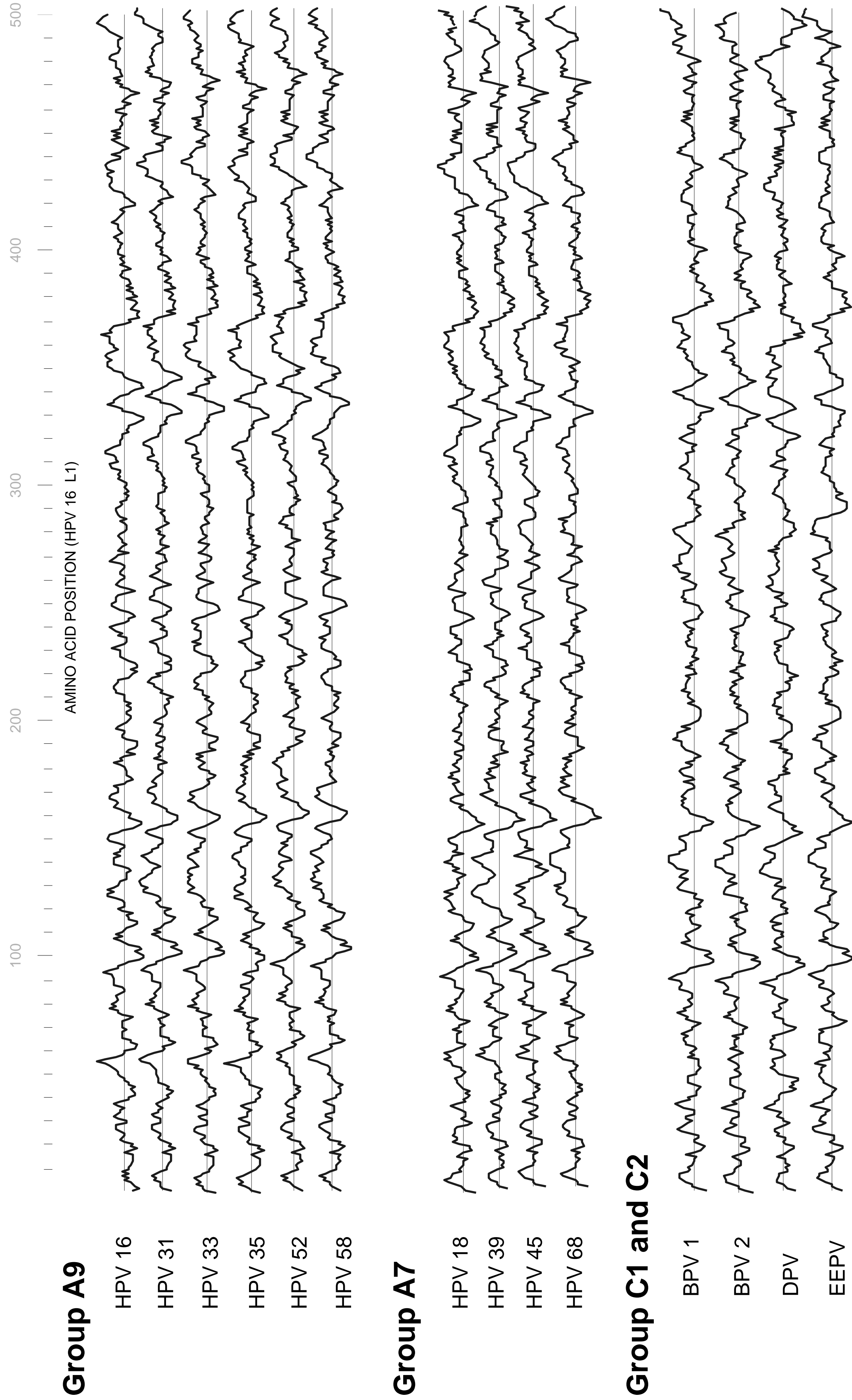
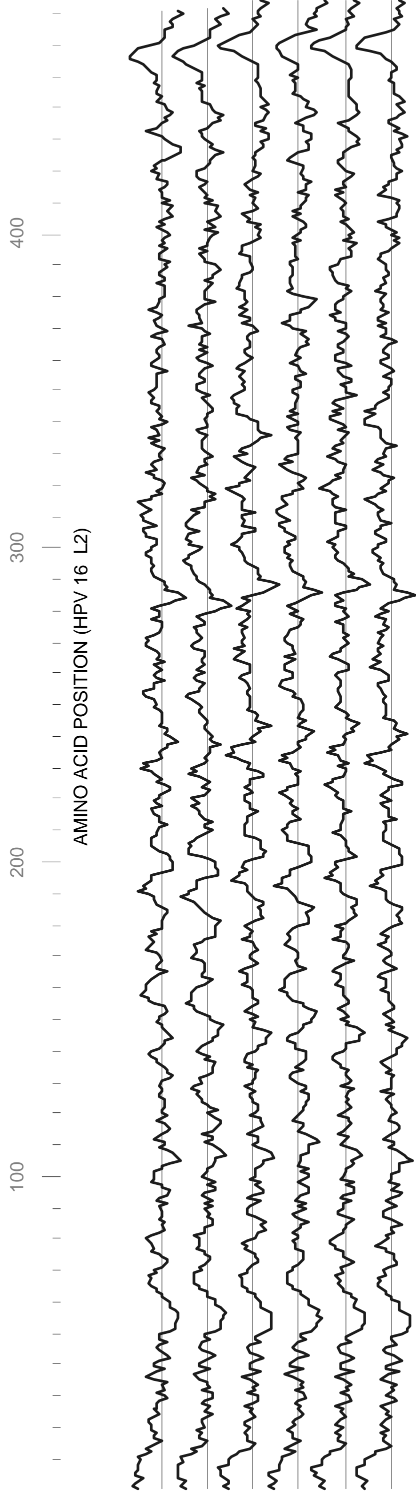

