The E1 proteins

Monika Bergvall a,b, Thomas Melendy c, Jacques Archambault a,b,d,*

a Molecular Virology Laboratory, Institut de recherches cliniques de Montréal (IRCM), 110 Pine avenue West, Montreal, QC, Canada H2W 1R7
b Department of Microbiology and Immunology, McGill University, Montreal, QC, Canada
c Departments of Microbiology and Immunology, Biochemistry and the Witebsky Center for Microbial Pathogenesis, University of Buffalo, 210 Biomedical Research Building, School of Medicine and Biomedical Sciences, Buffalo, NY 14214, USA
d Department of Biochemistry and Molecular Medicine, Université de Montréal, Montreal, QC, Canada

Available online 10 September 2013

Keywords:
Papillomavirus
E1
DNA replication
Helicase
Post-translational modifications
Phosphorylation
Episome

ARTICLE INFO

Abstract

E1, an ATP-dependent DNA helicase, is the only enzyme encoded by papillomaviruses (PVs). It is essential for replication and amplification of the viral episome in the nucleus of infected cells. To do so, E1 assembles into a double-hexamer at the viral origin, unwinds DNA at the origin and ahead of the replication fork and interacts with cellular DNA replication factors. Biochemical and structural studies have revealed the assembly pathway of E1 at the origin and how the enzyme unwinds DNA using a spiral escalator mechanism. E1 is tightly regulated in vivo, in particular by post-translational modifications that restrict its accumulation in the nucleus. Here we review how different functional domains of E1 orchestrate viral DNA replication, with an emphasis on their interactions with substrate DNA, host DNA replication factors and modifying enzymes. These studies have made E1 one of the best characterized helicases and provided unique insights on how PVs usurp different host-cell machineries to replicate and amplify their genome in a tightly controlled manner.

© 2013 Elsevier Inc. All rights reserved.

Introduction

The hexameric DNA helicase E1 is the only enzyme and the most conserved protein encoded by papillomaviruses (PVs). This high degree of conservation reflects the essential role that E1 plays during the viral life cycle, namely as the replicative helicase that orchestrates the faithful copying of the viral episome in the nucleus of infected keratinocytes. E1 is thought to be required throughout the viral replicative cycle, first to increase the copy number of the viral episome upon infection of basal keratinocytes, then to maintain a constant level of episomes in cells that are displaced upward in the epithelium and begin to differentiate, and, finally, to promote amplification of the viral genome during the productive phase of the viral life cycle that takes place in the uppermost differentiated epithelial layers. Studies in cell culture have demonstrated that E1 is essential for establishing the viral genome as a multicopy episome in the nucleus of transfected keratinocytes, a process that is believed to mimic, albeit only partially, what happens to the viral DNA during the early stages of infection. Cell culture studies have also supported the notion that E1 is required for the amplification of the viral genome that occurs upon cellular differentiation; some of the evidence coming from the identification of E1 mutant proteins that are specifically defective for this process. However, there remains some controversy as to whether E1 is also needed to maintain the viral episome at a constant copy number in undifferentiated cells; some studies suggesting that once the viral genome has been established as a multicopy episome, its replication and maintenance in proliferating keratinocytes become E1-independent (Egawa et al., 2012; Kim and Lambert, 2002).

E1 is a classical initiator protein and, as such, plays several roles in the initiation and catalysis of viral DNA synthesis. E1 must first recognize a specific segment of the viral genome known as the “origin of DNA replication”, or “ori” for short. For most PV types examined to date, the minimal ori sequence that can support viral DNA replication in transient assays maps to the 3’ portion of the viral long control region (LCR), upstream of the early genes, and is typically comprised of two to three E2-binding sites, a palindromic E1-binding region and an AT-rich sequence, all of which are required for optimal ori function (Lee et al., 1997; Lu et al., 1993; Raj and Stanley, 1995; Russell and Botchan, 1995; Santucci et al., 1995; Spalholz et al., 1993; Sun et al., 1996; Sverdrup and Khan, 1995; Ustav et al., 1991). As reviewed in detail below, the key step in the initiation of viral DNA replication at the ori is the assembly of E1 into its enzymatically active form, a double-hexameric helicase capable of unwinding the ori and the DNA ahead of the replication fork, in an ATP-dependent manner. E1 also engages in multiple interactions with specific host factors to orchestrate
the assembly of a functional replisome needed for bi-directional replication of the viral genome. The absolute reliance of E1 on host DNA replication factors for function has contributed to making PV DNA replication a model system for the study of eukaryotic DNA synthesis. Furthermore, the central role of E1 in this process and the availability of crystal structures of the enzyme have made E1 the best-studied member of the superfamily III (SF3) of helicases (Hickman and Dyda, 2005).

In this article, we will review the structure and function of E1 with an emphasis on the mechanism of DNA unwinding and the identification of functional domains of the protein that mediate key protein–protein and protein–DNA interactions required for replication of the PV genome. Like most DNA helicases, E1 is very stringently regulated, in particular by post-translational modifications that prevent its over-accumulation in the nucleus. These regulatory mechanisms and how their failure may contribute to integration of the viral genome during carcinogenesis will also be reviewed.

**Domain structure of E1**

The first insights into the structure and function of E1 came from the observation that it shares sequence similarity with the initiator proteins of other DNA viruses such as the large T antigen (LT-Ag) of simian virus 40 (SV40) and of other polyomaviruses (Cler tant and Seif, 1984). Four conserved regions, termed A, B, C and D, were identified in the C-terminal regions of E1 and LT-Ag (Fig. 1, lower panel), which suggested that E1 may function as an ATPase (Cler tant and Seif, 1984). Subsequent studies confirmed that purified recombinant E1 displays ATPase activity and is in fact a hexameric DNA helicase with 3′ to 5′ directionality (Fouts et al., 1999; Hughes and Romanos, 1993; Jenkins et al., 1996; Raj and Stanley, 1995; Rocque et al., 2000; Santucci et al., 1995; Sedman and Stenlund, 1998; Seo et al., 1993; Sheikh et al., 2003; White et al., 2001; Yang et al., 1993). Much of our understanding of the structure and function of E1 has been gathered from the study of the prototypical bovine papillomavirus type 1 (BPV1) E1 and, more recently, of E1s from prevalent anogenital human papillomavirus types (HPV6, 11, 16, 18, 31 and 33) and the cutaneous virus HPV1. Thus, one must keep in mind that our understanding of the structure and function of E1 emanates from E1 proteins from a very limited set of viruses. Although some of the basic principles underlying the mechanisms of viral DNA unwinding and replication have likely been highly-conserved during evolution, the fine-tuning and regulation of E1 activity may have evolved more rapidly to accommodate differences in the life cycles of different PV types.

A significant issue that has limited the biochemical characterization of the E1 helicase to only a few PV types has been the difficulty in expressing and purifying large quantities of the protein in a recombinant form that can support cell-free DNA replication. This is in part due to the fact that E1 must assemble from monomers at the ori in order to be active for DNA replication but is typically purified as large pre-formed oligomers when overexpressed in heterologous systems such as bacteria and insect cells. A notable exception has been the BPV1 E1 protein which can be readily purified from bacteria or insect cells in monomeric form or as oligomers that are in a monomer–hexamer equilibrium; a feature that has contributed to making BPV1 E1 the preferred enzyme for biochemical studies (Bonne-Andrea et al., 1995a; Fouts et al., 1999; Melendy et al., 1995; Mohr et al., 1990; Müller et al., 1994; Sedman and Stenlund, 1998). In contrast and as mentioned above, overexpression of E1 from other PV types often results in enzyme preparations comprised mostly of oligomers rather than monomers. For example, HPV11 E1 overproduced in insect cells using a baculovirus-expression system is purified mostly as hexamers that are not easily dissociated into monomers and thus poorly active in cell-free DNA replication, despite displaying significant levels of ATPase and short-duplex DNA unwinding activities (Dixon et al., 2000; Rocque et al., 2000; White et al., 2001). Fortunately, this issue has not been insurmountable and protocols have been developed to obtain HPV11 E1 from insect cells in a replication-competent form (Kuo et al., 1994). Monomeric and active HPV11 E1 can also be obtained by expression of the protein in vitro, by coupled transcription and translation in a rabbit reticulocyte lysate, a facile approach that should be easily adaptable to the E1 proteins from other PV types (Amin et al., 2000).

E1 is encoded by the largest and most conserved open-reading frame (ORF) of the PV genome. The protein ranges in size from 600 to 650 amino acids, depending on the PV type. Overall, the protein

---

**Fig. 1. Domain structure of E1.** Diagram of the BPV1 E1 protein highlighting the functions and amino acid boundaries of its various domains. The N-terminal regulatory region is shown in light purple while the domains required for viral DNA-replication in vitro are colored in deep purple. The diagram shows the locations of the bipartite nuclear localization signal (NLS), DNA binding domain (DBD), minimal oligomerization domain (O), AAA+ ATP-binding domain and C-terminal brace. The latter three regions comprise the helicase domain (HD), as indicated. Below the HD are shown the positions of the Walker A (WA), Walker B (WB), β-hairpin (β), sensor 1, 2 and 3 (S1, S2 and S3, respectively), and arginine-finger (R). The locations of the four regions (A–D) of homology with the large T antigen of SV40 are also indicated. The bottom part of the figure summarizes the main functions of the DBD and HD.
can be divided into three functional segments: an N-terminal regulatory region that is essential for optimal replication in vivo but dispensable in vitro (Amin et al., 2000; Ferrand and McBride, 1998; Morin et al., 2011; Sun et al., 1998), a central origin-binding domain (known as the DNA-binding domain, DBD) that recognizes specific sites in the ori (Auster and Joshua-Tor, 2004; Chen and Stenlund, 1998; Leng et al., 1997; Sarafi and McBride, 1995; Sun et al., 1998; Thorner et al., 1993; Titolo et al., 2003), and a C-terminal enzymatic domain sufficient for self-assembly into hexamers that display ATPase activity and are capable of unwinding short DNA duplexes (Fig. 1) (Castella et al., 2006b; Enemark and Joshua-Tor, 2006; Titolo et al., 2000; White et al., 2001). The DBD and C-terminal helicase domain (HD) are sufficient for ori-dependent DNA replication in vitro and form the core of the molecular motor that drives viral DNA replication (Amin et al., 2000).

N-terminal regulatory region

The N-terminal region of E1, comprised of approximately 200 amino acids, is the least conserved segment of the protein. This region contains a number of short amino acid sequence motifs that are variably conserved amongst different PVs, including a bipartite nuclear localization signal (NLS), a Crm1-dependent nuclear export signal (NES), a cyclin-binding motif (CBM) that interacts with cyclin A/E in complex with cyclin-dependent kinase 2 (Cd2k) and phosphorylation sites for Cd2k and other kinases (Deng et al., 2004; Fradet-Turcotte et al., 2010a; Leng et al., 1994; Lentz et al., 1993, 2006; Ma et al., 1999; McShan and Wilson, 1997; Yu et al., 2007). The less stringent evolutionary conservation of the E1 N-terminal region likely reflects the adaptation of different PVs for their respective host and the type of epithelium they infect. For instance, it is easy to imagine that E1 may be regulated slightly differently in cutaneous versus mucosal tissue. Examples of how the N-terminal region regulates the nucleo-cytoplasmic shuttling of E1 slightly differently for different PV types are provided later in this review article. From a structural biology perspective, little is known about this N-terminal region, likely because it does not assume a compact structure as suggested by its accessibility to proteases. Using NMR, a short 82 amino acid region at the N-terminus of HPV31 E1 has been found to be unstructured and may constitute a genuine intrinsically disordered domain (IDD) (Lehoux et al., 2012). IDDs are common in viral proteins as they can accommodate interactions with several different protein partners using a folding-upon-binding mechanism. IDDs tend to be enriched in charged residues and sites of post-translational modifications, such as phosphorylation, as is the case for the E1 N-terminus. The putative IDD of HPV31 E1 contains a binding site for the cellular protein p80/Uaf1 (Côté-Martin et al., 2008, discussed later in this review) followed by a short and conserved hydrophobic motif, α-ω-φ-ω-φ (where φ are hydrophobic amino acids) with a propensity to fold as an amphipathic α-helix (Morin et al., 2011). A triple amino acid substitution in this motif (ααα 45–51, MVDFI changed to AVDAA) was found to reduce transient viral DNA replication approximately 2-fold, at a step following assembly of the E1–E2–ori pre-initiation complex, consistent with this putative amphipathic helix playing a stimulatory, yet unknown role in viral DNA replication (Morin et al., 2011). Thus, a picture is emerging where the N-terminal region of E1 may be a privileged segment of the protein for easily accessible molecular recognition features (MoRfs) and short linear motifs (SiMs) involved in regulating E1.

