The papillomavirus major capsid protein L1

Christopher B. Buck\textsuperscript{a,\ast}, Patricia M. Daya\textsuperscript{a}, Benes L. Trus\textsuperscript{b}

\textsuperscript{a} Lab of Cellular Oncology, Center for Cancer Research, NCI, USA
\textsuperscript{b} Imaging Sciences Laboratory, Center for Information Technology, National Institutes of Health, Bethesda, MD, USA

\textbf{Abstract}

The elegant icosahedral surface of the papillomavirus virion is formed by a single protein called L1. Recombinant L1 proteins can spontaneously self-assemble into a highly immunogenic structure that closely mimics the natural surface of native papillomavirus virions. This has served as the basis for two highly successful vaccines against cancer-causing human papillomaviruses (HPVs). During the viral life cycle, the capsid must undergo a variety of conformational changes, allowing key functions including the encapsidation of the ~8 kb viral genomic DNA, maturation into a more stable state to survive transit between hosts, mediating attachment to new host cells, and finally releasing the viral DNA into the newly infected host cell. This brief review focuses on conserved sequence and structural features that underlie the functions of this remarkable protein.

\textcopyright 2013 Elsevier Inc. All rights reserved.

\section*{Introduction}

The papillomavirus major capsid protein, L1, is a ~55 kD protein with the ability to spontaneously self-assemble into virus-like particles (VLPs). These VLPs present an exterior surface essentially indistinguishable from the native 60 nm non-enveloped papillomavirus virion. Purified recombinant L1 proteins can achieve this complex assembly reaction in the absence of any chaperones (Schiller and Lowy, 2012). Assembled VLPs are potent immunogens, likely due to innate B-cell recognition of the regular icosahedrally displayed spacing of surface epitopes (Bachmann et al., 1993; von Bubnoff, 2012). These discoveries laid the foundation for the development of the current VLP-based vaccines that offer highly effective protection against infection with the cancer-causing human papillomavirus (HPV) types 16 and 18.

The assembled VLP immunogens used in current HPV vaccines represent the state of L1 in only one phase of the viral life cycle—namely, the rigid mature form of the virion during its transmission from one cell to another. In contrast to the mature virion state, during the process of virion assembly L1 interactions must be flexible enough to allow selective uptake of the viral genomic DNA into the virion lumen. Transmission of some papillomavirus species is thought to occur via deposition on environmental surfaces, such that the initially fragile immature virion must gain the ability remain infectious in a desiccated state for days or longer to achieve transmission (Roden et al., 1997). This high degree of stability is achieved through a maturation process in which the flexible immature virion gradually achieves a more rigid state that is stabilized by disulfide crosslinks between neighboring L1 molecules (Buck and Trus, 2012).

Since L1 forms the entire exterior surface of the stabilized mature virion, it obviously must mediate initial attachment to host tissues or cells. After attachment to cells, L1 must again become pliable enough to ultimately allow release of the viral genome into a new target cell. This review will focus on the conserved sequence and structural features of L1 that govern its range of functions during the viral life cycle.

\section*{The L1 gene}

The papillomavirus life cycle is closely tied to the biology of keratinocytes, which form the outer layer of the skin (epidermis), as well as the surface of other stratified squamous epithelia, including the genitals, oral cavity, and esophagus. The late phase of the viral life cycle, during which new virions are assembled, occurs only in keratinocytes undergoing terminal differentiation steps that ultimately lead to desquamation (shedding of dead keratinocytes into the environment; Doorbar, 2005).

Restriction of the late phase of the life cycle appears to serve the virus as a form of immune evasion. In this model, restriction of virion production to the desquamating upper layers of the epithelium prevents routine contact between the highly immunogenic...
virion and patrolling immune cells. The idea that restriction of the late phase is of paramount importance for papillomavirus evolutionary fitness is consistent with the observation that the late genes are subject to an extraordinary degree of regulation at essentially every phase of gene expression (reviewed in the current issue of Virology, cite: Schwartz/Ozbun and Bernard articles).

