Molecular biology of human papillomavirus infection and cervical cancer

John DOORBAR
Division of Virology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, U.K.

ABSTRACT

HPVs (human papillomaviruses) infect epithelial cells and cause a variety of lesions ranging from common warts/verrucas to cervical neoplasia and cancer. Over 100 different HPV types have been identified so far, with a subset of these being classified as high risk. High-risk HPV DNA is found in almost all cervical cancers (> 99.7%), with HPV16 being the most prevalent type in both low-grade disease and cervical neoplasia. Productive infection by high-risk HPV types is manifest as cervical flat warts or condyloma that shed infectious virions from their surface. Viral genomes are maintained as episomes in the basal layer, with viral gene expression being tightly controlled as the infected cells move towards the epithelial surface. The pattern of viral gene expression in low-grade cervical lesions resembles that seen in productive warts caused by other HPV types. High-grade neoplasia represents an abortive infection in which viral gene expression becomes deregulated, and the normal life cycle of the virus cannot be completed. Most cervical cancers arise within the cervical transformation zone at the squamous/columnar junction, and it has been suggested that this is a site where productive infection may be inefficiently supported. The high-risk E6 and E7 proteins drive cell proliferation through their association with PDZ domain proteins and Rb (retinoblastoma), and contribute to neoplastic progression, whereas E6-mediated p53 degradation prevents the normal repair of chance mutations in the cellular genome. Cancers usually arise in individuals who fail to resolve their infection and who retain oncogene expression for years or decades. In most individuals, immune regression eventually leads to clearance of the virus, or to its maintenance in a latent or asymptomatic state in the basal cells.

HPVs (HUMAN PAPILLOMAVIRUSES) AND DISEASE

HPVs cause a diverse range of epithelial lesions. Over 100 different HPV types have been identified based on DNA sequence analysis [1], with each being associated with infection at specific epithelial sites [2]. At an evolutionary level, HPVs fall into a number of distinct groups or genera (Figure 1), and the lesions they cause have different characteristics.

The two main HPV genera are the Alpha and Beta papillomaviruses, with approx. 90% of currently characterized HPVs belonging to one or other of these groups. Beta papillomaviruses are typically associated with inapparent cutaneous infections in humans but, in immunocompromised individuals and in patients

Key words: cervical cancer, epithelial cell, human papillomavirus, immunity, infection, neoplasia.
Abbreviations: ARF, ADP-ribosylation factor; BPV, bovine papillomavirus; CIN, cervical intraepithelial neoplasia; COPV, canine oral papillomavirus; DAPI, 4,6-diamidino-2-phenylindole; Dlg, the Drosophila discs large protein; E6AP, E6-associated protein; EV, epidermodysplasia verruciformis; HPV, human papillomavirus; HSIL, high-grade squamous intraepithelial neoplasia lesions; Hsp, heat-shock protein; hTERT, human telomerase reverse transcriptase; LSIL, low-grade squamous intraepithelial neoplasia lesions; MCM, multicopy maintenance; ORF, open reading frame; PCNA, proliferating-cell nuclear antigen; PML, promyelocytic leukaemia; Rb, retinoblastoma; pRb, Rb protein; ROPV, rabbit oral papillomavirus; SCC, squamous cell carcinoma; URR, upstream regulatory region.
Correspondence: Dr John Doorbar (email jdoorbar@nimr.mrc.ac.uk).
suffering from the inherited disease EV (epidermodysplasia verruciformis), these viruses can spread unchecked and become associated with the development of non-melanoma skin cancer [3,4]. EV patients carry mutations in their TMC6 (previously known as EVER1) or TMC8 (previously known as EVER2) genes, which renders them susceptible to these viruses [5,6].

The largest group of HPVs comprise the Alpha papillomaviruses, and it is this group that contains the genital/mucosal HPV types (Figure 1). The Alpha papillomaviruses also include cutaneous viruses such as HPV2, which cause common warts, and which are only very rarely associated with cancers [7]. More than 30 different HPV types are known to infect cervical epithelium, with a subset of these being associated with lesions that can progress to cancer. These cancer-associated HPVs are classified as high-risk HPV types. HPV16 is the most prevalent high-risk HPV in the general population, and is responsible for approx. 50% of all cervical cancers. The remaining mucosal types are classified as intermediate or low risk depending on the frequency with which they are found in cancers. Low-risk HPV types, such as HPV11, are associated with cervical cancer only very rarely, but are still medically important because they cause genital warts. Genital warts are a major sexually transmitted disease in many countries and can affect 1–2% of young adults [8].

The remaining HPVs come from three genera (Gamma, Mu and Nu) and generally cause cutaneous papillomas and verrucas that do not progress to cancer (but see [9,10]). The relationship between the different genera of HPVs, and in particular the high-risk HPV types associated with cervical cancer, is shown in Figure 1.

**NORMAL PRODUCTIVE INFECTION**

Many HPV types produce only productive lesions following infection and are not associated with human cancers. In such lesions, the expression of viral gene products is carefully regulated, with viral proteins being produced at defined times and at regulated levels as the infected cell migrates towards the epithelial surface. The events that lead to virus synthesis in the upper epithelial layers appear common to both the low- and high-risk HPV types, with protein expression patterns in low-grade CIN (cervical intraepithelial neoplasia) resembling the patterns found in benign warts caused by other papillomaviruses. Productive infection can be divided into distinct phases, with the different viral proteins playing specific roles.