DNA-binding domain (DBD)

The DBD, sometime referred to as the origin-binding domain (OBD), has been extensively characterized in vitro and was the first E1 domain to be crystallized (Auster and Joshua-Tor, 2004; Enemark et al., 2000, 2002). As its name indicates, the DBD is a DNA binding domain that recognizes specific sequences in the ori, albeit with low affinity. The E1 binding region within the ori was first characterized as an 18-nucleotide AT-rich imperfect palindrome (Chen and Stenlund, 2001; Holt et al., 1994; Holt and Wilson, 1995; Mendoza et al., 1995; Sun et al., 1996; Ustav et al., 1991). Further analysis of this region indicated that it contains six E1-binding sites (E1BS 1–6) of the consensus sequence 5′-ATTGT-3′, although some sites can be degenerate and of lower affinity (Chen and Stenlund, 2001; Titolo et al., 2003). Systematic mutagenesis of a single E1BS in vitro revealed that the E1 DBD binds with higher affinity to sites of the following consensus sequence: 5′-AT(A/G)/TGC(C/T)/(C/T)-3′, highlighting the importance of the A1, T2 and G4 positions for binding (Titolo et al., 2003). The six E1BS present in the BPV1 origin are shown in Fig. 5A. Four of those six E1BS correspond to the 18-bp palindrome mentioned above and are arranged as two pairs of overlapping inverted repeats (E1BS1 and 3 or E1BS2 and 4), with both E1BS within each repeat being separated from each other by three nucleotides (Chen and Stenlund, 2001; Titolo et al., 2003). As will be further discussed below, these two pairs of E1BS can each support dimerization of the E1 DBD. The remaining two sites, E1BS 5 and 6, are also conserved in the ori of several PVs and extend by 3-bp on either side of the palindrome; however, because they are not paired, they do not support the dimerization of E1 (Fig. 5A) (Chen and Stenlund, 2001). Rather, E1BS 5 and 6 are thought to cooperate with the other E1BS to orchestrate the assembly of E1 as a double-trimer at the ori; a key intermediate complex in the assembly of a functional double-hexamer (see section on the initiation of PV DNA replication for more details) (Chen and Stenlund, 2001; Titolo et al., 2003). Thus, it is believed that the arrangement of the six E1BS in the ori forms the basis for how six E1 molecules assemble into a double-trimer at the ori.

The crystal structure of the E1 DBD from BPV1 (residues 159–303) and, subsequently, that of the HPV18 DBD (210–354) revealed that this domain adopts a fold similar to that of the SV40 LT-Ag OBD, despite very limited sequence similarity between both domains (Auster and Joshua-Tor, 2004; Enemark et al., 2000). The BPV1 E1 DBD consists of a central five-stranded anti-parallel β-sheet flanked by loosely packed α helices on one side (α1, α2, α5 and α6) and more tightly packed helices (α3 and α4) on the other (Fig. 2A) (Enemark et al., 2000). Previous mutational studies had demonstrated the importance of several key residues for DNA-binding (Gonzalez et al., 2000; Thorner et al., 1993), which, when mapped onto the structure of the DBD, defined two conserved regions involved in DNA recognition. These two regions are comprised, respectively, of a loop located between α2 and β1 and designated as the “DNA-binding loop”, and of helix α4 referred to as the “DNA-binding helix” (Fig. 2A) (Enemark et al., 2000). Together, these two regions form a continuous and positively charged area on the surface of the protein. Subsequent crystal structures of the DBD bound to DNA indicated that the DNA-binding loop contacts the first three nucleotides of the E1BS (5′-ATT-3′) while the DNA binding helix interacts with the remaining three residues of the opposite DNA strand (5′-AAC-3′) (Enemark et al., 2002). Most of DBD-DNA contacts involved electrostatic interactions with the phosphate backbone, explaining the relatively poor specificity of the DBD for its target site. For the DBD from BPV1, and likely also for that of HPV11, most of the sequence specificity arises from Van der Waal interactions between the DNA-binding loop and the highly-conserved second thymidine (T2) of the E1BS (Enemark et al., 2002). This mode of protein–DNA interaction may not apply to all PV types, however. Indeed, the crystal structure of the HPV18 E1 DBD suggested that this DBD lacks specific contacts with T2, a finding consistent with the increased affinity of this protein for
non-specific DNA (Auster and Joshua-Tor, 2004). Several amino acid substitutions that impair the DNA-binding activity of the DBD have been reported (Gonzalez et al., 2000; Thorner et al., 1993; West et al., 2001; West and Wilson, 2002).

While monomeric in solution, the DBD can dimerize in vitro on DNA substrates containing two inverted E1BS spaced by 3 base pairs, as found in the ori (Chen and Stenlund, 2001; Enemark et al., 2000; Titolo et al., 2003). Dimerization increases the affinity of the DBD for its target sequence by approximately 10-fold (Fig. 2B) (Enemark and Joshua-Tor, 2002). The structure shows the mode of DNA-binding and dimerization. It also shows that the two DBD dimers are bound on separate faces of the double-helix and do not interact with each other.

Non-specific DNA (Auster and Joshua-Tor, 2004). Several amino acid substitutions that impair the DNA-binding activity of the DBD have been reported (Gonzalez et al., 2000; Thorner et al., 1993; West et al., 2001; West and Wilson, 2002).

While monomeric in solution, the DBD can dimerize in vitro on DNA substrates containing two inverted E1BS spaced by 3 base pairs, as found in the ori (Chen and Stenlund, 2001; Enemark et al., 2000; Titolo et al., 2003). Dimerization increases the affinity of the DBD for its target sequence by approximately 10-fold (Titolo et al., 2003). As most PVs contain two such pairs of inverted E1BS, their ori could in principle support the assembly of two separate DBD dimers. Crystal structures of the BPV1 DBD bound to two pairs of E1BS confirmed this possibility and indicated that the two DBD dimers assemble on separate faces of the DNA and do not interact substantially with each other (Fig. 2B) (Enemark et al., 2002). The DBD dimerization interface is relatively small (∼500 Å²), hydrophobic and involves conserved helix α3 (Fig. 2A) (Enemark et al., 2000). Mutation of a key residue within α3 of the BPV1 DBD (A206R) was found to prevent dimerization in vitro and to greatly reduce the ability of full-length E1 to support transient DNA replication (Enemark et al., 2000; Schuck and Stenlund, 2005b). Similar results were obtained for the DBDs of HPV11 and HPV31 E1 (Fradet-Turcotte et al., 2010b; Titolo et al., 2003). The crystal structure of the HPV18 DBD revealed that it also dimerizes via its α3 helix, although the latter is shorter than in other DBDs. This structural difference may contribute to the higher affinities of the HPV18 DBD for non-specific DNA, for example if it were to increase its propensity to dimerize (Auster and Joshua-Tor, 2004; Titolo et al., 2003). Collectively, these findings suggest that the DBD dimerization-interface promotes viral DNA replication by enhancing the affinity and specificity of E1 towards the ori. As will be further discussed below, the ability of the DBD to bind DNA as a dimer is required for the assembly of E1 as a replication-competent double-hexamer (DH) at the ori.

Surface mutagenesis of the BPV1 E1 DBD was used to identify other functions of this domain unrelated to DNA-binding (Schuck and Stenlund, 2006). This approach led to the identification of 12 amino acid substitutions that impaired the ability of E1 to support transient DNA replication while having little to no effect on expression of the protein in transfected cells. Interestingly, 8 of the 12 substitutions did not affect the ability of E1 to bind the ori in vitro, suggesting that the function of the DBD in replication is indeed not limited to DNA-binding. These substitutions could be subdivided into three classes based on functional assays. One class comprised substitutions that, despite affecting the activity of E1 in vivo, had little effect on its activity in vitro in a cell-free DNA replication system (T188A, N252A, S280A, S281A). Accordingly, these substitutions did not affect the helicase and unwinding activities of E1 measured in separate assays. It was hypothesized that these substitutions could affect the regulation of E1 in vivo, perhaps by phosphorylation as 3 of the 4 affected residues are potential phospho-acceptor sites. The second class comprised substitutions that impaired the ability of E1 to support ori-dependent DNA replication in vitro but which had little effect on its helicase and unwinding activities (V256A, L268A). These substitutions may affect the ability of E1 to interact with specific host DNA replication factors such as replication protein A (RPA), topoisomerase I (Topo I) and DNA polymerase α-primase (Pol α-prim). The third class contained amino acid substitutions that also affected the replication activity of E1 in vivo but in this case by specifically inhibiting the ability of E1 to assemble as a double-hexamer at the ori (R269A, K279A). As will be discussed later in this review, assembly of the E1 double-hexamer proceeds via a double-trimer intermediate, whose formation was also impaired by the R269A, K279A substitutions. Collectively, these studies highlighted the many different roles that the E1 DBD plays in viral DNA replication in addition to DNA-binding and dimerization.

Helicase domain (HD)

When produced in recombinant form, the HD of E1 is sufficient for assembly into hexamers that have short duplex DNA unwinding activity, as shown for the BPV1 and HPV11 enzymes (Castella et al., 2006b; White et al., 2001). This domain can be subdivided into three functional regions that cooperate to form an active enzyme: the minimal oligomerization domain, the AAA+-ATP-binding module and the C-terminal brace. As Two crystal structures of the BPV1 E1 HD in its hexameric form have been reported: one of the hexamer bound to ssDNA and nucleotide cofactors and the other of the free (i.e. apo) hexamer (Enemark and Joshua-Tor, 2006; Sanders et al., 2007). The different functional regions of the E1 HD are discussed below in light of these structures.

Oligomerization domain

In vitro, E1 can oligomerize either into a double-hexamer in the presence of ori DNA and ATP (reviewed in a later section) or into a
DNA unwinding. Mutagenesis of the Walker A, Walker B, C motif, energy of ATP hydrolysis to conformational changes required for helicases, the arginine motifs provided by the adjacent monomer. By analogy with other monomer, each ATPase site is comprised of an (Fig. 3). In addition to the Walker A, Walker B and C motif from one contribute amino acids involved in ATP-binding and/or hydrolysis (Abbate et al., 2004; Castella et al., 2006b; Titolo et al., 2000; White et al., 2001). This domain contains the four highly conserved regions termed A, B, C and D that share significant similarities with the LT-Ag of SV40 and of other polyomaviruses (Clerant and Seif, 1984). Three of these conserved regions have been identified as important for oligomerization of E1, namely regions A, B and C, the latter two forming part of the ATPase catalytic site (Fig. 1) (Titolo et al., 2000). The smallest fragment of HPV11 E1 that could self-assemble in the yeast two-hybrid system was comprised of amino acids 353–431 (corresponding to aa 309–386 in BPV1 E1); this region spans conserved region A and constitutes the minimal oligomerization domain (Fig. 1) (Titolo et al., 2000). In the crystal structures of the BPV1 E1 hexamer, the six oligomerization domains form a rigid collar N-terminal to the ATP-binding sites (Fig. 3). Mutations of highly conserved amino acids within region A inhibit the ability of E1 to oligomerize in vitro and to support transient DNA replication in vivo (Titolo et al., 2000). Likewise, mutations in conserved regions B and C, which lie outside or the oligomerization domain and form part of the ATP-binding site, confirmed their importance for E1 oligomerization, consistent with ATP playing a stimulatory role in this process (Titolo et al., 2000).

