Virology 386 (2009) 32–43

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Human papillomavirus type 16 E2 and E6 are RNA-binding proteins and inhibit in vitro splicing of pre-mRNAs with suboptimal splice sites

Sohrab Bodaghi, Rong Jia, Zhi-Ming Zheng *

HIV and AIDS Malignancy Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA

ARTICLE INFO

Article history:
Received 3 July 2008
Returned to author for revision
18 August 2008
Accepted 26 December 2008
Available online 1 February 2009

Keywords:
Human papillomavirus type 16
RNA splicing
RNA–protein interaction
SR proteins
Protein–protein interaction
Viral proteins

ABSTRACT

Human papillomavirus type 16 (HPV16) genome expresses six regulatory proteins (E1, E2, E4, E5, E6, and E7) which regulate viral DNA replication, gene expression, and cell function. We expressed HPV16 E2, E4, E6, and E7 from bacteria as GST fusion proteins and examined their possible functions in RNA splicing. Both HPV16 E2, a viral transactivator protein, and E6, a viral oncoprotein, inhibited splicing of pre-mRNAs containing an intron with suboptimal splice sites, whereas HPV5 E2 did not. The N-terminal half and the hinge region of HPV16 E2 as well as the N-terminal and central portions of HPV16 E6 are responsible for the suppression. HPV16 E2 interacts with pre-mRNAs through its C-terminal DNA-binding domain. HPV16 E6 binds pre-mRNAs via nuclear localization signal (NLS3) in its C-terminal half. Low-risk HPV6 E6, a cytoplasmic protein, does not bind RNA. Notably, both HPV16 E2 and E6 selectively bind to the intron region of pre-mRNAs and interact with a subset of cellular SR proteins. Together, these findings suggest that HPV16 E2 and E6 are RNA binding proteins and might play roles in posttranscriptional regulation during virus infection.

Published by Elsevier Inc.

Introduction

RNA splicing is an essential step in the control of viral and mammalian gene expression. It occurs immediately after a nascent primary message is transcribed and consists of a series of cascaded biochemical reactions that take place in a spliceosome. Spliceosome-mediated pre-mRNA splicing involves five small U RNAs (U1, U2, U4, U5, and U6) and many splicing factors. The first step in the accurate recognition of intron splice sites involves interaction of the 5′ splice site (5′ ss) with U1, of the branch site with U2, and of the 3′ splice site (3′ ss) with U2AF (U2 auxiliary factor). These interactions are modulated by many cellular splicing factors including SR proteins (Graveley, 2000). SR proteins are a growing family of structurally related and highly conserved cellular splicing factors that are characterized by the presence of an RNA-recognition motif (RRM) and RS dipeptides. A group of classical SR proteins contains extensive phosphorylated RS domains and can be recognized by the monoclonal antibody mAb104 (Zahler et al., 1993). SR proteins are essential splicing factors that modulate the selection of a suboptimal splice site (Zheng, 2004). Many of them have redundant functions.

Human papillomaviruses (HPVs), a group of small DNA tumor viruses, usually infect keratinocytes of skin or epithelial cells of mucosa and cause benign warts or occasionally malignancies (Lowy and Howley, 2001; zur Hausen, 2002). Viral gene expression in infected cells depends on cell differentiation and usually leads to the expression of six nonstructural viral regulatory proteins (E1, E2, E4, E5, E6 and E7) from early regions of the virus genome and two structural viral capsid proteins (L1 and L2) from late regions of the genome. Expression of each of these genes requires extensive RNA splicing. However, the factors that determine when a specific splicing pathway will be used to express a specific gene remain largely unexplored.

The papillomavirus E2 protein is a 42-kDa nuclear protein containing two defined functional domains that are relatively conserved among all papillomaviruses. The N-terminal domain, consisting of ~200 amino acid (aa) residues, is crucial for transcriptional activation, whereas the C-terminal domain, consisting of ~100 aa residues, possesses the DNA binding and dimerization properties of the protein. These two domains are linked by a hinge region that lacks a conserved aa sequence and varies in length among papillomaviruses (Hegde, 2002). The hinge in epidermodysplasia verruciformis-associated HPV E2 contains ~200 aa residues and multiple RS dipeptide repeats, but the hinge in anogenital HPV E2 is ~40–80 aa residues and lacks RS repeats (Sakai et al., 1996; Hegde, 2002). Besides its involvement in papillomavirus DNA replication (Hughes and Romanos, 1993; Sakai et al., 1996; Frattini and Laimins, 1994), E2 is also a transcriptional activator or repressor that regulates the E6 promoter through four consensus E2-binding sites (E2-BSs), ACC(N_6)GGT (Androphy et al., 1987; Hawley-Nelson et al., 1988; Sousa et al., 1990; Romanczuk et al., 1990), upstream of the viral E6 promoter. However, E2 functions as a repressor at steps after TBP or TFIID binding (Hou et al., 2000) and its transcriptional repression occurs only in cells harboring integrated,