**Establishment of HPV infection**

All papillomaviruses share a number of characteristics and contain double-stranded circular DNA within an icosahedral capsid (Figure 2). Although the viral genome can vary slightly in size between different HPV types, it typically contains around 8000 bp [7904 bp for HPV16 (GenBank® accession number NC_001526)], and encodes...
eight or nine ORFs (open reading frames) (Figure 2). The virus shell is made up of two coat proteins. The L1 protein is the primary structural element, with infectious virions containing 360 copies of the protein organized into 72 capsomeres [11]. L2 is a minor virion component, and it is thought that a single L2 molecule may be present in the centre of the pentavalent capsomeres at the virion vertices [11,12]. Both proteins play an important role in mediating efficient virus infectivity.

Infection by papillomviruses requires that virus particles gain access to the epithelial basal layer and enter the dividing basal cells. At present there is some controversy as to the precise nature of the receptor for virus entry [13], but it is thought that heparan sulphate proteoglycans may play a role in initial binding and/or virus uptake [13–15]. As with other viruses [16,17], it seems that HPV infection requires the presence of secondary receptors for efficient infection, and it has been suggested that this role may be played by the α6 integrin [18–20]. Papillomavirus particles are taken into the cell relatively slowly following binding [21] and, for HPV16, this occurs by clathrin-coated endocytosis [22]. This mode of entry may not be conserved among all HPV types, however, and it has been suggested that HPV31 may gain entry via caveolae [23]. Papillomavirus particles disassemble in late endosomes and/or lysosomes, with the transfer of viral DNA to the nucleus being facilitated by the minor capsid protein L2 [22,24]. In experimental systems, viral transcripts can be detected as early as 12 h post-infection, with mRNA levels increasing over the course of several days [24]. Infection leads to the establishment of the viral genome as a stable episome (without integration into the host cell genome) in cells of the basal layer, and this is thought to require expression of the viral replication proteins, E1 and E2. The E2 protein plays several roles during productive infection and, in basal cells, is required for the initiation of viral DNA replication and genome segregation. E2 is a DNA-binding protein that recognizes a palindromic motif [AACCg(Nn)cGGTT] in the non-coding region of the viral genome ([25], and Figure 2). HPV16 has four such motifs, one of which lies adjacent to the viral origin of replication. E2 binding is necessary for the recruitment of the E1 helicase to the viral origin, which binds in turn to cellular proteins necessary for DNA replication, including RPA (replication protein A) and DNA polymerase α primase [26–29]. The E2 protein subsequently dissociates from the viral origin, allowing the assembly of E1 into a double hexameric ring that functionally resembles the hexameric ring structures that assemble at cellular replication origins and which are built from the cellular MCM (multicopy maintenance) proteins.

In the basal cells, it appears that the viral genome replicates with the cellular DNA during S-phase, with the replicated genomes being partitioned equally during cell division. The role of E2 in anchoring viral episomes to mitotic chromosomes is critical for correct segregation [30] and, in some papillomavirus types, involves the cellular Brd4 protein, which directly associates through its C-terminus with the viral E2 protein [30,31]. For the high-risk HPV types that are associated with human cancers, association appears to be via the spindle, with additional cellular proteins being involved. In addition to its role in replication and genome segregation, E2 can also act as a transcription factor and can regulate the viral early promoter (p97 in HPV16; p99 in HPV31) and control expression of the viral oncogenes (E6 and E7). At low levels, E2 acts as a transcriptional activator, whereas at high levels E2 represses oncogene expression by displacing SP1 transcriptional activator from a site adjacent to the early promoter (Figure 2). It is unclear whether E2-mediated transcriptional regulation is important in basal epithelial cells, where the nucleosomal structure and/or methylation status of the viral DNA may not be compatible with efficient activation of the early promoter [32,33]. The analysis of genome maintenance in BPV (bovine papillomavirus)-transformed keratinocytes and cell lines derived from cervical lesions [34,35] has suggested that viral episomes may be maintained at 10–200 copies in basal cells, although, as yet, this has not been adequately demonstrated in lesions. Knockout mutants in the viral genome have suggested that both E1 and E2 are required for viral genome maintenance in the basal layer [36,37].

**Stimulation of cell proliferation**

A number of model systems have been used to examine the papillomavirus productive cycle during *in vivo* infection. Following experimental inoculation of mucosal epithelial tissue by ROPV (rabbit oral papillomavirus) or COPV (canine oral papillomavirus), an increase in cell proliferation in the basal and suprabasal cellular compartments is readily apparent, with mature warts becoming visible by 4 weeks post-infection [38,39]. The time between initial infection and the appearance of productive papillomas can vary depending on the titre of the infecting virus and possibly also on the nature of the infecting papillomavirus type, and it has been suggested that latency may result when inoculating titres are low [40,41]. In cervical lesions caused by HPVs, the increased proliferation of suprabasal epithelial cells is attributed to the expression of the viral oncogenes, E6 and E7. During natural infection, the activity of these genes allows the small number of infected cells to expand, increasing the number of cells that subsequently go on to produce infectious virions. The ability of E6 and E7 to drive cells into S-phase is also necessary, along with E1 and E2, for the replication of viral episomes above the basal layer. Suprabasal cells normally exit the cell cycle and begin the process of terminal differentiation in order to produce the protective barrier that is normally provided by the skin [42]. In HPV-infected keratinocytes,
Figure 2  Organization of the HPV genome and the virus life cycle