AAA+ ATP-binding module

E1 is a member of the SF3 family of helicases, which also include other helicases encoded by small DNA and RNA viruses (Gorbulew et al., 1990; Hickman and Dyda, 2005). This family is characterized by several highly conserved sequence motifs, namely the Walker A, Walker B, and C motifs (not to be confused with conserved regions A-D, Fig. 1) that are characteristic of AAA+ proteins (ATPase associated with various cellular activities) (Koonin, 1993; Ogura and Wilkinson, 2001). For BPV1 E1, this AAA+ module is located within amino acids 387–579, between the minimal oligomerization domain and the C-terminal brace (Fig. 1). The Walker A motif (consensus amongst E1 sequences: G-x3-G-K- [T/S]), aa 433–440 in BPV1 E1), also known as phosphate-binding loop or P-loop, and the Walker B motif (consensus Φ-Φ-D-D, aa 476–479) form the core of the catalytic site that binds and hydrolyzes ATP as a magnesium chelate. The C motif (consensus Φ-[T/S]-[T/S]-N, aa 520–523), also known as sensor 1, is characterized by the presence of an invariant asparagine residue that contacts the Walker B and the γ-phosphate of ATP. In hexameric DNA helicases such as E1, each of the six ATPase catalytic sites is actually formed at the junction of two monomers, which both contribute amino acids involved in ATP-binding and/or hydrolysis (Fig. 3). In addition to the Walker A, Walker B and C motif from one monomer, each ATPase site is comprised of an “arginine finger” (R538 in BPV1 E1) and of sensor 2 (K425) and sensor 3 (R493) motifs provided by the adjacent monomer. In analogy with other helicases, the arginine finger likely plays a key role in coupling the energy of ATP hydrolysis to conformational changes required for DNA unwinding. Mutagenesis of the Walker A, Walker B, C motif, “arginine finger” and sensors 2 and 3 confirmed their importance for ATP-binding and/or hydrolysis (Abbate et al., 2004; Castella et al., 2006a; Hughes and Romanos, 1993; Jenkins et al., 1996; Lin et al., 2002; Liu and Stenlund, 2010; MacPherson et al., 1994; Mansky et al., 1997; Raj and Stanley, 1995; Rocque et al., 2000; Titolo et al., 1999; White et al., 2001). In addition to its role in ATP-binding and hydrolysis, the AAA+ module of E1 also contains specific determinants involved in non-specific binding to single- and double-stranded DNA, in particular a β-hairpin structure (aa 504–508 for BPV1 E1) and a hydrophobic loop (aa 457–467). These DNA-binding elements are critical for the assembly of E1 at the origin and for its subsequent DNA unwinding activity, as discussed in later sections of this review article.

C-terminal brace

The HD contains a short flexible region at its extreme C-terminus that plays a role in the assembly and stabilization of the E1 hexamer (Fig. 1) (Whelan et al., 2012). In the crystal structure of the complete HD of BPV1 E1, the C-terminal 26 amino acids are not resolved indicating that they are flexible and unstructured, a notion that is further supported by the sensitivity of these residues to protease digestion (Sanders et al., 2007). Deletion of this C-terminal extension was found to impair, albeit not completely, the assembly and stability
of E1 hexamers in vitro and, consequently, to reduce the unwinding activity of E1 particularly on longer DNA substrates (Whelan et al., 2012). Analysis of the solution structure of BPV1 E1 hexamers by Small Angle X-ray Scattering (SAXS) indicated that the negatively charged C-terminal extension from each monomer contacts a positively charged cleft on the adjacent subunit and acts as a brace to stabilize the E1 hexamer (Whelan et al., 2012). It was suggested that this brace may be important for maintaining the oligomeric state of E1 during conformational changes induced by ATP-binding and hydrolysis. As the E1 proteins from all PV types contain a C-terminal extension similar to that of BPV1 E1, this mechanism of hexamer stabilization appears to be generally conserved.

Crystal structure of the E1 DNA helicase and the mechanism of DNA unwinding

Staircase mechanism

The crystal structure of the BPV1 E1 DNA helicase in its hexameric, nucleotide-bound and ssDNA-bound state was a watershed event in the understanding of how members of this family of AAA+ replicative DNA helicases function (Enemark and Joshua-Tor, 2006). Until the publication of this structure, there was much confusion about whether the central pore in these hexameric DNA helicases encircled ssDNA or dsDNA. With the publication of this structure, and re-analyses of other AAA+ DNA helicase structures in light of this new information, the general consensus is that these hexameric replicative DNA helicases form a donut or torus around a ssDNA template and operate as motor proteins that travel along one of the two DNA strands (Enemark and Joshua-Tor, 2006). As such, the unwinding activity of these helicases is not so much that of “strand-separating” as it is the resultant displacement of any nucleic acid annealed to the DNA strand upon which the helicase is translocating.

The structure by Enemark and Joshua-Tor (2006) clearly demonstrated that the central pore of the hexameric BPV1 E1 helicase contains ssDNA. Within this pore, six sequential sugar-phosphate residues of the ssDNA backbone get bound by the AAA+ DNA binding β-hairpins of the six E1 monomers comprising the hexamer (albeit with some interaction of histidine 507 with the phosphate of the adjacent nucleotide) (Fig. 4). Each of the six DNA-binding hairpins exists in a slightly different position relative to the DNA helicase torus, with the E1 monomer in the ATP-bound conformation having the hairpin positioned most closely to the face where the 5′ end of the ssDNA extends. The adjacent E1 monomer exists in an ADP-bound state, with the DNA binding hairpin shifted slightly towards the opposite face of the torus. The next E1 monomer also exists in an ADP-bound state, but with the hairpin shifted further away from the 5′ face of the torus. Subsequent E1 monomers generally also have ADP bound, but in an increasingly looser conformation, in preparation for ADP release. Conformational changes in E1 associated with the more loosely bound nucleotide result in their respective DNA binding hairpins being shifted further toward the 3′ face of the helicase torus. (Note that due to space limitations this is a somewhat abbreviated and simplistic description of the complex subunit structures described in the original reference, with additional intermediate structures described for the monomers of the associated second hexamer—interested parties are directed to consult the original reference for more detail (Enemark and Joshua-Tor, 2006)). The authors characterized this right-handed helical structure of DNA binding hairpins within the central pore as a “spiral staircase”, and noted that the process of ATP-hydrolysis and the coincident conformational changes in each E1 monomer would serve to push each associated sugar-phosphate residue of the ssDNA template from the 5′ to the 3′ face of the helicase torus, producing the well-established 3′ to 5′ directionality of the E1 helicase. While the “spiral staircase” description is an apt analogy for the static picture produced by the crystal structure, it does not reflect the dynamic nature of the translocating helicase. A more apt analogy might be that of a “spiral escalator”, where each sugar-phosphate residue of the ssDNA template “steps” onto a DNA binding loop closest to the 5′ face of the torus, and is lifted to the 3′ face via the conformational changes in that E1 monomer associated with ATP hydrolysis and release. It should also be noted that due to the dynamic nature of this model, nucleotides are constantly loading onto and unloading from the “spiral escalator”; hence, each helicase hexamer is simultaneously contacting only four or five ssDNA nucleotides, even though every nucleotide

Fig. 4. Location of the β-hairpins and single-stranded DNA within the E1 helicase central channel. Crystal structure of the hexameric BPV1 E1 HD bound to ADP and single-stranded DNA (PDB accession number 2GXA) (Enemark and Joshua-Tor, 2006). The E1 hexamer is colored as in Fig. 3 but with the six DNA-binding β-hairpins (amino acids 504 to 508) highlighted in different colors. The single-stranded DNA is omitted to allow for easier visualization of the β-hairpins.
residue is assigned its own DNA binding hairpin or “step” during translocation.

This model of DNA unwinding by E1 implies that the transport of each and every nucleotide residue through the HD would require hydrolysis of one ATP (Enemark and Joshua-Tor, 2006). This energetic cost would be at the upper estimates of what has been determined for other DNA helicases and may, in reality, be lower. Indeed, a subsequent structure of the BPV1 E1 helicase showed that this same “spiral staircase” arrangement can form in absence of both nucleotide and ssDNA (Sanders et al., 2007). While Enemark and Joshua-Tor had certainly noted that E1 monomer–monomer interactions played a role in the slight conformational changes that create the “spiral staircase” (Enemark and Joshua-Tor, 2006), it was not until the publication of the E1 structure in absence of substrate that it could be appreciated that E1 monomer–monomer interactions play probably the most vital (and apparently sufficient) role in creating this spiral arrangement (Sanders et al., 2007). The implication of this finding is that one can now envision that an ATP molecule may not be directly required for the transport of each ssDNA nucleotide residue up the “spiral escalator”; conformational shifts caused by ATP-binding and hydrolysis by only a subset of E1 monomers could be sufficient to induce the necessary conformational shifts in the other monomers within the hexameric helicase. Indeed, it is possible that as few as one ATP molecule might be required for the passage of up to six nucleotide residues through the central pore. This could account for the varying step-size estimates determined for E1-related helicases (Singleton et al., 2007).

These results on the E1 helicase are particularly important in regards to the human replicative helicase complex. The major motor component of the eukaryotic replicative DNA helicase complex is a hetero-hexamer composed of the six mini-chromosome maintenance proteins (MCM 2 through 7), which are all members of the AAA+ DNA helicase superfamily (Bochman and Schwacha, 2009). While all six MCM proteins are believed to be essential components of this motor helicase, hetero-hexamers assembled with an ATPase-dead (Walker motif) version of MCM 2, 3 or 5, or with all three mutant subunits, have been shown to remain fully functional for ATPase activity; conversely, hetero-hexamers with Walker B mutations in the MCM 4, 6 and 7 subunits show dramatically decreased ATPase activity (Schwacha and Bell, 2001). As for their effect on the unwinding function of these mutant complexes, Walker A mutations in MCM 4, 5, 6 or 7 have been found to ablate DNA helicase activity, while Walker A mutations in MCM 2 or 3 retain moderate to wild-type levels of activity (Bochman and Schwacha, 2008). While these have been confusing results for the eukaryotic DNA replication field, the findings on E1 suggesting that only a subset of ATPase sites actually hydrolyzing ATP may be sufficient for translocation and helicase action, may explain why not all MCM ATPase domains are required for the MCM complex to be functional.

Small-molecule inhibitors of E1 enzymatic activity

The enzymatic activity of E1 has been considered an attractive target for antiviral drug discovery, specifically for the treatment of HPV-associated lesions in which the viral genome is not integrated but remains episomal, such as in benign anogenital warts caused by HPV6 and HPV11 infections. However, despite intense high-throughput screening efforts, only two classes of E1 inhibitors have been reported. Benzodiazepine-like inhibitors of the ATPase activity of HPV11 E1 have been claimed in the patent literature (US patent 6703,387 B2; Hurst et al., 2004)). Potent biphenylsulfonylacetamide inhibitors of the ATPase activity of HPV6 E1 have also been reported (Faucher et al., 2004), but those were inactive against HPV11 E1 due to the replacement of a single non-conserved residue within the NTP-binding site of this enzyme (tyrosine 486 in HPV6 E1 replaced by a cysteine in HPV11 E1) (Faucher et al., 2004; White et al., 2005). Despite these inhibitors being active in vitro, they were not effective in cell culture assays, thus limiting their potential as anti-HPV lead compounds. Why E1 has remained refractory to pharmacological inhibition is not completely understood but may be due, at least in part, to the poorly-discriminating nature of its nucleotide-binding site which, in the case of HPV11 E1, has been shown to accommodate and hydrolyze a variety of different NTPs and dNTPs in vitro (Rocque et al., 2000; Titolo et al., 2000; White et al., 2001).

Initiation of papillomavirus DNA replication

Although it is generally accepted that replication of the viral genome in undifferentiated cells, in transient assays or in vitro involves a bi-directional replication fork, it has been suggested that amplification of the PV genome during the productive-phase of the life cycle occurs through a rolling-circle mechanism (Flores and Lambert, 1997). Because the vast majority of published studies have investigated how the viral genome is replicated during the establishment and maintenance stages of the viral life cycle, much more is known about the bi-directional mode of DNA replication that is characteristic of these stages. In this section, we will review how bi-directional replication of the PV genome is initiated by E1 and E2 and the mechanism by which E1 assembles at the ori as a functional double-hexamer. Most in vitro studies on the assembly of E1 and E2 at the origin have been performed with the BPV1 proteins and a fragment of the BPV1 origin, designated as the minimal ori, which contains a single E2-binding site (E2BS), an E1-binding region and an AT-rich sequence (Fig. 5A). This minimal ori supports BPV1 DNA replication in transient assays, albeit at slightly lower levels that the complete ori which contains an additional E2BS (Fig. 5A) (Ustav et al., 1993). The current model of initiation of BPV1 DNA replication is schematized in Fig. 5B and described below together with relevant data obtained with the E1 and E2 proteins of other PV types.