L1 is translated from an RNA species whose transcription is thought to begin in the common upstream regulatory region (URR). To reach the late region of the genome, the transcriptional machinery must pass through the early region genes, then fail to engage the early polyadenylation signal that terminates early gene mRNAs. Messenger RNAs encoding L1 are generated through splicing events that remove early gene sequences present on the pre-mRNA (Wang et al., 2011). There are no known examples of splicing events occurring within the L1 open reading frame (ORF).

The ATG codon presumed to initiate the translation of L1 is invariably located immediately 3' of a consensus splice acceptor site, suggesting that the translational initiation codon for L1 always lies immediately downstream of an intron. With very few exceptions, the N-terminus of L1 carries a consensus MxxWx7YLPP motif. Although computational analysis of papillomavirus genomic sequences might suggest that a minority of papillomavirus species encode an N-terminal extension upstream of the MxxW motif, these hypothetical leader peptides are not conserved, even among members of individual papillomavirus species. Furthermore, there are rarely splice acceptor signals shortly upstream of the hypothetical non-consensus leader sequences. Thus, it appears that the predicted “MxxW” isoform of L1 is only conserved protein product of this gene.

L1 structure

The exterior surface of papillomavirus virions is distinctively knobby (Fig. 1). Each of the 72 knobs is composed of a pentameric L1 capsomer. The N- and C-termini of L1 are arranged as extended “invading arms” that form the floor between the capsomer knobs (Modis et al., 2002; Wolf et al., 2010). The C-terminus of L1 is particularly elaborate (Fig. 2). Each C-terminal invading arm wraps up the canyon-facing surface of the invaded capsomer to a point near the outer apex of the knob. At this high apical point, an inter-L1 disulfide bond is formed between the neighboring capsomers.

HPV16, which causes a majority of cases of cervical cancer, as well as a significant fraction of cancers of the anus, genitals, and tonsils, is the most extensively studied member of the family Papillomaviridae. In HPV16, the inter-capsomeric disulfide bond involves Cys175 in the L1 of the invaded capsomer and Cys428 in the invading L1 C-terminus. Homologs of these two cysteines are conserved among all known papillomavirus L1 proteins. In the mature virion, all L1 molecules are predicted to be involved in either reciprocal disulfide-linked dimers or topologically ring-shaped L1 trimers (Modis et al., 2002). Dimer linkages occur between neighboring hexavalent capsomers, while trimer linkages surround the pentavalent (vertex) capsomer. In a few papillomavirus species, such as bovine papillomavirus type 1 (BPV1), formation of an additional pair of disulfide bonds results in covalent crosslinking of all L1 molecules in the virion into a unimolecular cage (Wolf et al., 2010). Formation of disulfide bonds is a slow process that is thought to occur during the gradual process of desquamation (Buck et al., 2005; Conway et al., 2009). Studies using papillomavirus reporter vectors (pseudovirions) show that while immature capsids that have not yet formed disulfide bonds are infectious, they are physically fragile. Thus, these key disulfide bonds appear to be essential for virion stability.

Fig. 1. Cryo-EM reconstruction of an HPV16 pseudovirion calculated at 18 Å resolution, as described in Buck and Trus (2012). The figure shows a view down the axis of a capsomer at the icosahedral vertex.

Fig. 2. Structure of an L1 capsomer. Images were drawn with PyMOL software (<www.pymol.org>) using the atomic coordinates of BPV1 L1 (Wolf et al., 2010). Each L1 chain was assigned a distinct color. The left panel shows a top view of a non-vertex capsomer, revealing the interdigitation of L1 loops on the surface of the capsomer knob. The right panel shows a side view of the capsomer in which the core beta jellyroll fold of L1 is visible. Conserved cysteines involved in inter-capsomeric disulfide bonds are highlighted as yellow spheres.
The core L1 fold that forms the capsomer knob is an 8-stranded \( \beta \)-jellyroll (Chen et al., 2000) of a type that is common in icosahedral capsid proteins of other viruses including the VP1 major capsid protein of polyomaviruses (Stehle et al., 1996) and a wide variety of other viral families. The 8 antiparallel-stranded \( \beta \)-jellyroll capsid fold is often referred to as “nucleoplasmin-like,” based on its similarity to the structure of a family of eukaryotic proteins called nucleoplasmins (gene symbol NPM), which serve as chaperones for cellular histones. A unique feature of both papillomaviruses and polyomaviruses is that the viral genomic DNA is packaged together with a full complement of cell-derived histones. Although the major capsid proteins of papillomaviruses and polyomaviruses have no linear homology to one another or to nucleoplasmins, we have previously speculated that the common core fold of the three classes of protein might reflect common descent from an ancestral histone chaperone protein (Buck et al., 2008). It remains unknown whether L1 or the VP1 capsid protein of polyomaviruses interact with histones.