(A) The HPV16 genome (7904 bp) is shown as a black circle with the early (p97) and late (p670) promoters marked by arrows. The six early ORFs [E1, E2, E4 and E5 (in green) and E6 and E7 (in red)] are expressed from either p97 or p670 at different stages during epithelial cell differentiation. The late ORFs [L1 and L2 (in yellow)] are also expressed from p670, following a change in splicing patterns, and a shift in polyadenylation site usage [from early polyadenylation site (PAE) to late polyadenylation site (PAL)]. All the viral genes are encoded on one strand of the double-stranded circular DNA genome. The long control region (LCR from 7156–7184) is enlarged to allow visualization of the E2-binding sites and the TATA element of the p97 promoter. The location of the E1- and SP1-binding sites is also shown. (B) The key events that occur following infection are shown diagrammatically on the left. The epidermis is shown in colour with the underlying dermis being shown in grey. The different cell layers present in the epithelium are indicated on the left. Cells in the epidermis expressing cell cycle markers are shown with red nuclei. The appearance of such cells above the basal layer is a consequence of virus infection, and in particular, the expression of the viral oncogenes, E6 and E7. The expression of viral proteins necessary for genome replication occurs in cells expressing E6 and E7 following activation of p670 in the upper epithelial layers (cells shown in green with red nuclei). The L1 and L2 genes (yellow) are expressed in a subset of the cells that contain amplified viral DNA in the upper epithelial...
However, the restraint on cell-cycle progression is lost and normal terminal differentiation does not occur [43]. The basic mechanism by which papillomaviruses stimulate cell-cycle progression is well known and is similar to the way that other tumour viruses deregulate cell growth. E7 associates with pRb [Rb (retinoblastoma) protein] and other members of the pocket protein family and disrupts the association between pRb and the E2F family of transcription factors, irrespective of the presence of external growth factors (Figure 3). E2F subsequently transactivates cellular proteins required for viral DNA replication such as cyclins A and E. E7 also associates with other proteins involved in cell proliferation, including histone deacetylases [44], components of the API transcription complex [45] and the cyclin-dependent kinase inhibitors p21 and p27 [46]. During natural infection, however, the ability of E7 to drive cell proliferation is inhibited in some cells, depending on the levels of the p21 and p27 cyclin-dependent kinase inhibitors (Figure 3). High levels of p21 and p27 in differentiating keratinocytes can lead to the formation of inactive complexes with E7 and cyclinE within the cell. It appears that the ability of E7 to drive cells through mitosis in differentiating epithelium may be limited to those cells which express p21 and p27 at low level or which express sufficient E7 to overcome the block to cell-cycle progression [47]. This is important given that deregulated expression of the viral oncoproteins is a predisposing factor in the development of HPV-associated cancers (Figure 3). The function of the viral E6 protein complements that of E7 and, in the high-risk HPV types, the two proteins are expressed together from a single polycistronic mRNA species [48]. A primary role of E6 is its association with p53 which, in the case of the high-risk HPV types, mediates p53 ubiquitination and degradation. This is thought to prevent growth arrest or apoptosis in response to E7-mediated cell-cycle entry in the upper epithelial layers, which might otherwise occur through activation of the ARF (ADP-ribosylation factor) pathway (see Figure 3). The general role of E6 as an anti-apoptotic protein is emphasized further by the finding that it also associates with Bak [49] and Bax [50]. This role of E6 is of key significance in the development of cervical cancers (discussed in more detail below), as it compromises the effectiveness of the cellular DNA damage response and allows the accumulation of secondary mutations to go unchecked. The E6 protein of the high-risk HPV types also plays a role in mediating cell proliferation independently of E7 through its C-terminal PDZ ligand domain [the name PDZ is derived from the first three proteins in which these domains were found: PSD-95 (a 95 kDa protein involved in signalling), Dlg (the Drosophila discs large protein), and ZO1 (the zonula occludens 1 protein which is involved in maintaining epithelial cell polarity)] [51]. E6 PDZ binding can mediate suprabasal cell proliferation [52,53] and may contribute to the development of metastatic tumours by disrupting normal cell adhesion. In productive lesions, cells are driven into cycle only in the lower epithelial layers and extend towards the epithelial surface to varying extents depending on lesion grade and the nature of the infecting HPV type [54]. In benign warts caused by HPV1, proliferating cells are restricted to the epithelial basal layer, with genome amplification beginning as soon as the infected cell enters the suprabasal cell layers. HPVs, such as HPV6 and HPV16 (Alpha papillomaviruses), typically have a region that may be many cell layers thick, where cells are retained in cycle prior to the onset of vegetative viral genome amplification [54]. In the case of the high-risk HPV types that are associated with cervical cancer, the relative thickness of these layers increases with the grade of neoplasia, while the extent of epithelial differentiation decreases.

**Genome amplification**

Although cell proliferation is required for lesion formation and the maintenance of viral episomes, all papillomaviruses must eventually amplify and package their genomes if infectious virions are to be produced. What triggers the onset of late events is not yet fully understood, but appears to depend in part on changes in the cellular environment as the infected cell moves towards the epithelial surface. Critical for this is the up-regulation of the differentiation-dependent promoter, which for many HPV types is contained within the E7 ORF (P670 in HPV16; p742 in HPV31) [55,56]. The activation of the differentiation-dependent promoter depends on changes in cellular signalling, rather than on genome amplification [55,56], and leads to an increase in the level of viral proteins necessary for replication (i.e. E1, E2, E4 and E5). Cells supporting productive infection can be visualized using antibodies to E4 (Figure 4), whereas those supporting genome amplification also contain E7 [57].