Assembly of the E1–E2–ori ternary complex

In vitro, E1 can bind to the ori and drive PV DNA replication independently of E2 if used at high concentrations and in absence of non-specific competitor DNA (Bonne-Andrea et al., 1995a). In contrast, the modest specificity of E1 for its target binding sites prevents it from binding to the ori without the help of E2 when challenged with large amounts of non-specific DNA in vitro or in presence of excess cellular genomic DNA in vivo (Dixon et al., 2000; Sedman and Stenlund, 1995; Sedman et al., 1997; Titolo et al., 2003; Yang et al., 1991). The only known exception being the E1 protein from HPV type 1a, which can support significant levels of viral DNA replication in absence of E2, in transient assays (Gopalakrishnan and Khan, 1994). For the other PV types, in vivo recognition of the ori by E1 requires E2, the only other viral protein needed to initiate DNA replication (Ustav and Stenlund, 1991; Yang et al., 1991). The role of E2 during replication has been extensively characterized. Its primary role is to act as loading factor to recruit E1 specifically at the ori; a task that it accomplishes through its ability to bind simultaneously to E1 and to the E2-binding sites (E2BS) present in the ori (Berg and Stenlund, 1997; Blitz and Laimins, 1991; Frattini and Laimins, 1994; Gillette et al., 1994; Lusky and Fontane, 1991; Lusky et al., 1993, 1994; Mohr et al., 1990; Sanders and Stenlund, 1998, 2000, 2001; Sedman and Stenlund, 1995; Sedman et al., 1997; Seo et al.,
As such, E2 acts as a molecular tether to direct E1 to the two pairs of E1BS located in the ori, leading to the assembly of an E1–E2–ori ternary complex. As will be further discussed in the next two sections, this E1–E2–ori complex serves as a template for the assembly of an E1 double-hexamer with unwinding activity (Sanders and Stenlund, 1998). Two separate domains of E2 are involved in directing E1 to the origin. One is the C-terminal dimerization and DNA-binding domain (DBD) that is sufficient for high-affinity interaction with the E2BS in the ori. The second is the N-terminal region, better known as the transactivation domain (TAD) for its role in regulating viral gene transcription, which is the domain that directly binds to E1. Interestingly,
the E2 TAD interacts with the C-terminal HD of E1 (Berg and Stenlund, 1997; Bonne-Andréa et al., 1997; Lusky and Fontane, 1991; Masterson et al., 1998; Moscufo et al., 1999; Müller and Sapp, 1996; Sarafi and McBride, 1995; Stenlund, 2003; Sun et al., 1998; Titolo et al., 1999; Yasugi et al., 1997a), an interaction that not only helps to recruit E1 to the ori but which also inhibits its non-specific binding to DNA (Bonne-Andréa et al., 1997; Stenlund, 2003). Thus, E2 promotes the binding of E1 specifically at the ori by two different mechanisms, namely by tethering E1 to the ori and by preventing its non-specific interaction with competing cellular DNA. For BPV1, a second interaction has also been reported that involves the E1 DBD and E2 DBD (Berg and Stenlund, 1997; Chen and Stenlund, 1998; Leng et al., 1997; Moscufo et al., 1999; Woytek et al., 2001). This additional protein–protein interaction appears to be specific to BPV1 and reflects the peculiar arrangement of E1 and E2 binding sites within the ori of this virus, specifically the close juxtaposition of E2BS12 to the E1-binding region (Fig. 5A) (Berg and Stenlund, 1997). Complex formation between the DNA-binding domains of both E1 and E2 likely serves to promote the subsequent interaction of the helicase region with the TAD (Berg and Stenlund, 1997; Gillitzer et al., 2000), which, based on functional studies and on its high-degree of conservation amongst PVs, is thought to be the more “productive” interaction in the assembly of the E1–E2–ori ternary complex. In vitro, studies using a BPV1 ori fragment containing a single E2BS indicated that it can nucleate the assembly of an E1–E2–ori comprised of a single E2 dimer and two E1 molecules bound at one of the two pairs of E1BS (E1BS 2 and 4, the highest affinity pair; Fig. 5) (Chen and Stenlund, 1998).

Because of its central role in the initiation of PV DNA replication, the E1–E2 interaction has received considerable attention both from an academic standpoint and as a candidate antiviral target. A crystal structure of a complex between the E2 TAD (aa 1–215) and E1-HD (aa 428–631) of HPV18 has been reported (Fig. 6) (Abbate et al., 2004). This structure revealed that both domains interact in a 1:1 ratio; an observation confirming that the E1–E2 interaction does not require dimerization of the TAD as previously suggested by the stoichiometry of the E1–E2–ori complex (Chen and Stenlund, 1998). The E1–E2 interaction surface is of medium size (940 Å²) and involves residues located in several structural elements of both proteins (Abbate et al., 2004). The E2-binding surface on HPV18 E1 is formed by helices α2 (aa 446–457), α3 (aa 460–472) e9 (aa 613–624) and an extended loop (loop 2, aa 595–612) preceding e9. Of particular interest is an ionic interaction (salt bridge) between arginine 454 in α2 and a highly conserved glutamic acid in the E2 TAD (aa 43 in HPV18, Fig. 6) (Abbate et al., 2004). Previous mutational analyses of E2 from different PV types had identified this glutamic acid residue as important for E1-interaction and viral DNA replication. As anticipated, substitution of R454 for alanine abolished the ability of E1 to interact with E2 in vitro and to support transient DNA replication in vivo (Abbate et al., 2004). The R454A substitution did not affect the ATPase activity of E1 demonstrating the selective role that R454 plays in interaction with E2.

The E1–E2 interaction has also received attention as a candidate antiviral target. A series of small molecule inhibitors of this interaction has been reported that binds to the E2 TAD and competitively inhibits its binding to the E1 HD (White et al., 2003). Although these inhibitors bind on the same surface of the TAD as E1, they make very different contacts with the protein and, as such, can bind specifically and with high affinity only to the E2 TADs of the low-risk HPV6 and HPV11 viruses. Like mutations in E1 and E2 that disrupt their interaction, these compounds inhibit viral DNA replication in cellular assays. As such, they represent an interesting starting point for the development of antiviral agents to treat diseases caused by the low-risk HPV6 and 11 such as benign genital warts (condylomas) and recurrent respiratory papillomatosis.

Assembly of the E1 double-hexamer

Although E1 can assemble into hexamers in solution, these hexamers cannot support ori-dependent DNA replication in vitro. It is now recognized that bi-directional unwinding and replication of the viral genome requires the assembly, at the ori, of a double-hexamer (DH) of E1 in which each of the two hexamers encircles one of the two DNA strands. For the past decade or so it was believed that the two hexamers comprising each DH would remain associated during DNA replication, with the template DNA being threaded through this dodecameric structure. Only recently has that paradigm begun to shift back, based on single-molecule experiments with SV40 large T-antigen indicating that the two hexamers can separate and track along the opposite strands during SV40 DNA replication (Yardimci et al., 2012). This model emphasizes the need for the two hexamers within each DH to assemble around the two opposite template DNA strands at the ori. As such, it also provides an explanation for why E1 hexamers formed in solution cannot support ori-dependent DNA replication, namely because of their inability to gain access to and encircle one of the two DNA strands.

Much of our current understanding of the assembly of the E1 DH comes from studies on the BPV1 enzyme by Stenlund and co-workers. Key to their success was their ability to divide the highly-cooperative process of DH assembly into discrete steps in vitro, by trapping different intermediate complexes using mutant or truncated ori substrates, specific E1 mutant proteins and conditions permitting ATP-binding versus hydrolysis. The mechanism by which BPV1 E1 assembles as a DH at the ori in vitro, in absence of E2 and in a reaction that requires ATP hydrolysis, is reviewed below.

Assembly of the E1 DH proceeds via the formation of an intermediate complex comprised of two E1 trimers bound in a head-to-head configuration, known as the E1 double-trimer (DT) (Enemark et al., 2000, 2002; Sedman and Stenlund, 1996). Formation of the E1 DT requires the two pairs of E1BS in the ori that can promote the dimerization of E1 (E1BS-2 and E1BS 2–4) and likely also involves the two unpaired sites E1BS 5 and 6; the specific arrangements of these six E1BS providing a scaffold for the assembly of six E1 molecules into a DT (Chen and Stenlund, 2001; Enemark et al., 2000, 2002; Sedman and Stenlund, 1996). Assembly of the E1 DT also requires sequences on both sides of the
E1BS region including in the AT-rich region; these sequences likely being contacted by the E1 helicase domain (Enemark et al., 2000, 2002; Sedman and Stenlund, 1996). DT assembly also requires several functional domains and activities of E1, namely the ability of the DBD to bind and dimerize on both pairs of E1BS and to also interact with adjacent E1 molecules, the non-specific DNA-binding activity of the HD to contact sequences outside of the E1BS region, and ATP-binding but not hydrolysis (Chen and Stenlund, 2002; Enemark et al., 2002; Liu et al., 2007; Schuck and Stenlund, 2005a, 2006). Critical in the assembly of the BPV1 E1 DT are regions of the HD that mediate non-specific DNA binding, specifically the β-hairpin motif, the hydrophobic loop spanning the invariant F464 and a charged loop which contain K461 (Liu et al., 2007, 2010). As mentioned previously, amino acid substitutions in the DBD have also been identified that specifically interfere with the assembly of the E1 DT, without affecting the capacity of the mutant E1 protein to bind DNA (Schuck and Stenlund, 2006). An interesting property of the DT complex is that it is sufficient to melt the ori. Melting is a prerequisite for the transition of the DT into a replication-competent DH, suggesting that ssDNA is the substrate on which the DH assembles (Schuck and Stenlund, 2005a, 2007, 2011). Ori melting requires specific TA base pairs flanking the E1 BS region and has been proposed to occur as a result of untwisting of the DNA, caused in part by the movement of the β-hairpins in response to ATP-binding and hydrolysis (Schuck and Stenlund, 2005a). Thus, the DNA-binding activities of both the DBD and HD are required for DT assembly and to melt the ori. It was suggested that a third DNA-binding activity, associated with the E1 oligomerization domain and involving conserved K356, also participates in ori melting (Sanders, 2008), although it has also been suggested that mutation of this residue impairs the ability of E1 to oligomerize and interact with E2 (Liu et al., 2007, 2010). Following its assembly, the E1 DT can be converted into a replication-competent DH in a process that requires ATP-hydrolysis. Although the exact mechanistic details by which this conversion occurs remain to be fully elucidated, it almost certainly requires, beside ori melting, the assembly of six additional E1 molecules and the rearrangement of the two newly formed head-to-head hexamers into ring-like structures that each encircle one of the two DNA strands; these two events being required to generate a DH complex with non-specific DNA helicase activity. Surface mutagenesis of the BPV1 E1 helicase domain has identified several amino acid substitutions that specifically impair the transition of the E1 DT to E1 DH (I423A, N436A, N444A, S456A, N459A, T490A, N494A, S537A) (Schuck and Stenlund, 2011). Interestingly, these substitutions all affect residues of the AAA+ module that are located at the interface between monomers in the BPV1 E1 hexameric structure (Schuck and Stenlund, 2011).

The findings described above have highlighted the key roles that the DBD and HD play during the assembly of BPV1 E1 at the ori. Studies on HPV11 E1 have confirmed the importance of these two domains in ori-binding but also hinted that the E1 N-terminal region may be involved. Specifically, it was found that deletion of the first 71 amino acids of HPV11 E1 increases its binding to the ori approximately 5-fold in vitro (Titolo et al., 2000). How the N-terminal region affects the binding and oligomerization of E1 at the ori remains to be determined.

ATP as an allosteric inhibitor of the E1–E2 interaction

The studies presented above revealed how the E1–E2–ori ternary complex is assembled at the ori and how E1, in absence of E2, can assemble into a DH in vitro. EMSA experiments performed with BPV1 E1 and E2 indicated that these two types of E1 complexes do not form independently at the ori but, rather, that assembly of the E1–E2–ori ternary precedes formation of the E1 DH and is in fact a prerequisite for DH assembly (Sanders and Stenlund, 1998). Furthermore, they revealed that E2 is not an integral part of the final E1 DH complex, suggesting that the interaction between E1 and E2 is disrupted, and E2 displaced, as part of the oligomerization process leading to DH formation (Lusky et al., 1994; Sanders and Stenlund, 1998). The notion that oligomerization of E1 is incompatible with E2-binding was further demonstrated in gel filtration studies indicating that HPV18 E2 can only bind to monomeric E1 and not to a hexamer (Abbate et al., 2004). Modeling studies and examination of the E1–E2 and E1 hexamers crystal structures suggest that this specificity is dictated in part by the loop 2 region of E1, whose interactions with the E2 TAD or with another E1 molecule during the oligomerization process are mutually exclusive (compare the position of loop 2 in Figs. 3 and 6) (Abbate et al., 2004).

