Human papillomavirus gene expression is controlled by host cell splicing factors

Tetyana Klymenko and Sheila V. Graham1
MRC-University of Glasgow Centre for Virus Research, Institute of Infection, Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, Garscube Estate, Glasgow G61 1QH, U.K.

Abstract
HPVs (human papillomaviruses) infect stratified epithelia and cause a variety of lesions ranging from benign warts to invasive tumours. The virus life cycle is tightly linked to differentiation of the keratinocyte it infects: papillomaviruses modulate host gene expression to ensure efficient virus replication. For example, the viral transcription factor E2 can directly up-regulate, in an epithelial differentiation-dependent manner, cellular SRSFs [SR (serine/arginine-rich) splicing factors] that control constitutive and alternative splicing. Changes in alternative splicing and the mechanisms controlling this for viral mRNAs have been the subject of intense exploration. However, to date experiments have only been carried out in model systems because the genetic systems suitable for studying alternative splicing of viral RNAs in the context of the virus life cycle are relatively recent and technically challenging. Now using these life cycle-supporting systems, our laboratory has identified SR proteins as important players in differentiation-dependent regulation of HPV gene expression. Better understanding of the role of cellular factors in regulating the virus life cycle is needed as it may help development of novel diagnostic approaches and antiviral therapies in the future.

Introduction
Papillomaviruses are small non-enveloped viruses that infect various animals, including mammals. They have strict tissue tropism and host specificity: thus HPVs (human papillomaviruses) infect only squamous epithelium in the human body. HPVs are classified by genotype and to date over 120 different types have been identified [1]. These viruses can also be grouped by site of infection as some preferentially infect cutaneous epithelium whereas others are found in mucosal epithelium, especially in epithelial cells of the genital tract [2]. The vast majority of HPV infections are benign and manifest as warts, for example verrucae or genital warts. However, a subset of the virus types infecting the genital tract are associated with the development of cancers and these HPVs are referred to as oncogenic or HR (high risk) HPVs. HR-HPVs are regarded as the causative agent of cervical cancer [2]. Approximately 25–35% of head and neck cancers are also linked to HR-HPV infection meaning that the carcinogenic effect of these pathogens is not limited to the genital tract [3].

All HPVs contain a double-stranded circular DNA genome of approximately 8 kb in size, which can be broadly divided into three regions: early, late and LCR (long control region). The early region contains ORFs (open reading frames) encoding non-structural proteins that have been designated from E1 to E8. Two late genes (L1 and L2) encode structural viral capsid proteins, necessary for virion formation, transmission and spread. The LCR contains the origin of replication and various elements that regulate viral gene expression. The virus life cycle is tightly linked to the physiological state of the host keratinocytes [4]. The viral early proteins are expressed in undifferentiated or intermediate differentiated keratinocytes present in the lower epithelial layers. The only notable exception is E4 that continues to be expressed in cells undergoing terminal differentiation in the uppermost epithelial layers along with the L1 and L2 capsid proteins [5]. It is generally believed that restriction of L1 and L2 protein expression to the uppermost epithelial layers is needed to avoid the host immune response. Both proteins are highly immunogenic, indeed the current HPV vaccine contains virus-like particles composed of recombinant L1 protein [6]. It is therefore very important to understand how capsid protein expression is regulated and to uncover links to keratinocyte differentiation.

Capsid gene expression is regulated at both transcriptional and post-transcriptional levels, but the latter is the key in the link between virus gene expression and keratinocyte differentiation [7]. In time it may be possible to design therapies to induce capsid protein expression in the lower epithelial layers, which would result in virus clearance by the immune system.

Regulation of transcription
In HR-HPV types, transcription is initiated at two major promoters, the ‘early promoter’ located upstream of the E6 ORF and the ‘late promoter’ in the E7 ORF [8].
Several other promoters which might play an important role in the viral life cycle have been identified [9–13]. Transcription of HPV late genes is only activated as HPV-positive cells start to differentiate [14,15]. Both viral and host cellular factors are thought to play a role in this, since promoter regions bear multiple transcription factor binding sites that regulate RNA polymerase II transcription [16]. It is established that viral genomic DNA is packaged into chromatin utilizing cellular histones, which form specifically positioned nucleosomes [17,18] and that chromatin structure is rearranged upon differentiation favouring activation of late gene expression [19]. Several studies have documented dynamic differentiation-dependent changes in transcription factor binding to both the LCR and late promoter [20,21]. The binding pattern of a number of transcription factors including c-Myb, Pax5, NFATx (a nuclear factor of activated T-cells family member), C/EBPβ (CCAAT/enhancer-binding protein β) and YY1 (Yin and Yang 1) was changed upon keratinocyte differentiation and it is likely that combinations of these factors act to regulate late gene expression.