Layers. Cells containing infectious particles are eventually shed from the epithelial surface (cells shown in green with yellow nuclei). In cutaneous tissue, this follows nuclear degeneration and the formation of flattened squames. The timing and extent of expression of the various viral proteins are summarized using arrows at the right of the Figure. The consequence of expressing viral gene products in this ordered way is shown on the far right. The expression of E6 and E7 in the presence of low levels of E1, E2, E4 and E5 allows maintenance of the viral genome (genome maintenance/cell proliferation). Elevation in the levels of these replication proteins facilitates viral genome amplification. The first appearance of L2 allows genome packaging to begin, with the expression of L1 allowing the formation of infectious virions (virus assembly). The accumulation of E4 close to the epithelial surface may improve the efficiency of virus release.
Figure 3  Stimulation of cell-cycle progression by high-risk HPy types

HPy infection leads to deregulation of the cell cycle. Regulation of protein expression in uninfected epithelium is shown in (A). In the presence of high-risk HPy (B), the regulation of proteins necessary for cell proliferation is altered, allowing HPy to stimulate S-phase entry in the upper epithelial layers. (A) Uninfected epithelium. The expression of proteins necessary for cell-cycle progression is controlled by pRB, which in non-cycling cells associates with members of the E2F transcription factor family (centre). In the presence of growth factors, cyclinD/Cdk4/6 is activated, which leads to Rb phosphorylation and the release of the transcription factor E2F, which drives the expression of proteins involved in S-phase progression. p16 regulates the levels of active cyclinD/Cdk in the cell, providing a feedback mechanism that regulates the levels of MCM, PCNA (proliferating-cell nuclear antigen) and cyclinE. p14Arf, whose expression is directly linked to that of p16, regulates the activity of the MDM (murine double minute) ubiquitin ligase, which maintains p53 at a level below that required for cell cycle arrest and/or apoptosis. (B) HPy-infected epithelium. In cervical epithelium infected by high-risk HPy types, progression through the cell cycle is not dependent on external growth factors, but is stimulated by the E7 protein, which binds and degrades pRB and facilitates E2F-mediated expression of cellular proteins necessary for S-phase entry. Although p16 levels rise, normal feedback is by-passed, as HPy-mediated cell proliferation is not dependent on cyclinD/Cdk4/6. The rise in the level of p14Arf, which occurs in the absence of p16-mediated feedback, leads to the inhibition of MDM function and an increase in the level of p53. This is countered by E6, which associates with the E6AP ubiquitin ligase in order to stimulate the degradation of the p53 protein and prevent growth arrest and/or apoptosis. In low-grade cervical disease, where E7 levels are carefully regulated, it is thought that E7-mediated cell proliferation can sometimes be inhibited as a result of association with p21 and cyclin E/Cdk. The high levels of E7 found in cervical cancer cells are thought to overcome this block by binding and inactivating the Cdk inhibitor p21.

The role of E1 and E2 in viral genome amplification is well established. E1 is highly conserved among papillomaviruses and has a weak affinity for a consensus motif (AACNAT) repeated 6 times in the viral origin. During natural infection, E1 is expressed at very low levels and requires the presence of E2 in order to be efficiently targeted to its binding sites. E2 associates with E1 primarily through its N-terminus and binds to DNA as a dimer through its C-terminus [58,59] (Figure 2). The formation of an E1–E2 complex at the origin of replication induces localized distortion in the viral DNA, which facilitates the recruitment of additional E1 molecules and the eventual displacement of E2 [60]. E1 can also associate with cellular Hsps (heat-shock proteins), in particular Hsp40 and Hsp70, and this contributes to the formation of E1 dihexamers [61]. E1 and E2 also act to regulate the viral early promoter (p97 in HPV16 and p99 in HPV31), with high levels of E2 acting to down-regulate the expression of E6 and E7 in experimental systems. The ability of E2 to either repress or activate early viral gene expression according to its abundance is thought to result from differences in the affinity of E2 for its various binding sites [62]. In HPV16, it is thought that binding site 4 is the primary site that is occupied when E2 is present at low levels and that binding to this site and to binding site 3 (Figure 2) leads to promoter activation [63]. As E2 increases in abundance, occupancy of the remaining sites leads to the displacement...
Figure 4 Characterization of events during productive papillomavirus infection by immunostaining

The pattern of viral protein expression is conserved in productive lesions caused by papillomaviruses of diverse origins. Typical immunofluorescence staining patterns are shown. (A) Productive papilloma showing the typical expression profile of PCNA (a surrogate marker of E6/E7 gene expression in red) and E4 (in green), which is thought to reflect the sites of expression of E1 and E2. Nuclei are counterstained with DAPI (4,6-diamidino-2-phenylindole: in blue). A small area of uninfected tissue is apparent on either side of the papilloma. (B) Comparison of the sites of viral genome amplification (in red; detected by FISH: fluorescence in situ hybridization) and E4 expression (in green) reveal that the two events correlate very closely. Nuclei are counterstained with DAPI (blue). (C) The viral capsid protein (L1; in red) is expressed in a subset of the cells that express E4 (in green) in the upper epithelial layers. Nuclei are counterstained with DAPI (in blue).

expression that is important in stimulating viral genome amplification will lead eventually to the down-regulation of E6/E7 expression and to the eventual loss of the replicative environment necessary for viral DNA synthesis. This curious link between replication and transcription provides a mechanism by which the virus can limit the timing and duration of genome amplification. In cervical lesions caused by HPV16, cells supporting viral genome amplification are often scarce and vary greatly in their location within the lesion [57]. By contrast, in verrucas caused by HPV1, cells supporting genome amplification are relatively prevalent and are consistently found immediately above the basal layer.