Given that ATP promotes the oligomerization of E1, which in turn disrupts its interaction with E2, one can think of ATP as an allosteric inhibitor of the E1–E2 interaction. The converse phenomenon, namely that E2 can act as an inhibitor the ATPase activity of E1 (which depends on its oligomerization), has also been verified experimentally. Specifically, it was observed for HPV11 E1 and E2 that the TAD could increase the $K_m$ for ATP by approximately 7-fold (White et al., 2001). Thus, ATP and the E2 TAD compete with each other for binding to E1, although they interact on different surfaces of the helicase domain. These findings provide a mechanistic basis to explain how ATP-binding could favor the transition of the E1–E2–ori ternary complex into the E1 DT/DH at the ori. This transition is likely further facilitated by ATP-hydrolysis and the interaction of the E1 HD with DNA, in particular with the ssDNA that is generated as part of the ori melting process and upon which a replication-competent DH is assembled, as discussed above. From a practical standpoint, the notion that ATP can regulate different steps in the initiation of viral DNA replication suggests that inhibitors of the ATPase function of E1 could affect not only the unwinding activity of the enzyme but also its assembly at the origin.

Interaction of E1 with the cellular DNA replication machinery

Small DNA tumor viruses, including PVs, encode only a few proteins and must therefore rely extensively on their host for the other functions required to complete their life cycles. Consequently, most papillomavirus proteins engage in multiple protein interactions with cellular factors to carry out their activities. E1 is no exception and has been reported to interact with several members of the cellular DNA replication machinery (Table 1). These include the DNA polymerase α-primase (Pol α-prim) complex, replication protein A (RPA), and topoisomerase I (Topo I). In vitro, papillomavirus DNA replication also requires replication factor C (RFC), proliferating-cell nuclear antigen (PCNA) and DNA polymerase δ, even though interaction of these factors with E1 has not been reported (Kuo et al., 1994; Melendy et al., 1995; Müller et al., 1994). The process of viral DNA synthesis is schematized in Fig. 7 and the various interactions that E1 makes with key replication factors are reviewed below.

DNA polymerase α-primase complex (Pol α-prim)

Pol α-prim is composed of four subunits of 180 (p180), 70 (p70), 58 (p58), and 48 (p48) kDa, respectively (Wang, 1991). P180 is the polymerase catalytic subunit, whereas p48 and p58 are the primase catalytic and auxiliary subunits, respectively. P70 has no known catalytic function. Pol α-prim is essential for initiation of DNA synthesis and for the repeated re-initiation events required for lagging-strand DNA synthesis of the host genome (Wang, 1991). E1 has been shown to interact with Pol α-prim, but when it
<table>
<thead>
<tr>
<th>E1-associated protein</th>
<th>Cellular function of associated protein</th>
<th>Interaction region on associated protein</th>
<th>Interaction region on E1</th>
<th>Function of interaction</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pol α-prim DNA replication</td>
<td>p70 and/or p180 subunits</td>
<td>Helicase domain interacts with p70 (aa 353–649 of HPV11 E1 and aa 337–583 of HPV16 E1)</td>
<td>Replication of viral DNA, direct role in viral DNA synthesis</td>
<td>Park et al. (1994), Bonne-Andrea et al. (1995a), Masterson et al. (1998), Conger et al. (1999), Amin et al. (2000)</td>
<td></td>
</tr>
<tr>
<td>RPA Single-stranded DNA-binding protein</td>
<td>Major ssDNA-binding domain of 70-kDa subunit (aa 181–291)</td>
<td>nd</td>
<td>Replication of viral DNA, direct role in DNA synthesis, may help load RPA onto nascent ssDNA</td>
<td>Han et al. (1999), Loo and Melendy (2004)</td>
<td></td>
</tr>
<tr>
<td>Topo I DNA relaxation</td>
<td>nd</td>
<td>DBD and helicase domain of BPV1 E1 independently bind Topo I</td>
<td>Increases Topo I activity and E1 binding to ori</td>
<td>Clower et al. (2006), Hu et al. (2006)</td>
<td></td>
</tr>
<tr>
<td>p80/Uaf1 WD-repeats protein associated with deubiquitinating enzymes</td>
<td>WD-repeats (aa 1–573)</td>
<td>N-terminal region of E1 from anogenital HPV types (aa 10–40 of HPV11 and HPV31 E1)</td>
<td>Replication of viral DNA and episomal maintenance</td>
<td>Côté-Martin et al. (2008), Lehoux et al. (2012)</td>
<td></td>
</tr>
<tr>
<td>Histone H1 Chromatin component</td>
<td>nd</td>
<td>N-terminal region (aa 1–185 of HPV11 E1)</td>
<td>E1 can displace histone H1 from DNA, this may alleviate the repressive effect of H1 on viral DNA replication</td>
<td>Swindle and Engler (1998)</td>
<td></td>
</tr>
<tr>
<td>Ini1/hSNF5 Subunit of SWI/SNF chromatin remodeling complex</td>
<td>nd</td>
<td>aa 147–444 of HPV18 E1</td>
<td>Replication of viral DNA, may be involved in chromatin remodeling during viral DNA synthesis</td>
<td>Lee et al. (1999)</td>
<td></td>
</tr>
<tr>
<td>Hsp40/70 Molecular chaperones</td>
<td>J-domain of Hsp40; nd for Hsp70</td>
<td>nd</td>
<td>Promotes binding of E1 to ori and assembly into active double hexamers</td>
<td>Liu et al. (1998), Lin et al., (2002)</td>
<td></td>
</tr>
<tr>
<td>E1-BP/TRIP13 Putative ATPase</td>
<td>nd</td>
<td>nd</td>
<td>Replication of viral DNA, mechanism unknown</td>
<td>Yasugi et al. (1997b)</td>
<td></td>
</tr>
<tr>
<td>p56 Interferon-inducible protein</td>
<td>N-terminal region (aa 29–152)</td>
<td>Helicase domain, aa F399 of HPV18 E1 essential for interaction</td>
<td>Inhibits helicase activity, interferes with E1 binding to E2 and to ori, sequesters E1 in cytoplasm</td>
<td>Terezeni et al. (2008), Saikia et al. (2010)</td>
<td></td>
</tr>
<tr>
<td>Caspases-3/-7 Mediators of apoptosis</td>
<td>nd</td>
<td>Caspase-cleavage motifs (D-x-x-D) in N-terminal region</td>
<td>Viral genome amplification in differentiated cells</td>
<td>Moody et al. (2007)</td>
<td></td>
</tr>
<tr>
<td>Cyclin A/E-Cdk2 Cyclin-dependent kinases that regulate Caspase-cleavage motifs (D-x-x-D) in N-terminal region</td>
<td>Cyclin A and cyclin E</td>
<td>Caspase-cleavage motifs (D-x-x-D) in N-terminal region</td>
<td>Phosphorylation of HPV11 and HPV31 E1 inhibits their nuclear export, phosphorylation of BPV1 E1 promotes its shuttling</td>
<td>Cueillete et al. (1998), Ma et al. (1999), Deng et al. (2004), Hsu et al. (2007), Fradet-Turcotte et al. (2010a)</td>
<td></td>
</tr>
<tr>
<td>ERK1 and JNK2 MAP kinases (MAPK)</td>
<td>nd</td>
<td>MAPK docking motifs in helicase domain (aa 505–520 and 541–545 of HPV11 E1)</td>
<td>Promotes nuclear import of E1</td>
<td>Yu et al. (2007)</td>
<td></td>
</tr>
<tr>
<td>Importins α3/α4/α5 Nuclear import of proteins</td>
<td>nd</td>
<td>NLS</td>
<td>Mediates nuclear import of E1</td>
<td>Bian et al. (2007)</td>
<td></td>
</tr>
<tr>
<td>Crm-1 exportin Nuclear import of proteins</td>
<td>nd</td>
<td>NES</td>
<td>Mediates nuclear export of E1</td>
<td>Deng et al. (2004), Fradet-Turcotte et al. (2010a), Hsu et al. (2007)</td>
<td></td>
</tr>
</tbody>
</table>

nd, Not determined.
Fig. 7. Proposed model for the papillomavirus DNA replication fork. The E1 protein (in purple) is shown oriented with its N-terminal DNA-binding domain (DBD) facing towards the unwound dsDNA. For simplicity, only one of the two hexamers is shown. The E1 helicase encircles the ssDNA template for leading strand DNA synthesis, pumping the ssDNA template through the E1 complex from the E1-DBD side towards the E1 HD side. Interaction of topoisomerase I (Topo I) (in pink) with the DBD assists both in E1 origin binding/specificity and targets Topo I to the incoming dsDNA, where its action is necessary for progression of the replication fork. Interaction of replication protein A (RPA) (in blue) with the DBD is involved in loading RPA onto the lagging strand ssDNA template newly unwound by the E1 helicase action. The interaction of the HD with polymerase α-primase (Pol α-prim) (in yellow) may stimulate primer synthesis. As each short RNA-DNA primer is synthesized by Pol α-primase, replication factor C (RFC) (in orange), in coordination with RPA, prevents Pol α-prim re-association and loads PCNA (in green) and DNA polymerase δ (in green) onto the recessed 3' DNA end, assembling a processive DNA polymerase complex. The various interactions of E1 with Topo I, RPA and Pol α-prim, as well as the interactions between the cellular factors themselves, are coordinated physically and temporally in a highly-organized sequential manner necessary for replication fork assembly and function.

Replication protein A (RPA)

Human RPA is a single-stranded DNA-binding protein involved in many aspects of DNA metabolism, including DNA replication, repair, and recombination (Wold, 1997). It is a highly structured, heterotrimeric complex consisting of three subunits: a 70-kDa subunit (RPA70), a 32-kDa subunit (RPA32), and a 14-kDa subunit (RPA14), containing multiple ssDNA binding motifs. E1 has been shown to interact with the 70-kDa subunit of RPA, and specifically within the major ssDNA-binding domain of this subunit (Han et al., 1999; Loo and Melendy, 2004). Mechanistic studies performed in vitro suggested how E1 may facilitate the loading of RPA onto ssDNA at the replication fork. The binding of RPA to ssDNA and E1 appear to be mutually exclusive; when RPA is bound to ssDNA it can no longer bind to E1. Also RPA preferentially binds to ssDNA molecules that have been pre-bound to E1; when RPA is introduced to ssDNA molecules where some of them are bound to E1 and some of them are free, the creation of the RPA-ssDNA complex occurs at the expense of the E1-ssDNA population, not at the expense of the free ssDNA population. Interestingly, no ternary E1-RPA-ssDNA complex has been detected. Together these findings led to the first described model for active “loading” of RPA onto ssDNA being expelled from a DNA helicase, and may describe a general process for DNA helicases and their cognate ssDNA binding proteins (Loo and Melendy, 2004). Indeed, very similar findings and a nearly identical model were subsequently proposed for the SV40 LT-Ag helicase (Jiang et al., 2006).
Topoisomerase 1 (Topo I)

Topo I is an ATP-independent protein that modifies the topology of supercoiled DNA, acting as a swivel to decrease torsional stress on DNA. Topo I has been shown to interact with both the DBD and the C-terminal domain of BPV1 E1 in vitro (Clower et al., 2006). These interactions increase the activity of Topo I (Clower et al., 2006; Hu et al., 2006). Once again similarities are seen between E1 and SV40 LT-Ag. LT-Ag also interacts with Topo I (Simmons et al., 1996) and this interaction has been shown to play several roles in SV40 DNA replication, including stimulating the binding of LT-Ag to the origin (Gai et al., 2000; Simmons et al., 2004; Simmons et al., 1998; Trowbridge et al., 1999). Similarly, Topo I has been shown to increase the binding of BPV1 E1 to the ori and its assembly into the oligomeric helicase in vitro (Clower et al., 2006). The interaction between Topo I and LT-Ag has been shown to be important for SV40 as LT-Ag mutants compromised specifically in binding Topo I are defective for SV40 DNA replication both in vitro and during virus infection (Khopde et al., 2008; Khopde and Simmons, 2008). It is currently unknown whether the E1-Topo I interaction will play as important a role for DNA replication.