Regulation of pre-mRNA processing

Although transcriptional regulation occurs, appropriate production of viral proteins is achieved largely by post-transcriptional means. All papillomavirus genes are transcribed to give multiple polycistronic RNAs containing two or more ORFs, which then undergo extensive processing including alternative splicing and polyadenylation. Tight cell-mediated regulation of the viral pre-mRNA processing mechanisms was first proposed nearly 25 years ago due to the observation that late region transcripts can be detected in the nuclei of the mid-epithelial layers, suggesting that these transcripts are synthesized but are not fully processed [22,23]. Indeed, it was established in our laboratory that late mRNAs produced in less differentiated cells are not polyadenylated and confined to the nucleus [12,24]. In differentiated cells of the upper epithelial layers, these pre-mRNAs are then fully processed [12]. For HPV16, at least 13 different mRNAs with the capacity to encode capsid proteins are produced by alternative splicing [12]. During differentiation of uninfected keratinocytes, levels of RNA-processing factors either do not change significantly or decrease [25]. In contrast, a number of cellular splicing regulators are significantly up-regulated in HPV-positive differentiating keratinocytes, probably to ensure availability of these key splicing factors for viral late RNA processing [26].

Role of SR (serine/arginine-rich) proteins in the HR-HPV life cycle

We previously identified the SR protein family among the splicing-related proteins up-regulated by HPV infection. The levels of the three smallest SR proteins, SRSF1–3 (SR splicing factor 1–3), but not other ‘classical’ SR proteins, are increased upon differentiation of HR-HPV-positive keratinocytes [27,28]. This is mediated by the virus transcription factor E2 since a single point mutation that inactivates the E2 transactivation domain in an otherwise intact HR-HPV genome is enough to abrogate SRSF1–3 up-regulation in cervical epithelial cells ([27], and T. Klymenko and S.V. Graham, unpublished work). E2 can activate transcription of SRSF1 directly by binding regulatory sequences in the SRSF1 gene promoter [25,27], and studies are underway to dissect any SRSF2 and SRSF3 promoter sequences regulated by E2. HPV E2 is known to interact with a number of host proteins including basal- and gene-specific transcription factors and could therefore also regulate SR protein genes indirectly via protein–protein interactions [29]. SRSF1–3 are also highly expressed in HPV-related malignant tumours, however, E2 expression is abrogated in these cells, pointing to a different mechanism of up-regulation in cancer progression [28]. SR proteins are phosphorylated and their subcellular location and activity is controlled by cycles of phosphorylation/dephosphorylation [30]. HR-HPV infection increases not only total levels of SRSF1 but also its phosphorylation in response to epithelial differentiation [25]. The consequences of these changes are under investigation.

SR proteins regulate alternative splicing by binding ESEs (exonic sequence enhancers) to direct and enhance binding of components of the spliceosome at upstream and downstream splice sites [30]. Previous studies using subgenomic HPV16 expressing plasmids identified an ESE in the HPV16 E4 ORF that might regulate splicing of late mRNAs. Using plasmid constructs containing BPV1 (bovine papillomavirus type 1) and HPV16 sequences, Jia et al. [31] identified SRSF3 as a protein factor binding to this ESE. Point mutations that disrupted SRSF3 binding had a negative effect on alternative splicing of L1-encoding mRNAs although the authors did not determine which RNAs were disrupted. A different study utilizing a similar HPV16 subgenomic construct approach identified multiple SRSF1-binding sites in the same region [32]. Inactivation of all putative binding sites redirected splicing to the downstream-located late 3′-splice site suggesting that late mRNA processing is dependent on SRSF1. The subgenomic construct approach yields data useful for mapping of the HPV16 splicing regulatory elements (Figure 1). However, in order to ensure HPV gene expression in the undifferentiated cervical cancer cell lines used in these transfection studies the gene constructs must be under control of a strong heterologous promoter and mutated to remove a negative regulatory element that suppresses late gene expression in undifferentiated epithelial cells [33]. Therefore it is difficult to extrapolate such data to elucidate what regulates alternative splicing from the entire virus genome in the context of the viral life cycle completed in normal differentiated keratinocytes. To address this technical issue, we knocked down cellular SR proteins in undifferentiated and differentiated untransformed keratinocyte cell lines (W12 [34] and NIKS HPV16 [35]) harbouring wild-type episomal HPV16 genomes. Depletion of individual SR family members confirmed their importance for L1 protein expression in
**Figure 1 | Schematic diagram of HPV 16 RNA regulatory elements**