Although E1 and E2 are key players in viral genome amplification, the viral E4 and E5 proteins also contribute [65–69]. E5 mutant genomes of HPV16 and 31 exhibit lower levels of genome amplification than wild type, and it is thought that the ability of E5 to modulate cell signalling is responsible for this. HPV E5 is a transmembrane protein that resides predominantly in the ER (endoplasmic reticulum), but which can associate with the vacuolar proton ATPase and delay the process of endosomal acidification [70–72]. It is thought that this affects the recycling of growth factor receptors on the cell surface, leading to an increase in EGF (epidermal growth factor)-mediated receptor signalling and the maintenance of a replication competent environment in the upper epithelial layers [73]. By contrast, the role of E4 in genome amplification is not yet fully established. E4 accumulates in the cell at the time of viral genome amplification, and its loss has been shown to disrupt late events in a number of experimental systems [67–69]. Part of the effect may be related to the ability of certain HPV types to associate with cyclinB/Cdk2 and to relocate the complex to the cytoplasm [74,75]. This prevents the nuclear accumulation of cyclinB/Cdk2, which is necessary for progression through mitosis. The ability of E4 to cause cell-cycle arrest in G2 and to antagonize E7-mediated cell proliferation is a common feature of the E4 proteins encoded by several HPV types, including HPV16, HPV11 [75], HPV18 [76] and HPV1 [77]. Interestingly, recent work has suggested that the E4 proteins of HPV16 and HPV18 can also associate with E2, which suggests an additional mechanism by which E4 may act [78].

Virus assembly and release

The final stage in the papillomavirus productive cycle requires that the replicated genomes are packaged into infectious particles. Capsid proteins (L1 and L2) accumulate after the onset of genome amplification, with L2 expression preceding the expression of L1 [79,80]. The events that link genome amplification to the synthesis of the capsid proteins are not yet fully understood, but are dependent on changes in mRNA splicing and on the generation of transcripts that terminate at the late
thought to limit the ability of the immune system to squame [104]. The retention of papillomavirus antigens by being shed from the epithelial surface as a cornified layer is thought to require E2 in addition to the capsid proteins L1 and L2 [92,93], and it has been suggested that E2 may improve the efficiency of genome encapsidation during natural infection [93,94]. L2 localizes to the nucleus by virtue of nuclear localization signals located at its N- and C-termini [95], which is also involved in recycling the viral E1 protein during replication [97], but does not appear to be strictly dependent on PML localization [98]. Although papillomavirus L2 proteins can associate directly with DNA [95], the specific recruitment of viral genomes to PML bodies is thought to require E2, which can associate with viral DNA through its specific recognition sites. L1 assembles into capsomeres in the cytoplasm prior to nuclear relocation and is recruited into PML bodies only after L2 has bound and has displaced the PML component sp100 [79]. The assembly of virus-like particles in experimental systems requires the presence of the cellular protein Hsp70 [96], which is also involved in recycling the viral E1 protein during replication [97], but does not appear to be strictly dependent on PML localization [98]. Although papillomavirus particles can assemble in the absence of L2, its presence contributes to efficient packaging [99] and enhances virus infectivity [100]. Loss of L2 in the context of the HPV31 genome results in a 10-fold reduction in packaging efficiency and a 100-fold reduction in virus infectivity when compared with wild-type HPV31 [101]. L2 associates with L1 through a hydrophobic region near the C-terminus of the protein that is thought to insert into the central hole in the pentavalent L1 capsomeres [102]. The interaction between capsomeres requires the C-terminus of the L1 protein [11] with virus maturation and stabilization occurring as the infected cells approach the epithelial surface as a result of disulphide cross-linking. Although papillomaviruses are resistant to desiccation [103], it is thought that their survival may be enhanced by being shed from the epithelial surface as a cornified squame [104]. The retention of papillomavirus antigens until the infected cell reaches the epithelial surface is thought to limit the ability of the immune system to detect infection. Ultimately, virus release requires efficient escape from the cornified envelope at the cell surface, which may be facilitated by the E4 protein. E4 can disrupt the keratin network [105,106] and can affect the integrity of the cornified envelope [104,107].

ABORTIVE INFECTION AS A PRECURSOR TO CANCER

Virus-induced cancers often arise at sites where productive infection cannot be properly supported. CRPV (cotton tail rabbit papillomavirus) induces productive papillomas in its natural host (the cotton tail rabbit), but gives rise to lesions that cannot support virus synthesis when inoculated into domestic rabbits. Such abortive infections progress to cancer more frequently than do infections in cottontails. A similar situation is seen following the inoculation of SV40 (simian virus 40) and adenovirus type 5 (which are also considered to be DNA tumour viruses) into inappropriate hosts, such as hamsters and rats, which support early, but not late, viral gene expression. The papillomavirus types that usually cause benign cutaneous warts in humans and are considered low risk (e.g. HPV2 and 4) can be associated occasionally with cancers at mucosal sites. The restricted tissue tropism of the different HPV types, coupled with their ability to infect non-optimal sites, may provide a partial explanation as to why otherwise benign HPV types are occasionally associated with human cancers.

The high-risk papillomavirus types that are associated with human cancers more frequently come predominantly from the Alpha 9 and Alpha 7 groups (see Figure 1), with HPV16 and HPV18 being the most prevalent types [108]. These viruses are found in women with no cytological abnormalities, as well as in women with LSIL and HSIL (low- and high-grade squamous intraepithelial neoplasia lesions respectively) and/or cancer [109–111]. By PCR, HPV16 DNA is apparent in approx. 26% of LSIL [112], but can be seen in as many as 63% of SCCs (squamous cell carcinomas) of the cervix [113]. HPV18 is the second most common HPV type associated with this disease (causing 10–14% of all cases of cervical SCCs), but is the primary HPV type associated with cervical adenocarcinoma, causing 37–41% (HPV16 causes 26–37% of all such cases [113]). The prevalence of high-risk HPV infection amongst young women (mean age, 25 years) is typically approx. 20–40% depending on geographical location [111,114], with the incidence declining with age as infections are either resolved or brought under control by the host immune system. HPV infections are very common in this age group, with cumulative incidence of infection over 5 years being as high as 60% in some populations [110]. Most women (80%) infected with a specific HPV type will, however, show no evidence of that type after 18 months, and it
Figure 5  Changes in expression patterns that accompany progression to cervical cancer
During cancer progression, the pattern of viral gene expression changes. In CIN1 (LSIL), the order of events is generally similar to that seen in productive lesions (shown diagrammatically on the left). In CIN2 and CIN3, however, the onset of late events is retarded, and although the order of events remains the same, the production of infectious virions becomes restricted to smaller and smaller areas close to the epithelial surface. Integration of HPV sequences into the host cell genome can accompany these changes and can lead to further deregulation in the expression of E7 (and the loss of the E1 and E2 replication proteins). In cervical cancer (shown on the right), the productive stages of the virus life cycle are no longer supported and viral episomes are usually lost.