Interaction of E1 with other cellular proteins

In addition to the known DNA replication factors mentioned above, E1 has also been reported to interact with other cellular proteins to assist and/or regulate viral DNA replication (Table 1). Among these are p80, HsfN5, histone H1, E1BP and p56, described below.

p80/Uaf1

Tandem affinity purification of HPV11 E1 revealed that it interacts with the cellular WD-repeat-containing protein p80 (also known as WDR48 or USP1-associated factor 1 (UAF1)) (Côté-Martin et al., 2008). The p80-interaction domain on E1 was mapped to amino acids 10–40 of the N-terminal regulatory region. The E1-p80 interaction was found to be conserved for HPV6, 11, 16, 18 and 31 but not for HPV1, BPV1 or CRPV, suggesting that it is specific to anogenital types (Côté-Martin et al., 2008). Microscopy studies revealed that p80 is re-localized from the cytoplasm to discrete nuclear foci induced by E1 and E2 (Côté-Martin et al., 2008; Lehoux et al., 2012). Chromatin immunoprecipitation assays further revealed that p80 is recruited to the viral origin (Lehoux et al., 2012). Double amino acid substitutions in HPV31 E1 that abrogate p80-binding (W17A/F18A, V20A/E21A, and V23A/I24A) were found to reduce transient viral DNA replication by 70%, without affecting the interaction of E1 with E2 or its assembly at the origin in vivo (Lehoux et al., 2012). In the context of the complete viral genome, these same substitutions were found to inhibit the maintenance of the viral episome in immortalized keratinocytes, suggesting a key role for the E1-p80 interaction in this process (Côté-Martin et al., 2008). Interestingly, overexpression in trans of a 40 amino acid long p80-binding peptide, derived from HPV31 E1, was found to inhibit transient viral DNA replication by preventing the recruitment of endogenous p80 to the origin (Lehoux et al., 2012). Characterization of this peptide by NMR showed that the p80-binding domain of E1 is intrinsically disordered in solution (Lehoux et al., 2012). Collectively, these studies highlighted the importance of the E1-p80 interaction for the replication and maintenance of the viral episome. While the cellular function of p80 remains to be clearly established, recent studies have shown that it can associate with de-ubiquitinating enzymes (Cohn et al., 2009; Cohn et al., 2007; Kee et al., 2010). Whether these enzymes play a role in viral genome replication and maintenance remains to be determined.

Histone H1 and Ini1/hSNF5

The PV genome is associated with nucleosomes in both infected cells and in virions. Given that viral DNA replication takes place in the context of chromatin, it is not surprising that E1 was found to interact with proteins that can modify chromatin structure. In one study, the N-terminal 185 amino acids of HPV11 E1 were found to interact with histone H1, whose primary function is to bind nucleosome-organized chromatin to create an ordered, more compact DNA structure (Swindle and Engler, 1998). Interestingly, E1 could displace H1 bound to DNA in vivo. This result led the authors to hypothesize that E1 may interact with histone H1 as a means to relieve the repressive effect that chromatin exerts on viral DNA replication in vivo.

In another study, HPV18 E1 was found by yeast two-hybrid screening to interact with Ini1/hSNF5, a subunit of the SWI/SNF chromatin remodeling complex (Lee et al., 1999). The interaction was also observed with HPV11 E1 and, to a lesser extent, with BPV1 E1. The minimal fragment of HPV18 E1 that could interact with Ini1/hSNF5 was comprised of amino acids 147–444, which spans the DBD and minimal oligomerization domain. HPV18 DNA replication was found to be stimulated by overexpression of Ini1/hSNF5 and, conversely, to be inhibited by an Ini1/hSNF5 antisense RNA. Furthermore, amino acid substitutions in E1 (S225P/F226S and L305K in the DBD, N395H/A396E and H418E/Y419H in and near conserved region A, respectively) were identified that reduced binding to Ini1/hSNF5 and abrogated transient HPV18 DNA replication. These functional results led the authors to conclude that the E1-Ini1/hSNF5 interaction is required for HPV DNA replication (Lee et al., 1999). As a cautionary note, it needs to be pointed out that most of the amino acid substitutions described in this study affect highly conserved residues of E1 and, hence, are likely to also impair other activities of the protein, such as ori-binding, oligomerization and ATP-hydrolysis, required for viral DNA replication.

Hsp40 and Hsp70

The two molecular chaperones, Hsp40 and Hsp70, have been found to promote the binding of HPV11 E1 to the ori in vitro (Liu et al., 1998). These chaperones enhance the binding of E1 to DNA, independently and by different mechanisms, such that their effects are additive. In a process requiring ATP-hydrolysis, Hsp70 was found to interact transiently with E1 and enhance its binding to DNA as a hexamer. On the other hand, Hsp40 promoted the binding of E1 to the ori as a double-hexamer (Liu et al., 1998). In a cell-free DNA replication assay based on recombinant HPV11 E1 and E2, Hsp40 and Hsp70 stimulated DNA synthesis by approximately 2-fold (Liu et al., 1998). A subsequent study revealed that these chaperones are also able to dissociate E2 from E1-ori complexes and to reverse the inhibitory effect that E2 imparts on the helicase activity of E1 (Lin et al., 2002). Thus, Hsp40 and Hsp70 stimulate viral DNA replication by at least two mechanisms, first by helping to dissociate E1 from E2 at the ori and, second, by facilitating the assembly of a replication-competent double-hexamer (Lin et al., 2002). It was also suggested that chaperone proteins may act to recycle E1 double-hexamers, by dissociating them into monomers capable of re-initiating DNA replication (Liu et al., 1998). Because these studies were all performed in vitro, it remains to be determined if the proposed mechanisms of action of Hsp40 and Hsp70 also apply in vivo.
The putative ATPase E1-BP (also known as thyroid hormone receptor interactor 13, TRIP13) was identified as an HPV16 E1-interacting protein from a yeast two-hybrid screen (Yasugi et al., 1997b). Amino acid substitutions in HPV16 E1 that prevent its association with E1-BP were identified (Y412F, W439R, W439L, G482D, G496R) and found to impair viral DNA replication, consistent with a role for E1-BP in this process (Yasugi et al., 1997b). Unfortunately, and as discovered by the authors, the interpretation of these findings was complicated by the fact that many of these amino acid substitutions also impaired other activities of E1, such as its ATPase activity, oligomerization, and interaction with E2 and/or Ubc9. Thus, further experiments will be required to determine the exact role of E1-BP in viral DNA replication. This study was amongst the first to highlight the extreme sensitivity of the E1 HD to structural perturbations, and specifically, that most amino acid substitutions in this domain are pleiotropic resulting in inhibition of E1 oligomerization, ATPase activity and interaction with E2 or with cellular proteins such as Ubc9 and In1/hSNF5. From a structural point of view, these pleiotropic effects can be explained by the interdependence of the minimal oligomerization domain, ATP-binding site and C-terminal brace for proper assembly and functioning of the HD.

Another yeast two-hybrid screen, this time using HPV18 E1 as the bait, identified p56, a cellular protein induced by interferon (IFN), as an E1-interacting protein. Interaction with p56 was also detected for HPV11 and HPV31 E1. Interestingly, p56 was found to inhibit HPV DNA replication by interacting directly with E1 and sequestering it in the cytoplasm (Terenzi et al., 2008). In addition, p56 was found to inhibit the helicase activity of E1 and to also interfere with its binding to E2 and to the viral origin in vitro (Saikia et al., 2010). The amino acid phenylalanine 399 (F399) in HPV18 E1 was identified as critical for p56 interaction. Deletion of F399 did not impair viral DNA replication but rendered the process insensitive to inhibition by IFN (Saikia et al., 2010). In contrast, other studies showed that alanine substitution of the analogous residue in HPV11 E1 (F393) and HPV31 E1 (F372) completely abrogated transient DNA replication even in absence of IFN treatment (Fradet-Turcotte et al., 2010b; Titolo et al., 2000). The reason for this discrepancy is unknown. F399 is highly conserved among E1 proteins of different PV types, with either F, W, and rarely Y being found at this position. From a structural point of view, these pleiotropic effects can be explained by the interdependence of the minimal oligomerization domain, ATP-binding site and C-terminal brace for proper assembly and functioning of the HD.

**E1 intracellular localization**

E1 accumulates primarily in the nucleus when expressed in transfected cells, in agreement with PV DNA replication taking place in this cellular compartment. Although E1 is located primarily in the nucleus at the steady-state, studies on HPV11, HPV31 and BPV1 E1 have shown that these proteins actually shuttle between the nucleus and cytoplasm, using a combination of nuclear localization and export signals. Importantly, this shuttling process appears to be highly regulated by phosphorylation, likely as a means to control the overall levels of viral DNA replication and ensure its appropriate timing during S-phase.

**E1 nuclear import**

Studies aimed at identifying the localization signal (NLS) of BPV1 E1 revealed that it lies in the N-terminal regulatory region of the

![Bipartite NLS](image)

**Fig. 8. Nucleo-cytoplasmic shuttling motifs in the N-terminal regulatory region of E1.** Sequence alignment of the E1 shuttling module from different PV types, as indicated. The position of the first amino acid in each sequence is indicated (number in parentheses). The location of the bipartite nuclear localization signal (NLS) is indicated by a double-arrow and highlighted in yellow. The nuclear export signal (NES) and cyclin-binding motif (CBM) are shaded in light blue and purple, respectively. Putative and known Cdk-phosphorylation sites, [S/T]-P, are colored in pink. Note the absence of some of these nucleo-cytoplasmic shuttling motifs in several E1 proteins from high-risk HPV types and from BPV1.
protein and is of the bipartite type, comprised of two regions of negatively charged amino acids separated by approximately 30 residues (Leng et al., 1994; Lentz et al., 1993; Yu et al., 2007). Later studies on HPV11 and HPV31 E1 confirmed these findings (Deng et al., 2004; Fradet-Turcotte et al., 2010a; Lentz et al., 1993; Yu et al., 2007). The N-terminal basic region of the NLS is the most highly conserved, with the sequence KRK being present in 95% of E1 types (Fig. 8). Substitution of the first two amino acids of this sequence by glycine, or of all three residues by alanine, was found to abrogate the nuclear accumulation of E1 from BPV1, HPV11 and HPV31 (Fradet-Turcotte et al., 2010a; Lentz et al., 1993; Yu et al., 2007).

Studies on BPV1 E1 revealed that the protein can bind to importins α3, α4 or α5 in vitro and that these importins can mediate nuclear import of E1 in digitonin-permeabilized cells (Bian et al., 2007; Castella et al., 2006b). Interestingly, the BPV1 E1 NLS is surrounded by several putative phosphorylation sites including T102 and S109 which flank the C-terminal basic region of the NLS (TPVKKKRS, aa 102–109). T102 can be phosphorylated by Cdk1 in vitro (Lentz et al., 1993). S109 can be phosphorylated by PKA and PKC in vitro (Zanardi et al., 1997) and is also one of several sites that are phosphorylated in recombinant BPV1 E1 purified from insect cells (Lentz et al., 2006). Results using mutant E1 proteins in which T102 and S109 were changed to either alanine to prevent phosphorylation or to aspartate as phosphomimetics, suggested that these residues are not required for E1 nuclear import but, rather, that phosphorylation of S109 may negatively regulate this process by disrupting the interaction of the NLS with its cognate importins (Bian et al., 2007; Castella et al., 2006b). Interestingly, a previous study had demonstrated that substitution of S109 by alanine (S109A) results in a mutant BPV1 genome with increased replication capacity, while substitution for glutamic acid decreases replication (Zanardi et al., 1997). This same study also demonstrated that the S109A E1, purified from insect cells, has a higher affinity for the ori in vitro than the wild type protein. Together, these findings suggest that phosphorylation of S109, presumably by either PKA or PKC, negatively regulates the import of E1 into the nucleus and its affinity for the viral origin, resulting in lower replication of the viral genome.

For HPV11 E1, nuclear import of the protein was found to be stimulated by phosphorylation. Indeed, HPV11 E1 was shown to interact with and to be phosphorylated by extracellular signal-regulated kinase 1 (ERK1) and c-Jun N-terminal protein kinase (JNK2), two mitogen-activated protein kinase (MAPK) family members (Yu et al., 2007). Two functional MAPK docking motifs were identified in the HD of HPV11 E1 (between aa 505–520 and aa 541–545, respectively), which are both conserved among many HPV E1 proteins and the BPV1 enzyme. Amino acid substitutions in either or both of these motifs were found to severely compromise the nuclear import of HPV11 E1 (Yu et al., 2007). ERK1 and JNK2 are proline-directed kinases and, as such, can phosphorylate substrates at similar sites as Cdk1 and Cdk2 (consensus sequence [S/T]-P). A double amino acid substitution of S89 and S93 by alanine was found to reduce nuclear accumulation of HPV11 E1 to the same extent as mutation inactivation of the MAPK docking motifs (Yu et al., 2007). In contrast, phosphomimetic substitutions had less of a deleterious effect. These results led to the suggestion that phosphorylation of S89 and S93 by MAPK is the key event that stimulates the nuclear import of HPV11 E1 (Yu et al., 2007). As will be discussed below, S89 and S93 can also be phosphorylated by Cdk2 to prevent the nuclear export of E1. Thus, both MAPK and Cdk2 act on S89 and S93 to promote the nuclear accumulation of E1, but by opposite mechanisms.