The beta strands that form the core of the L1 fold are assigned letters. The top surface of the capsomer knob is composed of loop structures named for the beta strands they cap. Most of the monoclonal antibodies known to neutralize the infectivity of various papillomavirus species bind one or more of these surface loops. For example, the well-studied HPV16-neutralizing monoclonal antibody (mAb) H16.V5 binds in the EP face of the L1 surface loop (L1 residues 262–291), with minor contributions from the HI loop (residues 348–360) (Christensen et al., 2001; Ryding et al., 2007). The surface loops are poorly conserved, even among closely related papillomavirus types. This is presumably the result of selective pressure for papillomavirus to accumulate mutations that prevent the binding of neutralizing antibodies raised by prior infections. The idea is supported by the observation that polyclonal antibody responses raised against L1 VLPs of one HPV type neutralize closely related HPV types with dramatically lower titers (Carter et al., 2003). Serological analyses strongly suggest that the great majority of neutralizing antibodies in HPV vaccine recipients bind epitopes that at least partially overlap the knob surface epitope recognized by H16.V5, suggesting that neutralizing antibodies of the class represented by H16.U4 are uncommon (Wang et al., 1997).

### L1 interactions with the minor capsid protein, L2

The papillomavirus virion can accommodate up to 72 molecules of a minor capsid protein called L2 reviewed in the current issue of Virology, cite: Roden). Although L2 is only minimally exposed on the surface of the mature virion, it emerges from the virion during the infectious entry process (Day et al., 2004, 2008a). L2 is also at least transiently exposed in immature virions (Richards et al., 2006).

Comparison of L1-only HPV16 VLPs to L1+L2 VLPs using computerized reconstruction of cryo-electron micrographs (cryo-EM) revealed that a portion of L2 is located within the interior lumen of the capsomer knob (Buck et al., 2008). This is the same general location in which the polyomavirus minor capsid proteins VP2 and VP3 are known to bind VP1 (Chen et al., 1998). Sequence-based predictions also suggest that conserved capsid-lumen side chains of HPV16 L1 residues 113–130 (\( \beta \)-sheet D) and 247–269 (\( \beta \)-sheet F) may be involved in binding L2 (Lowe et al., 2008).

The point or points through which L2 emerges from the L1 shell during maturation and infectious entry remain unknown.

### The role of L1 in infectious entry

The initial interaction of the papillomavirus capsid with the host is largely attributable to L1 interactions with heparan sulfate (HS) carbohydrates displayed on proteoglycans. This is supported by inhibition of virion binding and infectious entry by soluble heparin (a highly sulfated form of HS) or by enzymatic removal of HS with heparinase both in vivo, utilizing a murine vaginal challenge model, and in vitro with cultured cell lines (Giroglou et al., 2001; Johnson et al., 2009; Joyce et al., 1999). In vivo the HS proteoglycan (HSPG) interaction occurs on the extracellular basement membrane, whereas in vitro PVs interact with HSPG on the cell surface and, to a lesser degree, the extracellular matrix (ECM; Selinka et al., 2007). Although there exist conflicting reports about the dependence of HPV31 in vitro infection on HSPG (Patterson et al., 2005), in vivo studies found complete dependence upon HSPG binding (Johnson et al., 2009). While the Betapapillomavirus HPV5 appears to bind a different spectrum of HS moieties than other papillomaviruses, it nevertheless requires HS binding for infection in vivo (Johnson et al., 2009). At present, it appears that all papillomaviruses require initial attachment to HSPGs for successful infection in vivo.