L1 Protein Conservation (Information Density)
A5-A7,A9,A10,B1
Group A9
Group A10
Group A7
Groups A5,A6
Group B1
Groups C1,C2
1
0.5
0
0
200
400
Position in Alignment

Figure 3. Hydrophilicity profiles of L1 proteins and L2 proteins from different groups of PVs. The hydrophilicity profiles for the indicated L1 proteins and L2 proteins were calculated using the PeptideStructure and PlotStructure algorithms in the Genetics Computer Group (GCG) package [38]. The algorithm used to calculate hydrophilicity values is based on the procedure of Kyte and Doolittle using a 9 residue window [39].


L1 Hydrophilicity
AMINO ACID POSITION (HPV 16 L1)
100
200
300
400
500
Group A9
HPV 16
HPV 31
HPV 33
HPV 35
HPV 52
HPV 58
Group A7
HPV 18
HPV 39
HPV 45
HPV 68
Group C1 and C2
BPV 1
BPV 2
DPV
EEPV

L2 Hydrophilicity



Group A7

HPV18
HPV 39
HPV 45
HPV 68

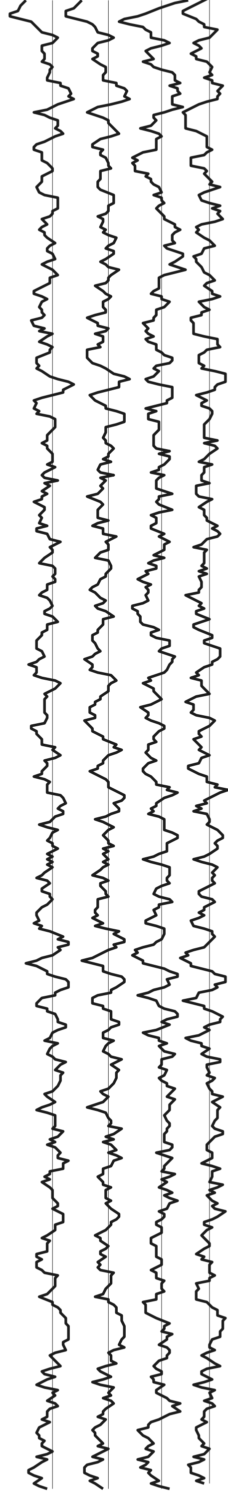
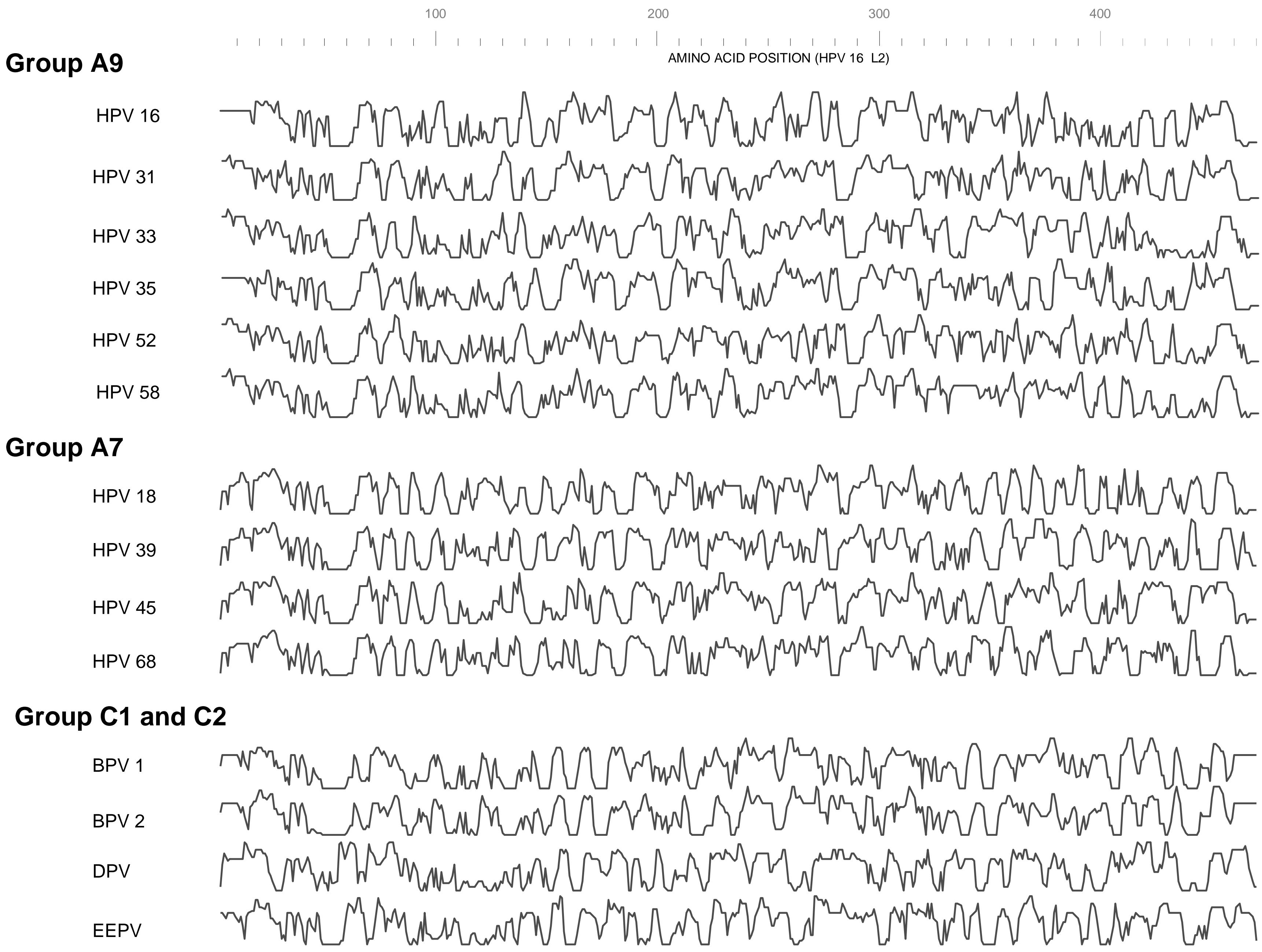


Figure 4. Antigenic index profiles of the L1 proteins and L2 proteins from different groups of HPVs. The antigenic index profiles for the indicated L1 proteins and L2 proteins was calculated using