Finally, it is worth keeping in mind that some E1 proteins may be differentially affected by MAPK phosphorylation. For example, it was found that a mutant BPV1 E1 in which all three [S/T]-P sites were changed to alanine could still accumulate in the nucleus of transfected cells to the same extent as its wild type counterpart (Hsu et al., 2007), a result indicating that phosphorylation of the protein neither is required nor stimulates its nuclear import. It would be of interest to determine if a similar result would be obtained by mutation of the MAPK docking motifs in BPV1 E1 and, more generally, if BPV1 is a substrate of MAPK. Furthermore, as the nuclear–cytoplasmic shuttling of BPV1 E1 is promoted by Cdk2-phosphorylation of the protein (discussed below), it would be important to determine if this process is also regulated by MAPK.

### Regulation of E1 nuclear–cytoplasmic shuttling by Cdk2-phosphorylation

The HPV11 and HPV31 E1 proteins were found to contain a leucine-rich nuclear export signal (NES) located within their N-terminal region (Deng et al., 2004; Fradet-Turcotte et al., 2010a). This NES is conserved in most, but not all, E1 proteins (Fig. 8). It is typically located between the two basic regions of the NLS and is of the bipartite type, comprised of two regions of the NES (TPVKRRKS, aa 102–109). Two functional MAPK docking motifs were identified in HPV11 and HPV31 E1 (L110A/I113A in HPV11 E1 and the analogous double substitution in HPV31 E1), or treatment of the cells with the Crm1 inhibitor leptomycin B (LMB), were shown to inhibit nuclear export of these proteins (Deng et al., 2004; Fradet-Turcotte et al., 2010a).

Interestingly, the Crm1–dependent nuclear export of HPV11 and HPV31 E1 was found to be inhibited by Cdk2 phosphorylation. Indeed, in addition to the NLS and NES, the N-terminal regulatory region of many HPV E1 proteins also contains a cyclin-binding motif (CBM) located next to the NLS/NES region as well as several putative Cdk2 phosphorylation sites (consensus [S/T]-P) (Fig. 8). For HPV11 and HPV31 E1, it has been demonstrated that the CBM (consensus R-x-L, aa 124–126 in HPV11 E1) can bind to cyclin A/E-Cdk2 and that a double alanine substitution within this interaction motif (RRL-x-L, aa 124–126 in HPV11 E1) almost completely inhibits its nuclear export. Thus, phosphorylation of HPV E1 by cyclin A/E-Cdk2 ensures that E1 is retained in the nucleus at the onset and during S-phase, and may serve as a mechanism to synchronize viral DNA replication with host DNA synthesis.

The findings summarized above are likely to apply to the E1 proteins of many PV types but perhaps not to all of them given that some lack an easily recognizable NES or CBM sequence. For example, the E1 protein of high-risk HPV33 lacks both the NES and CBM (Fig. 8). Other E1 proteins, like that of BPV1 E1, contain only a CBM but no obvious NES (Fig. 8). Despite lacking a NES in its N-terminal region, BPV1 E1 was shown to shuttle between the nucleus and cytoplasm in a Crm1-dependent fashion (Hsu et al.,
2007; Rosas-Acosta and Wilson, 2008). This prompted the search for a NES in other regions of the protein. A Crm1-dependent NES was identified in the HD of BPV1 E1, between aa 408–418 of the protein (Rosas-Acosta and Wilson, 2008). Substitution for alanine was identified for a NES in other regions of the protein. A Crm1-dependent NES during S-phase, perhaps as a mechanism to down-regulate viral S283 activates the nucleo-cytoplasmic shuttling of BPV1 E1.

Another interpretation, however, could be that the I413A/L416A substitution also affects a function of E1 required in vivo, but dispensable in vitro, such as its interaction with E2 for example. Along this line, we note that I413 and L416 lie in a classical heterokaryon assay, and to accumulate almost exclusively in the nucleus of the steady-state (Hsu et al., 2007). BPV1 E1 was found to shuttle between the nucleus and cytoplasm using this NES (I413A/L416A substitution on the E1–E2 interaction in order to confirm the conclusion that nuclear export of E1 is required for its DNA replication activity in vivo.

The nucleo-cytoplasmic shuttling of BPV1 E1 was also shown to be regulated by Cdk2-phosphorylation of the protein. However, and in stark contrast to HPV11 and HPV31 E1, phosphorylation of BPV1 E1 by cyclin A-Cdk2 was found to promote, rather than inhibit, its shuttling (Hsu et al., 2007). BPV1 E1 was found to shuttle between the nucleus and cytoplasm using a classical heterokaryon assay, and to accumulate almost exclusively in the nucleus of the steady-state (Hsu et al., 2007). BPV1 E1 interacts with cyclin A/E-Cdk2 and is a substrate of both kinase complexes in vitro. The protein contains three putative Cdk2 phosphorylation sites, T102 and T126 in the N-terminal region and S283 in the DBD. Through the use of mutant proteins in which one or all three sites were changed to alanine, it was found that Cdk2-phosphorylation of BPV1 E1 is dispensable for its DNA replication activity but, instead, is required to activate its shuttling during S-phase. The S-phase kinase cyclin A-Cdk2 was found to phosphorylate E1 primarily on S283. Substitution of this residue by alanine resulted in a mutant E1 that was unable to shuttle between nuclei in a heterokaryon assay, unlike what is observed for the wild type protein. Conversely, substitution of S283 by glutamic acid as a phosphomimetic greatly stimulates shuttling of the protein and abrogates its ability to support viral DNA replication in vivo. These results led Hsu et al. (Hsu et al., 2007) to conclude that Cdk2-phosphorylation of S283 activates the nucleo-cytoplasmic shuttling of BPV1 E1 during S-phase, perhaps as a mechanism to down-regulate viral DNA replication and prevent amplification of the viral genome in non-differentiated cells.

**Nuclear accumulation of E1 arrests cells in S-phase and triggers a DNA damage response**

Why is the nuclear accumulation of E1 so tightly regulated? Part of the answer came from the finding that expression of E1 in transfected cells blocks cellular proliferation by causing cell-cycle arrest in S-phase (Fraden-Turcotte et al., 2011; Sakakibara et al., 2011). This anti-proliferative effect was observed with the E1 proteins from many different PVs (BPV1, HPV8, HPV11, HPV16 and HPV31) and even under conditions of very low E1 expression, suggesting that it is a general phenomenon and not an artifact of protein overproduction. Furthermore, this effect depends on the integrity of the NLS, DBD and ATPase domain of E1, indicating that cell-cycle arrest is caused by nuclear accumulation of enzymatically active E1. As might be expected when a rogue DNA helicase gains access to the host genome, the presence of E1 in the nucleus leads to the induction of DNA double-strand breaks and of a concomitant cellular DNA damage response (DDR), likely as a result of E1 trying to initiate DNA replication on cellular DNA (Fraden-Turcotte et al., 2010a, 2011). The E1-induced DDR is characterized by a robust activation of the ataxia-telangiectasia mutated (ATM) kinase signaling pathway (Fraden-Turcotte et al., 2010a, 2011) and, to a lesser extent, of the ATR (ATM and Rad3-related) pathway (Reinson et al., 2013). Complex formation with E2 attenuates the ability of E1 to induce a DDR but does not completely eliminate it, such that DDR markers are still observed in E1- and E2-induced nuclear foci. Remarkably, transient HPV DNA replication remains unaffected by the induction of this DDR, unlike host DNA synthesis which is shut-down by a checkpoint response (Fraden-Turcotte et al., 2011; King et al., 2010). Overall, these studies highlighted the importance of regulating the nuclear accumulation of E1 in order to achieve the optimal balance between promoting viral DNA replication without interfering with cellular proliferation. A particularly telling example is the observation that inactivating the E1 NES in the context of the HPV31 genome results in a mutant episome that induces DNA damage and is poorly maintained in undifferentiated keratinocytes (Fraden-Turcotte et al., 2010a).

Although it appears that the levels of E1 must be kept low in the nucleus of undifferentiated cells, this may not be the case in differentiated cells undergoing genome amplification. It was previously demonstrated that amplification of the HPV31 episome is dependent on the induction of a DDR within differentiated keratinocytes (Moody and Laimins, 2009). While this study clearly showed that E7 is sufficient for induction of a DDR within these cells, it remains entirely possible that nuclear accumulation of E1 also contributes to the magnitude of this response. More generally, the observations that PVs induce and require a cellular DDR to amplify their viral genome, and that mediators of the DDR and components of the DNA repair and recombination machineries are recruited to viral replication centers, have been both exciting and perplexing (Gillespie et al., 2012). One possibility to be investigated is whether these recruited repair and recombination factors might play a role in resolving the HPV concatamers produced during rolling-circle DNA replication into circular monomeric genomes suitable for viral packaging (Flores and Lambert, 1997). Future studies will undoubtedly address the molecular function of these host factors in viral DNA replication and amplification and how they may impinge on the different activities of E1.

Finally, the capacity of E1 to induce DNA damages raises the possibility that it may participate in the early stages of cancer induced by high-risk HPVs. Indeed, it is easy to imagine how the failure to control the nuclear levels of E1 would result in genomic instability and thus facilitate integration of the viral genome. As elegantly demonstrated by Kadaja et al., E1- and E2-dependent onion-skin replication of the integrated genome is an additional source of genomic instability that could favor cancer progression (Kadaja et al., 2009). Thus, while we often think of cancer progression having been solely driven by overexpression of E6 and E7, undoubtedly the two main drivers of carcinogenesis, we also need to keep in mind the potential contributions of E1 and E2 during the early stages of disease development.

**Regulation of E1 by other post-translational modifications**

**Phosphorylation**

In addition to the phosphorylation events described above which regulate the nucleo-cytoplasmic shuttling of E1, the protein can also be a substrate for other cellular kinases. For example,
BPV1 E1 can be phosphorylated by Casein kinase 2 (CK2) in vitro on amino acids S48 and S548 (Lentz et al., 2002; Lentz, 2002; McShan and Wilson, 1997). These sites are also phosphorylated when BPV1 E1 is expressed in insect cells using a baculovirus system (Lentz et al., 2006). Both sites conform to the CK2 consensus sequence (S-x-x-[E,D]). Substitution of S48 for a glycine (S48G) resulted in a mutant E1 protein that is defective for transient BPV1 DNA replication (McShan and Wilson, 2000). In contrast, substitution of S48 for aspartate, or for glutamate as a phosphomimetic, did not alter the replication activity of the protein (McShan and Wilson, 2000). One- and two-hybrid assays performed in yeast indicated that the S48G E1 protein retains the ability to bind to the ori and to interact with E2 and with itself (oligomerization). Although the precise molecular defect of the S48G E1 protein could not be identified, these results nevertheless support the idea that CK2-phosphorylation of S48 is required for BPV1 DNA replication. As for serine 584, its substitution by alanine (S584A) also resulted in a BPV1 E1 deficient for replication (Lentz et al., 2002). In functional assays in vitro, the S584A E1 retained the ability to interact with E2 and with Pol α-prim. It also displayed near wild type helicase activity when tested with a short DNA duplex substrate. The only defect that was reproducibly detected in vitro was a reduced capacity to bind to the ori (Lentz et al., 2002). S584 lies in the C-terminal brace of the HD. Given that deletion of this short region only partially reduces the assembly and stability of E1 hexamers, resulting in a decreased ability to hydrolyze ATP and to unwind long duplex DNA substrates, it would be interesting to determine if the S584A substitution also affects the ATPase and processive unwinding activities of E1. More generally, these findings raise the possibility that CK2 regulates the activity of the brace, a possibility supported by the fact most E1 proteins contain one or more putative CK2 sites in or near this region of the protein.

Lastly, an attempt was made to identify all of the phosphorylation sites in E1 by mass spectrometry. In this analysis, recombinant BPV1 E1 expressed and purified from insect cells was used to identify sites that are phosphorylated in vivo, albeit in a heterologous system. Two previously reported sites were confirmed (S48 and S584) and several new ones identified (S94, S95, S100, T126, S305) (Lentz et al., 2006). Although the function of the newly identified phosphorylation sites remains to be determined, three (S94, S95 and S100) lie between the two basic regions of the NLS suggesting that they may regulate the nuclear accumulation of BPV1 E1. These three sites are predicted to be substrates of CK2.