The HPV16 genome (horizontal line) encodes for eight partially overlapping ORFs (grey rectangles containing gene names). Cis-acting RNA processing regulatory elements (vertical bars) and binding proteins (annotated ovals) are shown. ESS, exonic sequence silencer; LRE, late regulatory element. Splice donor site (5′ splice site), broken arrows. Splice acceptor site (3′ splice site), broken arrows. Polyadenylation sites [early/late poly(A)], thin solid arrows. Note: only SR protein family members up-regulated in HR-HPV-infected keratinocytes and heterogeneous nuclear ribonucleoprotein (hnRNP) family members that have been demonstrated to bind viral RNA are shown.

**Figure 2 | Alternatively spliced late mRNAs expressed by HPV16**

(A) ORFs are represented by grey rectangles containing gene names. Promoters are indicated by arrows and polyadenylation sites [early/late poly(A)] are indicated with black vertical bars. (B) A selection of late mRNAs encoded by HPV16 showing the major alternative splicing events required to generate late protein-encoding mRNAs. Grey boxes represent exons and spliced out introns are represented by black lines. The ORFs in each mRNA are indicated to the right-hand side. Adapted with permission from Milligan, S.G., Veerapraditsin, T., Ahamat, B., Mole, S. and Graham, S.V. (2007) Analysis of novel human papillomavirus type 16 late mRNAs in differentiated WI12 cervical epithelial cells. Virology 360, 172–181.

differentiated keratinocytes (T. Klymenko and S.V. Graham, unpublished work).

**Regulation of translation of the capsid mRNAs**

Although it is clear that post-transcriptional control is key in regulating production of HR-HPV proteins, translational control may also be important. Codon usage in the L1 and L2 ORFs could be one of the determinants in differentiation-dependent regulation of translation [36]. A synthetic HPV11 L1 gene with codons substituted by those more common in human genes generated protein levels that were at least 100-fold higher compared with the wild-type HPV11 L1, whereas both genes were transcribed at similar levels in mammalian epithelial and fibroblast cells [37]. However, this interesting finding needs to be tested in the context of the viral life cycle, for example by introducing an episomal codon-modified genome in differentiation-competent keratinocytes.

Translation initiation control by leaky scanning [38] and re-initiation [39] has been proposed to explain efficient expression of polycistronic HPV transcripts. These models imply that a certain distance between the ORFs in a bicistronic transcript is needed for the downstream ORF to be efficiently translated. It is interesting in this regard that in cells where SRSF3 was depleted we noted an increased presence of splice isoforms which contain the L2 ORF followed by the overlapping L1 ORF (Figure 2). Such molecules are probably inefficient in L1 translation. We hypothesize that HPV-dependent up-regulation of SR proteins upon differentiation enables shift of the balance towards RNA splice isoforms more suitable for L1 protein production.

**Future perspectives**

Changes in alternative splicing of HPV mRNAs leads to altered HPV16 transcription patterns, which may be characteristic of HPV-associated disease. Recently, a novel
HPV16 RNA pattern for intelligible grading of cervical lesions was proposed [40]; however, design of this assay is limited to the previously characterized HPV transcriptome. The advent of RNA-Seq (next generation sequencing technology that allows unbiased whole transcriptome interrogation) has brought studies of viral transcriptomes to a whole new level. Recent literature has described various novel viral genes and transcript isoforms determined by RNA-Seq in active and latent infections [41,42]. Such experiments should expand our knowledge of gene expression patterns characteristic of active virus infection and help determine biomarkers useful for disease grading. This will ultimately lead to better disease diagnosis to allow choice of the most appropriate treatment regimes.

Acknowledgements

We thank Margret Stanley and John Doorbar for W12E and NIKS HPV16 cell lines. We are grateful to Hegel Hernandez-Lopez, Rebecca Devine and James Brown for a critical reading of the paper.

Funding

This work was funded by the Wellcome Trust [grant number 0088848/J/09/02].

References

17 Favre, M., Breitburd, F., Croissant, O. and Orth, G. (1977) Chromatin-like structures obtained after alkaline disruption of bovine and human papillomaviruses. J. Virol. 21, 1205-1209
32 Somborg, M. and Schwartz, S. (2010) Multiple ASF/SF2 sites in the human papillomavirus type 16 (HPV-16) E4 coding region promote splicing to the most commonly used 3′-splice site on the HPV-16 genome. J. Virol. 84, 8219-8230

© The Authors Journal compilation © 2012 Biochemical Society