the PeptideStructure and PlotStructure algorithms in the GCG package [38]. The algorithm used to calculate the antigenic index is based on the procedure of Jameson and Wolf using a 9 residue window [40]. Interpretation of antigenic index may be difficult [41].

L1 Antigenic Index
AMINO ACID POSITION (HPV 16 L1)
100
200
300
400
500
Group A9
HPV 16
HPV 31
HPV 33
HPV 35
HPV 52
HPV 58
Group A7
HPV 18
HPV 39
HPV 45
HPV 68
Group C1 and C2
BPV 1
BPV 2
DPV
EEPV

L2 Antigenic Index



[1] Hagensee ME, Yaegashi N, and Galloway DA: Self-assembly of human papillomavirus type 1 capsids by expression of the L1 protein alone or by coexpression of the L1 and L2 capsid proteins. J. Virol., 67, 315-22, 1993.

[2] Kirnbauer R, Taub J, Greenstone H, Roden R, Durst M, Gissmann L, Lowy DR, and Schiller JT: Efficient self-assembly of human papillomavirus type 16 L1 and L1-L2 into virus-like particles. J. Virol., 67, 6929-36, 1993.

[3] Orth G, Breitburd F, and Favre M: Evidence for antigenic determinants shared by the structural polypeptides of (Shope) rabbit papillomavirus and human papillomavirus type 1. Virology, 91, 243-55, 1978.