Ubiquitination

E1 is present at exceedingly low levels in PV infected cells suggesting that it is synthesized only in low amounts and/or subject to rapid degradation. Studies performed in Xenopus egg extracts and in transfected cells have indicated that BPV1 E1 is an unstable protein that is degraded by the ubiquitin-proteasome pathway (Malcles et al., 2002). Interestingly, it has been found that BPV1 E1 is stabilized when complexed with cyclin E/Cdk2, immediately prior to viral DNA replication, but becomes destabilized following DNA synthesis (Cueille et al., 1998; Malcles et al., 2002). Proteasomal degradation of E1 requires its polyubiquitination by the anaphase-promoting complex/cyclosome (APC/C), a multisubunit E3 ubiquitin ligase that regulates mitosis by triggering the degradation of regulatory proteins such as cyclins (Mechali et al., 2004). Abrogation of APC/C activity by the natural inhibitory protein Em1 led to stabilization of E1 and increased viral DNA replication. Like many APC/C substrates, BPV1 E1 contains a KEN motif (K-E-N, aa 28–30) and a functional destruction box (DB, consensus motif: R-x-x-L, aa 346–349) (Mechali et al., 2004). Mutation of the destruction box (RAFL to AAAF), and to a lesser extent of the KEN motif (KEN to AAN), stabilized the protein in transfected cells, providing further evidence that BPV1 E1 is a target of APC/C (Mechali et al., 2004). It was proposed that the APC/C-mediated degradation of BPV1 E1 helps to maintain the viral genome at a constant low copy number during latent infection (Malcles et al., 2002; Mechali et al., 2004). Because the presence and/or location of the KEN motif and destruction box are not highly conserved in the E1 proteins of HPV types, it remains to be determined if the turnover of these proteins is also regulated by the APC/C ubiquitin ligase.

Sumoylation

A two-hybrid screen with HPV16 E1 as the bait provided the first evidence that E1 can interact with the SUMO-conjugating enzyme Ubc9 (Yasugi and Howley, 1996; Yasugi et al., 1997b). Subsequent studies indicated that the E1 proteins from BPV1, HPV1, HPV11 and HPV18 also interact with Ubc9 (Fradet-Turcotte et al., 2009; Rangasamy and Wilson, 2000). Mapping studies performed with BPV1 and HPV11 E1 indicated that the oligomerization domain of the protein is required for Ubc9-binding in vitro but that the complete HD is needed for interaction in yeast. The DB, although not sufficient, was also found to contribute to the strength of the E1-Ubc9 two-hybrid signal in yeast. Accordingly, several amino acid substitutions in the DB and HD of E1 were identified that could reduce interaction with Ubc9 in yeast (K421A/ K421D; K484A, K484R for BPV1; K228A, Y380A, N389A, F393A, K484A, K484R for HPV11; S530R, Y412F, W439R, G482D, G496R for HPV16) (Fradet-Turcotte et al., 2009; Rangasamy and Wilson, 2000; Yasugi et al., 1997b). As all of these substitutions in the HD have been shown or are predicted to affect the enzymatic activity of E1 these results suggest that a functional helicase is required for interaction with Ubc9 in the yeast two-hybrid system and, more generally, that Ubc9 may prefer to interact with oligomeric E1 in vivo (Fradet-Turcotte et al., 2009).

The E1 proteins from BPV1, HPV1, HPV11 and HPV18 can be sumoylated in vitro and in vivo, albeit not as efficiently as PML, a well-characterized substrate of the SUMO pathway (Fradet-Turcotte et al., 2009; Rangasamy and Wilson, 2000; Rangasamy et al., 2000). Sumoylation of BPV1 and HPV11 E1 can be enhanced 2– to 3-fold in vitro by specific members of the PIAS family of E3 ligases, and in particular by Miz1, which imparts on BPV1 E1 a preference for SUMO1– over SUMO2– conjugation (Rosas-Acosta et al., 2005). The site of SUMO attachment in BPV1 E1 was identified by mutagenesis as lysine 514, which is located close to motif C (sensor 1) in the C-terminal HD (Rangasamy et al., 2000). Substitution of this residue by alanine or arginine was found to abrogate the ability of the protein to support transient DNA replication in vivo (Rangasamy et al., 2000). Initial characterization of the K514A and K514R mutant proteins indicated that they accumulated in the cytoplasm rather than in the nucleus of transfected cells (Rangasamy et al., 2000), but this observation was later corrected. Indeed, it was subsequently reported that the K514R mutant protein (the K514A E1 was not tested) accumulates almost exclusively in the nucleus although it may be slightly more enriched at the nuclear periphery than the wild type protein, which shows a more diffuse pattern throughout the nucleus (Rosas-Acosta and Wilson, 2008). Thus, unlike what was concluded previously, sumoylation of BPV1 E1 is not required for nuclear accumulation of the protein. A similar conclusion was reached for HPV11 and HPV16 E1 using mutant proteins defective for Ubc9-interaction coupled with inhibition of the SUMO pathway either by the Gam1 protein of the CELO adenovirus, the dominant-negative Ubc9 C93S or using shRNAs to deplete Ubc9 (Fradet-Turcotte et al., 2009). Thus, the biological consequences of E1 sumoylation remain largely unknown.
Proteolytic cleavage by caspases-3 and -7

The productive phase of the viral life cycle takes place in the uppermost layers of the epithelium. As part of this differentiation-dependent process, the viral episome is amplified to more than 1000 copies per cell. Surprisingly, it was found that HPV activates caspases-3, -7, and -9 upon differentiation but without any significant induction of apoptosis, suggesting that caspases may play a role in some aspect of the viral life cycle (Moody et al., 2007). Through the use of caspase inhibitors, it was discovered that these enzymes are required for the differentiation-dependent amplification of the viral genome. Examination of the sequence of the E1 helicase revealed that one or more putative caspase cleavage sites (consensus D-x-x-D) are present in the N-terminal region of the protein. For example, HPV31 E1 contains two overlapping sites, D-x-x-D-x-x-D between amino acids 46–52. This motif was found to be efficiently cleaved by caspase-3 and -7 in vitro, as well as in transfected cells treated with a chemical inducer of apoptosis to activate these caspases (Moody et al., 2007). As anticipated, substitution of the central aspartate residue for alanine (D49A) completely eliminated caspase cleavage. The D49A substitution had no effect on the ability of E1 to support transient DNA replication in transfected cells indicating that caspase cleavage of E1 is not required per se for its DNA replication activity. However, in the context of the complete HPV31 genome, the D49A substitution prevented amplification of the viral episome in differentiated keratinocytes (Moody et al., 2007). These results suggested that HPV activates caspases upon differentiation, in part to promote the cleavage of E1 and amplification of the viral episome during the productive phase of the viral life cycle. The molecular mechanism by which caspase-cleavage of E1 contributes to the process of genome amplification is currently unknown. More generally, the finding that the E1 D49A substitution specifically impairs amplification of the viral episome, while having little to no effect on its maintenance in undifferentiated cells, provided direct evidence that E1 is required for this amplification process.

Concluding remarks

Besides being the only enzyme, E1 is also the largest and arguably most complex protein encoded by papillomaviruses. As the repliative helicase, E1 plays a central role in the viral life cycle. To do so, E1 engages in multiple protein–protein and protein–nucleic acid interactions to create a novel and critical interface between the viral origin of replication and the cellular DNA replication machinery. In systems biology terminology, E1 could be described as a hub protein that rewires the DNA replication network of the host to maintain and amplify the viral episome. The last two decades have provided a detailed understanding of how E1 and E2 interact with each other to facilitate the assembly of a replication-competent E1 DH at the origin. Crystal structures of the E1 DBD, either free or bound to DNA, and of the E1 HD in complex with the E2 TAD or in its hexameric, helicase-active form have provided an unprecedented view of how E1 functions as a sequence-specific DNA-binding protein for ori-recognition and transits into a non-specific DNA helicase. Given that most studies on the assembly of E1 and E2 at the origin were performed with a fragment of the BPV1 ori containing a single E2BS, and that most PV origins (including that of BPV1) contain 2 or more E2BS, it is becoming of great interest to investigate if these extra E2BS can influence the nature of the E1 and E2 complexes formed at the ori and, more generally, to pinpoint their exact role in viral DNA replication. As anticipated from its essential functions, E1 contains some of the most conserved domains of all PV proteins, most of them located in the HD and being involved in DNA unwinding or in formation of this functional interface with host DNA replication factors. Region D, however, stands out as one of the highly conserved regions whose function is still poorly defined. While it forms part of the interaction interface with E2, its conservation in SV40 and polyomavirus LT-Ags suggests that it must also be important for another function of these helicases, perhaps in hexamerization or interaction with host factors as suggested by its location within the 3D-structure of these proteins. Viral DNA replication is also critically dependent on the activity of the DBD. In recent years, we have come to realize that the function of this domain is not limited to binding DNA but that it also participates in key protein–protein interactions required for assembly of the E1 DT and DH at the ori, or for the recruitment of specific host factors to the replication fork. The DBD is the second most conserved domain of E1. The reason why it exhibits slightly less conservation than the HD is likely because interaction with E2 also plays an important role in ori-recognition. Since formation of the E1–E2–ori complex relies entirely on virally encoded protein–protein and protein–DNA interfaces, it is less evolutionary constrained, as highlighted by the fact that the E1 and E2 proteins of different PV types cannot always be interchanged in functional assays (Chiang et al., 1992; Gopalakrishnan et al., 1999; Zou et al., 1998). It will be important to take into account this heightened genetic variability when considering the development of antiviral agents that target the E1–E2 interaction, as it will likely limit the activity of these drugs to a subset of PV types, as already observed (White et al., 2003). Together, the E1 DBD and HD are sufficient to support PV DNA replication in vitro and, as such, constitute the core of the molecular motor that drives PV DNA replication. In vivo, this process is tightly controlled, in part through the N-terminal part of E1. This in vivo regulatory region is the least conserved segment of E1; this greater evolutionarily divergence likely reflecting the adaptation of different PVs to their particular host. The E1 regulatory region is mostly unstructured and, like many other disordered protein domains, is rich in short regulatory motifs and sites of post-translational modifications arranged in a combinatorial fashion. An important function of this domain is to regulate the nuclear accumulation of E1, either by modulating its nuclear import, export, or both, most often in a phosphorylation-dependent manner. Although the details of this regulatory mechanism vary across PV types, it appears that its overall purpose is to synchronize replication of the viral genome with that of the host and to determine the magnitude of the replication/amplification process during the different phases of the viral life cycle, by controlling the amount of E1 in the nucleus. It is likely that other cellular factors that regulate and/or participate in PV DNA replication will be discovered in the future, in particular as it pertains to the roles of the cellular DNA damage and repair pathways in this process. Many host proteins and enzymes have already been found to interact with E1. While few of these interactions have been shown to be essential for PV DNA replication to date, it is likely that at least a subset of them will be. As further research identifies and characterizes these critical E1-host factor interactions, we anticipate that the E1 protein domains required for binding these cellular proteins will be highly conserved, perhaps even more so than those involved in E2-binding given that cellular targets exhibit far less genetic variation than E2. This may pave the way to the identification of small-molecule drugs that modulate these E1-host protein interactions and re-invigorate research into targeting E1 for the development of anti-PV drugs, which, unfortunately, has been met with little success so far due to the intrinsic plasticity of the enzyme catalytic site. In addition to remaining a valid drug target, E1 also continues to be a model enzyme for the study of helicase activity. We can be sure that additional research into the structure and function of E1 will lead to further insights into the mechanisms of DNA unwinding and eukaryotic DNA replication in general. The great number of
functional E1 enzyme sequences available in the PaVE database will undoubtedly be of great assistance to those interested in these basic and therapeutic research areas.

Acknowledgments

We apologize to those whose work was not included because of space considerations or whose papers were unintentionally omitted. We thank Dr Peter Bullock (Tufts University) and members of the Archambault and Melendy laboratories for critical reading of this manuscript. Work in the authors’ laboratories is supported by grants from the US National Institutes of Health (NIH - AI095632), Canadian Institutes for Health Research (CIHR), Canadian Cancer Society Research Institute (CCSRI) and the Cancer Research Society (CRS).

References

Ferrant, M.C., McBride, A.A., 1998.Transient viral DNA replication and repression of viral transcription are supported by the C-terminal domain of the bovine papillomavirus E1 protein. J. Virol. 72, 796–801.
Han, Y., Loo, Y.M., Milletelo, K.T., Melendy, 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.