* Corresponding author. HIV and AIDS Malignancy Branch, Center for Cancer Research, NCI/NIH, 10 Center Dr., Rm. 6N106, MSC-1868, Bethesda, MD 20892-1868, USA. Fax: +1 301 480 8250.
E-mail address: zhengt@exchange.nih.gov (Z.M. Zheng).

0042-6822/$ – see front matter. Published by Elsevier Inc.
doi:10.1016/j.virol.2008.12.037

but not episomal HPV16 DNA (Bechtold et al., 2003), raising a question about how E2 might function. In general, HPV16 E2 (16E2) by transient transfection is toxic to mammalian cells and is usually under detection level, suggesting other functions of E2 in the induction of cell toxicity. The finding that HPV5 E2 (5E2) facilitates RNA splicing and interacts with SR proteins (Lai et al., 1999) suggests that HPV E2 might play a role at post-transcriptional level. RNA transcription and splicing are coupled processes and many transcription factors have unexpected roles in this coupled network (Maniatis and Reed, 2002).

HPV E4 is expressed as an E1^E4 protein of 92 aa residues in which the N-terminal 5 aa residues are derived from the E1 ORF spliced to the E4 ORF. The E4 protein which is ~10 kDa in size is the most abundantly expressed HPV protein and accumulates in differentiating cells of the upper epithelial layers. E4 expression collapses the cytokeratin network (Doorbar et al., 1991) and mediates cell cycle arrest in G2 (Davy et al., 2002; Nakahara et al., 2002). The E1^E4 protein of HPV16 also binds to a DEAD-box containing RNA helicase, but the function of this association remains unknown (Doorbar et al., 2000). Since DEAD-box proteins regulate gene expression mostly at post-transcriptional levels, this observation implies that HPV16 E1^E4 might be involved in functions other than its effect on cytokeratin.

HPV16 E6 (16E6) and E7 (16E7) are viral oncoproteins that inactivate, respectively, cellular p53 and pRB, two tumor suppressor proteins essential for cell cycle control. 16E6 is an ~18 kDa nuclear protein and is composed of 151 aa residues. 16E6 contains two hypothetical zinc fingers involved in zinc binding (Kanda et al., 1991) and three nuclear localization signals (NLS) (Tao et al., 2003) as well as a PDZ-binding site in the N-terminus (Kiyono et al., 1997; Lee et al., 1997). Besides its ability to immortalize and transform cells and induce p53 degradation, 16E6 is also functionally involved in the regulation of gene transcription (Desaintes et al., 1992; Klingelhutz et al., 1996) through interaction with other transcription factors and coactivators (Patel et al., 1999; Ronco et al., 1998; Kumar et al., 2002; Veldman et al., 2001; Veldman et al., 2003; Thomas and Chiang, 2005). However, it remains to be determined what part of 16E6 is responsible for these protein–protein interactions because 16E6, like16E2, is usually under detection level in cells after transient transfection. 16E7 is a nuclear protein with 98 aa residues in size. The N-terminal 37 aa residues of 16E7 have been characterized as an important portion of the protein that contributes to pRB binding and degradation as well as cell transformation. Similar to 16E6, 16E7 also interacts with cellular transcription factors and coactivators (Massimi et al., 1997; Avvakumov et al., 2003; Bernat et al., 2003; Huang and McCance, 2002). Many of these interactions appear to involve the C-terminal half of 16E7, but their biological relevance remains to be understood. In addition to the protein–protein interactions of both 16E6 and 16E7, 16E6 has been shown to be a DNA binding protein (Imai et al., 1989; Ristriani et al., 2000; Ristriani et al., 2001; Nomine et al., 2003); the function of this DNA binding also remains unknown.