[4] Zhou J, Doorbar J, Sun XY, Crawford LV, McLean CS, and Frazer IH: Identification of the nuclear localization signal of human papillomavirus type 16 L1 protein. Virology, 185, 625-32, 1991.

[5] Zhou J, Sun XY, and Frazer IH: Glycosylation of human papillomavirus type 16 L1 protein. Virology, 194, 210-8, 1993.

[6] Xi SZ and Banks LM: Baculovirus expression of the human papillomavirus type 16 capsid proteins: detection of L1-L2 protein complexes. J. Gen. Virol., 72, 2981-88.

[7] Larsen PM, Storgaard L, and Fey SJ: Proteins present in bovine papillomavirus particles. J. Virol., 61, 3596-601, 1987.

[8] Hagensee ME, Olson NH, Baker TS, and Galloway DA: Three-dimensional structure of vaccinia virus-produced human papillomavirus type 1 capsids. J. Virol., 68, 4503-5, 1994.

[9] Baker TS, Newcomb WW, Olson NH, Cowsert LM, Olson C, and Brown JC: Structures of bovine and human papillomaviruses. Analysis by cryoelectron microscopy and three-dimensional image reconstruction. Biophysical Journal, 60, 1445-56, 1991.

[10] Volpers C, Schirmacher P, Streeck RE, and Sapp M: Assembly of the major and minor capsid protein of human papillomavirus type 33 into virus-like particles and tubular structures in insect cells. Virology, 200, 504-12, 1994.

[11] Chan SY, Delius H, Halpern AL, and Bernard HU: Analysis of genomic sequences of 95 papillomavirus types: uniting typing, phylogeny, and taxonomy. J. Virol., 69, 3074-83, 1995.

[12] Liddington RC, Yan Y, Moulai J, Sahli R, Benjamin TL, and Harrison SC: Structure of simian virus 40 at 3.8 resolution. Nature, 354, 278-84, 1991.

[13] Zhou J, Sun XY, Stenzel DJ, and Frazer IH: Expression of vaccinia recombinant HPV 16 L1 and L2 ORF proteins in epithelial cells is sufficient for assembly of HPV virion-like particles. Virology, 185, 251-7, 1991.

[14] Sasagawa T, Pushko P, Steers G, Gschmeissner SE, Hajibagheri MA, Finch J, Crawford L, and Tommasino M: Synthesis and assembly of virus-like particles of human papillomaviruses type 6 and type 16 in fission yeast Schizosaccharomyces pombe. Virology, 206, 126-35, 1995.

[15] Breitburd F, Kirnbauer R, Hubbert NL, Nonnenmacher B, Trin-Dinh-Desmarquet C, Orth G, Schiller JT, and Lowy DR: Immunization with virus-like particles from cottontail rabbit papillomavirus (CRPV) can protect against experimental CRPV infection. J. Virol., 69, 3959-63, 1995.

[16] Bell JA, Sundberg JP, Ghim SJ, Newsome J, Jenson AB, and Schlegel R: A formalin-inactivated vaccine protects against mucosal papillomavirus infection: a canine model. Pathobiology, 62, 194-8, 1994.

[17] Muller M, Gissmann L, Cristiano RJ, Sun XY, Frazer IH, Jenson AB, Alonso A, Zentgraf H, and Zhoe J: Papillomavirus capsid binding and uptake by cells from different tissues and species. J. Virol., 69, 948-54, 1995.

[18] Roden RB, Kirnbauer R, Jenson AB, Lowy DR, and Schiller JT: Interaction of papillomaviruses with the cell surface. J. Virol., 68, 7260-6, 1994.

[19] Volpers C, Unckell F, Schirmacher P, Streeck RE, and Sapp M: Binding and internalization of human papillomavirus type 33 virus-like particles by eukaryotic cells. J. Virol., 69, 3258-64, 1995.

[20] Kirnbauer R, Hubbert NL, Wheeler CM, Becker TM, Lowy DR, and Schiller JT: A virus-like particle enzyme-linked immunosorbent assay detects serum antibodies in a majority of women with human papillomavirus type 16. J. Nat. Cancer Inst., 86, 494-9, 1994.

[21] Rose RC, Reichman RC, and Bonnez W: Human papillomavirus (HPV) type 11 recombinant virus-like particles induce the formation of neutralizing antibodies and detect HPV-specific antibodies in human sera. J. Gen. Virol., 75, 2075-9, 1994.

[22] Zhou J, Sun XY, Louis K, and Frazer IH: Interaction of human papillomavirus (HPV) type 16 capsid proteins with HPV DNA requires an intact L2 N-terminal sequence. J. Virol., 68, 619-625, 1994.