Capsid interaction with HSPGs results in a subtle conformational change that causes exposure of an amino-terminal portion of the minor capsid protein L2. This exposure allows cleavage of L2 by cellular furin protease (or related cellular proprotein convertases) (Richards et al., 2006). The furin cleavage event occurs on the basement membrane that underlies basal keratinocytes in vivo (Kines et al., 2009). Furin cleavage of L2 appears to induce an additional conformational change that allows the virion to bind a secondary receptor on the keratinocyte cell surface. This secondary conformational change may reduce the affinity of the virion for HS. Evidence suggests that keratinocyte binding, like HSPG binding, is predominantly through L1 interaction (Day et al., 2008b). L1/L2 capsids do not directly interact with keratinocyte cell surfaces in vivo, although L1-only or furin-cleaved L1/L2 particles can bind (Kines et al., 2009). This implies that the receptor binding site is obscured on L1/L2 particles prior to furin cleavage (Day and Schiller, 2009).

The putative secondary receptor on keratinocytes has not been definitively identified. Alpha 6 beta 4 integrin has long been considered a candidate HPV receptor (Abban and Meneses, 2010; Evander et al., 1997). The strong expression of this integrin on basal keratinocytes and its association with HSPG make it an appealing candidate. However, expression is not obligatory for infection, as alpha 6beta 4 negative cell lines can be infected (Shafti-Keramat et al., 2003). This integrin is also non-essential for infection in the in vivo challenge model (Huang and Lambert, 2012).

More recently, CD151 was identified as mediating in vitro HPV internalization via tetraspanin-enriched microdomains (Spoden et al., 2008). Interestingly, integrin association with CD151 appeared to contribute to infection (Scheffer et al., 2013). Additionally, other researchers have suggested that HPV-HSPG-growth factor complexes on the cell surface represent a signaling platform that contributes to infectious endocytosis (Surviladze et al., 2012, 2013). A recent report has implicated annexin A2 as a possible secondary receptor through interaction with L2 (Woodham et al., 2012). However, as with alpha 6beta 4 integrin, some cell lines that lack annexin A2 are infectable (Puisieux et al., 1996; Raff et al., 2013). Except for alpha 6beta 4 integrin, none of these potential internalization mechanisms has been investigated and validated in vivo.
The mechanisms of papillomavirus endocytosis have not been fully resolved for all the different HPV types. Most studies have focused on HPV16 entry. Even here disparate results exist, but the most comprehensive work points to a non-clathrin, non-caveolin pathway that most closely resembles macropinocytosis (Schelhaas et al., 2012). L1 and L1/L2 particles appear to initially traffic similarly, passing through early endosomes to late endosomes (Day et al., 2003; Kämper et al., 2006). The low pH encountered within the late endosomal compartments is required for exposure of the viral genomic DNA, but it is unclear whether this is a direct effect on L1 or an indirect effect on cellular factors.

After trafficking of the virion to endosomal compartments, the contribution of L1 to the entry process seems to terminate. DNA encapsidated within L1-only HPV16 particles does not escape from late endosomes, whereas in L2-containing particles, L2 and the genome are transferred to the trans-Golgi network prior to nuclear delivery (Day et al., 2013). The separation of the L2/genome complex from L1 seems to be mediated by the action of cellular cyclophilin proteins and the interaction of L2 with sorting nexin-17 (SNX17) (Bergant and Banks, 2013; Bergant Marusic et al., 2012; Bienkowska-Haba et al., 2009, 2012).