We recently showed that GFP-16E6 fusion is a nuclear and specifically, a nucleolar protein (Tao et al., 2003), and its nuclear or nucleolar localization is controlled by its three NLSs. Since the nucleoli are enriched with various small RNAs and NLS3 seems to strengthen, rather than to promote, the nuclear localization of the protein, we proposed that the major function of NLS3 might be to retain the protein in the nucleus, perhaps through an interaction with a nucleic acid (DNA or RNA) (Tao et al., 2003). Based on these observations and difficulties to express detectable HPV16 E2 and E6 in mammalian cells, we expressed E2, E4, E6 and E7 as GST-fusion proteins from individual HPV16 ORFs in *E. coli* and analyzed their possible interactions with RNAs and involvements in post-transcriptional regulation. We report here that both 16E2 and 16E6 are RNA binding proteins. 16E2 binds RNA through its characteristic DNA-binding domain and 16E6 binds RNA via its C-terminal half. Perhaps most importantly, both 16E2 and 16E6 were found to suppress RNA splicing and to interact with SR proteins.

Results

Expression of HPV proteins as GST-fusion proteins in E. coli

Various expression conditions were used to efficiently express each viral protein as a GST fusion in two bacteria strains, BL21(DE3) pLysS and BL21 Codon Plus (DE3)-RP, a bacteria strain harboring extra copies of tRNA genes that are rare in *E. coli*, but common in humans. Optimized expression conditions were obtained empirically for each protein and are summarized in Table 1. HPV16 E2 and E6, as well as some of their truncated mutants, were difficult proteins to express in BL21(DE3)pLysS. When expressed in this bacteria strain at room temperature or at 37 °C, they tended to form inclusion bodies and were toxic to the bacteria. Especially when 16E2 was expressed, the transformed bacteria had difficulty reaching the OD_{600} value (0.6) required for the initiation of IPTG induction and frequently collapsed during early exponential growth. However, 16E2, 16E6, and some of their mutants were expressed relatively more efficiently, with fewer inclusion bodies, in the BL21 Codon Plus (DE3)-RP strain. Other proteins, such as E4 and E7 were relatively easy to express in either bacteria strains at room temperature or at 37 °C with a few hours of induction.

Effects of HPV16 E2, E4, E6 and E7 on RNA splicing

HPV16 E2-, E4-, E6- and E7-GST fusion proteins extracted to near homogeneity (Fig. 1A) were examined for their possible role in RNA splicing. We utilized an RNA substrate, HPV16 E6E7 pre-mRNA(Zheng et al., 2004), containing an exon 2 that was short but had a U1 binding site attached as a splicing enhancer (Fig. 1B). The viral fusion proteins (at 0.1, 1.0 or 5.0 μg) (5 μg of semipurified GST fusions were equivalent to approximately 500 ng of an expected pure viral fusion protein, see Materials and Methods) were pre-mixed with HeLa or 293 nuclear extracts (NE), followed by the addition of the RNA substrate to be tested in a splicing condition. As shown in Fig. 1C, HPV16 E6E7 RNA was spliced efficiently in the presence (lanes 2–4) or absence (lane 1) of 0.1, 1.0 or 5 μg of GST, but its splicing efficiency was reduced in the presence of 16E2- and 16E6-GST fusion proteins in a dose-dependent manner, reaching up to 3-fold reduction at 5 μg (equal to only 4% of total NE protein in a splicing reaction) (lanes 7 and 13), indicating that the 16E2- and 16E6-GST fusions are suppressive. HPV16 E4- and E7-GST fusions at 5 μg exhibited a smaller suppressive effect (less than 2-

Table 1
Optimized conditions for expressing HPV–GST fusion proteins in *E. coli*[a]

GST fusions	IPTG (mM)[b]	Induction at °C	Induction time (h)
16E2 wt	0.7	18	20
16E2ΔC227	1	25	2
16E2ΔN220	0.5	18	20
16E2ΔN259	1	18	20
16E2ΔC260	1	25	2
16E2ΔN220+C260	1	25	2
16E6 wt	0.25	25	3
16E6 mt[c]	0.25	25	3
16E6ΔN41	0.7	25	2
16E6ΔN102	1	25	2
16E6ΔN102 mt[d]	1	25	2
16E6ΔC103	1	25	2
16E6ΔC42	1	25	2
16E6ΔN41+ΔC103	1	18	20
16E4 wt	1	25	2
16E7 wt	1	37	2
5E2 wt	0.7	18	20
6E6 wt	0.5	25	3
GST	1	37	2

[a] BL21 Codon Plus (DE3)-RP strain.
[b] Final concentration.
[c] Point mutations in both NLS1 and NLS3 motifs (Tao et al., 2003).
[d] Point mutations in NLS3 motif (Tao et al., 2003).