Figure 6  Detection of viral proteins in cervical intraepithelial neoplasia
(A) The expression of surrogate markers of E7 (in this case MCM; in red) and E4 (in green) in HPV16-infected cervix resembles the pattern of expression seen in productive papillomas caused by other papillomavirus types (see also Figure 4). Nuclei are counterstained with DAPI (in blue) and uninfected tissue is apparent either side of the area of infection. (B) During cancer progression, the expression of surrogate markers of E7 (in this case MCM; in red) extends towards the epithelial surface and the expression of E4 is restricted to isolated pockets close to the upper epithelial layers. CIN3 is shown on the left, whereas the pattern seen in many productive CIN1 lesions is shown on the right. Nuclei are counterstained with DAPI (in blue).

is generally thought that re-infection by the same HPV type is uncommon [115]. The development of high-grade cervical neoplasia arises in women who cannot resolve their infection and who maintain persistent active infection for years or decades following initial exposure. In such women, the continuous stimulation of S-phase entry and cell proliferation by E7, coupled with the loss of p53-mediated DNA repair pathways as a result of E6 expression, allows the accumulation of secondary point mutations in the cellular genome that eventually lead to cancer. Previous work from a number of laboratories has suggested that not all cervical cancers develop through this route, however, and that, in some instances rapid onset of HSIL can arise after initial infection [110,116]. Given the prevalence of genital HPV types in the general population and the high life-time risk of infection (estimated at approx. 80%), the incidence of cervical cancers is very low [typically approx. 0.03% in the absence of screening (0.018–0.044%)] with most infections being successfully resolved [117].

Low-grade cervical lesions (i.e. LSIL or grade 1 CIN) caused by high- and low-risk HPV types are similar to productive infections in the pattern of viral gene expression, and viral coat proteins can usually be detected in cells at the epithelial surface [57] (Figure 5). High-grade lesions (HSIL or grade 2 or 3 CIN) have a more extensive proliferative phase, with the productive stages of the virus life cycle being supported only poorly (Figures 5 and 6). It has been estimated that approx. 20% of CIN1 will progress to CIN2, and that approx. 30% of these lesions will progress to more severe neoplasia if left untreated. Approx. 40% of CIN3 lesions can progress to cancer [118], with cervical neoplasia generally arising within the cervical transformation zone where the columnar cells of the endocervix meet the stratified squamous epithelial cells of the ectocervix. The extent of columnar...
and stratified epithelium in this region changes markedly during a woman’s life, and it is in this region of change that most cervical neoplasias develop. It has been suggested that the transformation zone may be a sub-optimal site for completion of the productive life cycle of high-risk HPV types such as HPV16. Interestingly, high-risk HPV infections can be found at many sites in the anogenital tract, including the vagina, vulva and penis, but, despite this distribution, the incidence of cancer at these sites is generally low (0.001%). Only at the anus, in men who have sex with men, does the incidence of HPV-associated cancer rise to levels that are similar to those found at the cervix (0.035%) [119], and it is interesting that at both these susceptible sites there is a transformation zone. In the cervix, the reserve cells of the transformation zone eventually form the basal cells of a stratified squamous epithelium, and the higher frequency of cancers at such sites may reflect the fact that the virus can access these basal cells more readily than those that are protected from infection by a permanent stratified epithelial layer. Despite this possibility, it appears that the transformation zone may in fact be an epithelial site where high-risk HPV types cannot properly regulate their productive cycle, and that variation both in the level and in the timing of expression of viral proteins may underlie the development of cancers at these sites.