Conserved sequence motifs

All papillomaviruses whose infectious entry processes have been studied require interactions with heparan sulfate during infectious entry. Using the X-ray crystal structures of truncated L1 capsomers lacking the N- and C-terminal invading arms, Knappe and colleagues selected surface-exposed lysine residues that might be involved in binding HS (Knappe et al., 2007). Individual or combinatorial mutation of HPV16 L1 lyses 54, 278, 356, and 361 resulted in mutant pseudovirions with diminished ability to bind cell surfaces or transduce cells, suggesting that these residues might be involved in HS interactions (Florin et al., 2012). The four lysines are located in the hypervariable surface loops of L1 and are not well conserved, even within the genus Alphapapillomavirus (Dasgupta et al., 2011). In particular, lyses 54, 278, and 356 are not conserved among HPV16’s nearest relatives (e.g., some isolates of HPV types 33, 52, and 58). When considering these studies, it is important to note that possible contributions of portions of L1 that were missing from the initial X-ray crystal structure have not been assessed. It has previously been shown that arginine and tyrosine are enriched in high-affinity HS-binding peptide motifs (Caldwell et al., 1996). Intriguingly, a Tyr—Arg dipeptide on the C-terminal invading arm (HPV16 L1 residues 418 and 419) is known to be surface-exposed along the canyon walls between capsomer knobs (Chen et al., 2001; Wolf et al., 2010). This dipeptide is absolutely conserved in all known papillomavirus species. One line of evidence for contribution of canyon wall motifs to HS binding is the canyon-binding mAb H16.U4 prevents attachment of HPV16 pseudovirions to cells (Day et al., 2007). It would be interesting to know whether the conserved Tyr—Arg or other conserved surface motifs play any role in the apparently universal ability of papillomaviruses to bind HS with high affinity.

Papillomavirus virion assembly occurs in the cell nucleus. A highly conserved polybasic patch at (or very near) the C-terminus of L1 mediates its import into the nucleus (Zhou et al., 1991). The highly conserved polybasic patch at (or very near) the C-terminus of L1 mediates its import into the nucleus (Zhou et al., 1991). The C-terminal due to occlusion of inter-capsomeric contacts by the cytoplasmic binding of karyopherins (Bird et al., 2008). The C-terminal

polybasic patch appears to play a role in encapsidation of the viral genomic DNA, at least for some papillomavirus species (Schafer et al., 2002). Since the polybasic patch is situated within the lumen of the mature virion (Wolf et al., 2010), it seems unlikely that it could be involved in interactions with HS, as suggested in an early study of HPV11 VLPs (Joyce et al., 1999).

There have not been any comprehensive reports on post-translational modifications found on native papillomavirus L1. In unpublished mass spectrometry work, we have observed phosphorylation and N-acetylglycosamine modifications at Thr129 and phosphorylation at Thr340 for HPV16 L1/L2 pseudovirions produced in cultured cells. Thr129 is conserved in most papillomavirus species and Thr340 is absolutely conserved. 

In addition to the conserved homologs of HPV16 L1 Cys175 and 428, which stabilize the mature virion, three additional cysteines (homologous to HPV16 L1 residues 161, 185, and 379) are highly conserved in all papillomaviruses. For the poliovirus SV40, uncoating of the viral genome depends on rearrangement of the disulfide bonds that link neighboring capsomers together into intramolecular disulfide bonds within individual VP1 molecules (Schelhaas et al., 2007). This disulfide shuffling process, which is facilitated by cellular protein disulfide isomerases, appears to allow detachment of vertex capsomers from the virion. Although indirect evidence is consistent with the idea that an analogous process (perhaps involving conserved cysteines flanking the 175—428 disulfide bond) might operate during papillomavirus infectious entry, this remains an area of uncertainty (Campos et al., 2012; Ishii et al., 2007).

Conclusion

Although a substantial amount is known about the dynamic functions of L1 in the papillomavirus life cycle, the protein has many conserved sequences whose functions remain unknown. The study of L1 biology is thus likely to continue to be a fertile area of research.

Acknowledgments

This research was supported in part by the Intramural Research Program of the NIH, the Center for Information Technology (CIT), the National Cancer Institute’s Center for Cancer Research, and the National Institute of Arthritis, Musculoskeletal and Skin Diseases. The authors are grateful to Alasdair Steven and Naiqian Cheng for their work on the HPV16 capsid structure shown in Fig. 1 and to Tom Magaldi for assistance in rendering the structures shown in Fig. 2.

References

Bergant, M., Banks, L. 2013. SNX17 facilitates infection with diverse papillomavirus types. J. Virol. 87, 1270–1273.