[23] Roden RB, Weissinger EM, Henderson DW, Booy F, Kirnbauer R, Mushinski JF, Lowy DR, and Schiller JT: Neutralization of bovine papillomavirus by antibodies to L1 and L2 capsid proteins. J. Virol., 68, 7570-4, 1994.

[24] Lin YL, Borenstein LA, Selvakumar R, Ahmed R, and Wettstein FO: Effective vaccination against papilloma development by immunization with L1 or L2 structural protein of cottontail rabbit papillomavirus. Virology, 187, 612-9, 1992.

[25] Cason J, Kambo PK, Jewers RJ, Chrystie IL, and Best JM: Mapping of linear B cell proteins of bovine papillomavirus: identification of three external type-restricted epitopes. J. Gen. Virol., 74, 2669-77, 1993.

[26] Selvakumar R, Borenstein LA, Lin YL, Ahmed R, and Wettstein FO: T-cell response to cottontail rabbit papillomavirus structural proteins in infected rabbits. J. Virol., 68, 4043-8, 1994.

[27] Hines JF, Ghim SJ, Christensen ND, Kreider JW, Barnes WA, Schlegel R, and Jenson AB: The expressed L1 proteins of HPV-1, HPV-6, and HPV-11 display type-specific epitopes with native conformation and reactivity with neutralizing and nonneutralizing antibodies. Pathobiology, 62, 165-71, 1994.

[28] Rose RC, Bonnez W, Da Rin C, McCance DJ, Reichman RC: Serological differentiation of human papillomavirus types 11, 16 and 18 using recombinant virus-like particles. J. Gen. Virol., 75, 2445-9, 1994.

[29] Christensen ND, Kirnbauer R, Schiller JT, Ghim SJ, Schlegel R, Jenson AB, and Kreider JW: Human papillomavirus types 6 and 11 have antigenically distinct strongly immunogenic conformationally dependent neutralizing epitopes. Virology, 205, 329-35, 1994

[30] Hines JF, Ghim SJ, Christensen ND, Kreider JW, Barnes WA, Schlegel R, and Jenson AB: Role of conformational epitopes expressed by human papillomavirus major capsid proteins in the serologic detection of infection and prophylactic vaccination. Gynecologic Oncology, 55, 13-20, 1994.

[31] Sapp M, Kraus U, Volpers C, Snijders PJ, Walboomers JM, and Streeck RE: Analysis of type-restricted and cross-reactive epitopes on virus-like particles of human papillomavirus type 33 and in infected tissues using monoclonal antibodies to the major capsid protein. J. Gen. Virol., 75, 3375-83, 1994.

[32] Heino P, Skyldberg B, Lehtinen M, Rantala I, Hagmar B., Kreider JW, Kirnbauer R, and Dillner J: Human papillomavirus type 16 capsids expose multiple type-restricted and type-common antigenic epitopes. J. Gen. Virol., 76, 1141-53, 1995.

[33] Yaegashi N, Jenison SA, Valentine JM, Dunn M, Taichman LB, Baker DA, and Galloway DA: Characterization of murine polyclonal antisera and monoclonal antibodies generated against intact and denatured human papillomavirus type 1 virions. J. Virol., 65, 1578-83, 1991.

[34] Zhou J, Sun XY, Davies H, Crawford L, Park D, and Frazer IH: Definition of linear antigenic regions of the HPV 16 L1 capsid protein using synthetic virion-like particles. Virology, 189, 592-9, 1992.

[35] Rho J, Roy-Burman A, Kim H, de Villiers EM, Matsukura T, and Choe J: Nucleotide sequence and phylogenetic classification of human papillomavirus type 59. Virology, 203, 158-61, 1994.

[36] Smith RF and Smith TF: Automatic generation of primary sequence patterns from sets of related sequences. Proc. Natl. Acad. Sci. U.S.A., 87, 118-21, 1990.

[37] Myers G, Bernard HU, Delius H, Favre M, Icenogle J, van Ranst M, and Wheeler C (ed). Human papillomaviruses 1994. A compilation and analysis of nucleic and amino acid sequences. Los Alamos National Laboratory, Los Alamos, NM. 1994.

[38] Genetics Computer Group. Program manual for the GCG package, version 8. Madison: Genetics Computer Group (1994).

[39] Kyte J and Doolittle RF: A simple method for displaying the hydropathic character of a protein. J. Mol. Biol., 157, 105-32, 1982.

[40] Jameson BA and Wolf H: The antigenic index: a novel algorithm for predicting antigenic determinants. CABIOS, 4(1), 181-6, 1988.

[41] Garnier J, Osguthorpe DJ, and Robson B: Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol., 120, 97 (1978).