MOLECULAR EVENTS DURING CANCER PROGRESSION

The identification of cervical lesions as flat condyloma, LSIL, HSIL or invasive cervical cancer reflects molecular changes in the normal programme of epithelial cell differentiation that occur following infection. Virus production at the epithelial surface depends on the ordered and timely expression of viral gene products (Figure 2) [54,57], with the timing of such events becoming progressively disturbed during neoplastic progression (Figures 5 and 6). In HSIL, viral genome amplification occurs closer to the epithelial surface than in condylomas and LSIL, and the expression of viral coat proteins is retarded [57] (Figure 5). The molecular bases for these changes are not fully understood, but may in some instances reflect changes in the levels of E6 and E7 expression that occur following integration of the viral genome into the host cell chromosome. Integrated HPV DNA is found in most invasive cancers and in a subset of high-grade lesions [120,121], but can also be found in some CIN1 lesions, and it has been suggested that integration may be an early event in cancer progression [122]. Indeed, p16INK4A expression, which is considered a marker of elevated E7 expression, can be detected in some CIN1 lesions, as well as in CIN2 and CIN3 lesions that show evidence of integration [123,124]. Integration of the HPV genome into the host cell chromosome is a critical event in the development of most cervical cancers and, although this can occur randomly throughout the genome, several studies have indicated a preference for integration at common fragile sites and have suggested that changes in the expression of genes at or near the integration site may participate in cancer development [125]. It is clear, however, that integration (which often results in the loss of E2) can lead to the deregulation of E6/E7 expression, and that this is critical for the enhanced growth characteristics of cervical cancer cells. The requirement of these genes for the maintenance of the cancer phenotype is shown most dramatically by studies that have aimed to inhibit viral oncogene activity in cervical cancer cells. Cell lines such as HeLa and SiHa, which have been grown in the laboratory for many decades, will undergo apoptotic cell death in the presence of molecules that inhibit E6 function [126–128]. Similar results are obtained following the reintroduction of E2 into such cell lines, which suppresses the expression of the viral oncogenes by binding to the URR (upstream regulatory region), preventing continued cell proliferation [129,130]. The ubiquitous retention of the E6/E7 region of the viral genome following integration into the host chromosome is usually accompanied by the loss or disruption of viral sequences encoding most of E1, as well as E2 and E4. The viral E2 protein has a negative effect on cell proliferation by regulating the viral URR and can also cause cell-cycle arrest at the G2 phase of the cell cycle [131,132]. Similarly, the E4 protein of HPV16 is able to inhibit mitosis by preventing the nuclear localization of cyclinB/Cdk1, and it is not surprising therefore that the expression of these proteins is abrogated in HPV-associated cancers. In addition to the loss of regulatory proteins, integration also leads to the loss of sequences at the 3’-end of the viral early transcripts that can suppress the production of viral mRNA species encoding E6 and E7 [133–135], contributing further to the deregulation of viral oncogene expression. Although integrated HPV DNA is found in the vast majority of cervical cancers, other factors are likely to influence the development of the precancerous changes seen in squamous intraepithelial neoplasia. These include exposure to glucocorticoids and progesterones, which can affect viral oncogene expression [136–138], and the regulation of viral gene expression by DNA methylation and chromatin organization [139]. Both factors can also regulate expression from integrated sequences and, in instances where integration occurs as concatenamers, it is often only one copy of the viral genome that is transcriptionally active [140].

Although many HPV types can infect the cervix, only the high-risk HPV types are consistently associated with cervical cancers because of the specific activity of their oncogenes. HPV16 gene expression facilitates integration of foreign DNA into the host cell chromosome, which may be a consequence of the increased level of host
Papillomavirus molecular biology 535

genome instability in these cells [141]. The high-risk E7 proteins, but not the E7 proteins encoded by the low-risk HPV types, induce centrosomal abnormalities in cell culture and in transgenic animals [142,143], and it has been suggested that high-risk E7 may act as a mitotic mutator, which acts to increase the chance of errors during each round of cell division [144]. Although the molecular basis for E7-mediated genome instability is not fully understood, it appears in part to be independent of the well-characterized association of E7 with the pocket proteins Rb, p107 and p130. The association of E7 with these proteins does, however, contribute to the ability of E7 to stimulate cell proliferation, with the high-risk E7 proteins binding Rb more efficiently than the E7 protein of the low-risk HPV types [145,146]. The high-risk E7 proteins are also capable of mediating Rb degradation through a proteosome-dependent mechanism [147,148], which is important for E7-mediated cell transformation [149]. As with E7, the E6 protein also differs in its function between high- and low-risk HPV types. One of the most important of these with regard to cancer progression is the ability of high-risk E6 proteins to form a tripartite complex with p53 and the cellular ubiquitin ligase E6AP (E6-associated protein), which leads to proteosome-mediated p53 degradation [150]. Low-risk E6 proteins bind p53 with a lower affinity than the high-risk types and have no significant ability to bind E6AP and to stimulate p53 degradation [150,151]. The loss of the p53-mediated DNA damage response in cells expressing high-risk HPV E6 predisposes to the accumulation of secondary changes in the host cell chromosome that eventually lead to cancer. The high-risk E6 proteins also differ from those encoded by the low-risk HPV types in having a C-terminal PDZ-binding domain. High-risk E6 proteins bind and stimulate the degradation of several cellular targets that contain PDZ motifs, such as hDlg (human homologue of Dlg) and hSrib (human homologue of the Drosophila scribble protein), which are thought to be involved in the regulation of cell growth and attachment [152]. PDZ binding is a conserved feature of all high-risk E6 proteins, and it has been suggested that the loss of cell–cell contacts mediated by tight junctions may contribute to the loss of cell polarity seen in HPV-associated cervical cancers [153]. PDZ binding, which is distinct from the ability of E6 to bind and degrade p53, is important for cell transformation [154] and has been shown to be necessary for the stimulation of epithelial hyperplasia in transgenic animals [53]. Another important function for high-risk E6 is its ability to activate the catalytic subunit of telomerase [hTERT (human telomerase reverse transcriptase)], which adds hexamer repeats to the telomeric ends of chromosomes [155]. Telomerase activity is usually absent in somatic cells, leading to the shortening of telomeres with successive cell divisions and, eventually, to cell senescence. Although the precise mechanism by which E6 mediates hTERT activation is not known, it is clear that such an activity may predispose to long-term infection and the development of cancer. Interestingly, recent studies have shown that the viral oncogenes E6 and E7 can antagonize BRCA-mediated inhibition of the hTERT promoter [156].

Although aberrant expression of high-risk oncogenes can predispose to the development of cervical cancer, their expression alone is not considered sufficient, and the viral proteins cannot fully transform human keratinocytes in culture [157]. It is generally accepted that papillomavirus-mediated oncogenesis requires the accumulation of additional genetic changes that occur over time following initial infection. The average age of women with invasive cervical cancer is approx. 50 years, whereas the mean age of women with HSIL is approx. 28 years, which suggests in most cases a long precancerous state that allows the accumulation of secondary genetic changes. Although secondary genetic changes may occur randomly, the presence of tobacco metabolites in cervical secretions is considered a risk factor in the development of cervical cancer [158], with certain smoking carcinogens being found at significantly higher levels in the cervical secretions of cigarette-smoking women [159]. Multiparity and the long-term use of oral contraceptives are also associated with increased risk [160,161].

REGRESSION, LATENCY AND ASYMPTOMATIC INFECTION

In contrast with our understanding of HPV-associated cancers, our knowledge of the molecular events that regulate the development of latent and/or asymptomatic infections is only poorly developed. Experimental infections using animal papillomaviruses such as ROPV or COPV lead to the development of lesions that persist for months rather than years [38,162]. Lymphocyte infiltration and lesion regression take place between 8 and 12 weeks post-infection and, by 16 weeks, infected sites regain the appearance of uninfected epithelium [162]. A similar pattern of events occurs in cattle [163], and regression of HPV-induced lesions in humans is also thought to follow this path [164]. The immune system is clearly important in controlling viral persistence, and patients with immune defects can develop widespread lesions that are refractory to treatment. Although the immune responses involved in clearance of HPV infections are considered in depth elsewhere [165,166], it is worth noting that Alpha papillomaviruses (and in particular those associated with cervical cancer) can cause persistent infections and have evolved a number of mechanisms to limit the chance of detection by the immune system. Among these are the regulation of E-cadherin expression and Langerhans cell density by E6 [167,168], the interference with MHC presentation by
E5 [167] and interference with the function of interferon response factor 3 by E7 [169]. Perhaps equally important, however, is the fact that papillomaviruses do not cause a lytic infection, with infectious particles being produced only in the upper epithelial layers in cells that are eventually lost from the epithelial surface at the end of their life span. The viral early proteins that mediate cell proliferation in the lower epithelial layers are thought to be expressed at levels below those required to reliably trigger an effective host immune response. When it occurs, the stimulation of a cell-mediated immune response that can clear infection appears to depend on cross-priming of dendritic cells by viral antigens expressed in keratinocytes. The frequent detection of HPV DNA in cervical lesions in the absence of any obvious disease may represent latent infection in which viral genomes are maintained in the basal layer in the absence of detectable productive infection. Immune surveillance is thought to be important for this, as lesions proliferate following immune suppression, but this may not be the only way in which viral latency is maintained. Recent studies have suggested that the generation of an asymptomatic infection in which viral DNA can be detected in the absence of any abnormal pathology may be a normal consequence of infection under some circumstances, and may result from the silencing of viral gene expression by methylation [139]. Latent infection is thought to require the expression of the viral E1 and E2 proteins necessary for genome maintenance in the basal layer, with the E6 and E7 genes not being required [40].

CONCLUSIONS

The consequence of papillomavirus infection depends on the infecting HPV type and site of infection, as well as on host factors that regulate virus persistence, regression and latency. The characteristics of different HPV types have been studied extensively and it is now well known that high-risk types, such as HPV16, encode genes that can contribute to cancer progression when aberrantly expressed. During productive infection, however, these genes are carefully regulated and play important roles in virus synthesis and in avoiding detection by the host immune system. It seems that papillomaviruses, like many other DNA tumour viruses, cause cancers when their regulated pattern of gene expression is disturbed. Viral oncoproteins such as E6 and E7, which are involved in cell proliferation and the regulation of cell death, can be extremely dangerous when inappropriately expressed. This can happen at the cervical transformation zone, where most cervical cancers originate, and it has been suggested that this may be an unstable site for productive infection by high-risk HPVs. Although our understanding of HPV-associated cancer progression and productive infection is now well developed, little is known as to the influence of the infected cell type and the consequence of different cellular environments on viral gene expression. The factors that regulate viral persistence and the events that lead to latency are other areas that are only poorly understood, but which are very important in fully understanding the consequences of infection in humans. Such topics are difficult to research, but will need to be addressed if our understanding of papillomavirus molecular biology is to be advanced further.

ACKNOWLEDGMENTS

J.D. is a programme leader in the Division of Virology at the MRC National Institute for Medical Research and is supported by the U.K. Medical Research Council. The leadership and vision provided by the present Director, Sir John Skehel FRS, is gratefully acknowledged, as is the continued support and advice provided by Dr Jonathan Stoye, Head of Virology. The commitment and enthusiasm of all members of the Papillomavirus laboratory in carrying out their work and in contributing to our understanding of papillomavirus biology is greatly appreciated.

REFERENCES

9 Shamanin, V., Glover, M., Rausch, C. et al. (1994) Specific types of human papillomavirus found in benign proliferations and carcinomas of the skin in immunosuppressed patients. Cancer Res. 54, 4610–4613

© 2006 The Biochemical Society


29 Han, Y., Loo, Y. M., Milielto, K. T. and Melendez, T. (1999) Interactions of the papovavirus DNA replication initiator proteins, bovine papillomavirus type 1 E1 and simian virus 40 large T antigen, with human replication protein A. J. Virol. 73, 4899–4907


38 Christensen, N. D., Cladel, N. M., Reed, C. A. and Han, R. (2000) Rabbit oral papillomavirus complete genome sequence and immunity following genital infection. Virology 269, 40–51


44 Brehm, A., Nielsen, S. J., Miska, E. A. et al. (1999) The E7 oncoprotein associates with Mi2 and histone deacetylase activity to promote cell growth. EMBO J. 18, 2449–2458


54 Titolo, S., Pelletier, A., Sauve, F. et al. (1999) Role of the ATP-binding domain of the human papillomavirus type 11 E1 helicase in E2-dependent binding to the origin. J. Virol. 73, 5282–5293


© 2006 The Biochemical Society


