The E1 Proteins

Francis Sverdrup[a] **and Gerald Myers**[b]

[a] *Department of Dermatology, New England Medical Center, Box 166,*
750 Washington Street, Boston, MA 02111
e-mail: fsverdru@opal.tufts.edu
[b] *Theoretical Biology and Biophysics, MS K710, Los Alamos National Laboratory,*
Los Alamos, NM 87545

Introduction

The papillomavirus E1 protein is the 'initiator' protein for viral DNA replication and thus plays a critical role in maintenance of the viral genome as an episome during the viral life cycle. E1 shows some sequence and structural similarities to the simian virus 40 and polyomavirus large T antigen initiators (12), and its biochemical properties are consistent with its role as a replication initiator protein. In BPV, E1 is a nuclear phosphoprotein with origin binding, ATPase, and DNA helicase activities.

In addition to E1, the full length viral E2 transactivator is also required for efficient initiation of viral DNA replication *in vivo*. E1 and E2 form a complex in solution and this complex binds with high affinity and specificity to the replication origin, which contains binding sites for both proteins. Therefore, part of the role of E2 is to stimulate binding of E1 to the replication origin. E2 also suppresses nonspecific initiation of replication by E1 *in vitro*. In this way, E2 serves as a specificity factor in promoting origin-dependent replication. The initial E1:E2:origin complex is thought to be inactive for replication initiation and is proposed to serve as a seed for formation of an active replication complex. In a poorly understood process, E2 is released from the complex and additional E1 molecules enter and oligomerize on the origin DNA. This active E1:DNA complex, in conjunction with the host cell replication machinery, results in localized melting of the DNA at the origin followed by initiation of new DNA strand synthesis.

Aside from its requirement for viral DNA replication, E1 also plays a role in the regulation of viral transcription and transformation. Presumably through interacting with E2, E1 modulates the ability of E2 to activate certain viral promoters. In BPV, E1 negatively regulates the activation of the major early promoter by E2 (27, 51) and this mechanism may be responsible for the suppression of transformation by E1 (26, 55, 72, 81). E1 additionally exerts direct effects on the host cell cycle (1, 2).

E1 is the most highly conserved of the papillomavirus genes (12, 32). Accordingly, the E1 gene products of different papillomaviruses are expected to have similar structure and functions and in general this appears to be the case (Appendix A). Since most detailed studies have been done with BPV-1 E1, this review will focus on and summarize the activities which have been demonstrated for this protein.

A. E1 gene products

The BPV-1 E1 ORF encodes two gene products expressed in BPV transformed cells from the major early promoter, P89. The full length product of the ORF is a 68- to 72-kDa nuclear phosphoprotein (5, 52, 63). This protein is required for viral DNA replication and its activities will be discussed in detail below. A smaller protein, termed E1-M, is encoded by a spliced transcript which fuses the 5′ third of the E1 ORF with a 1235∧3225 splice to 14 amino acids from a downstream exon. E1-M exists as a 23-kDa phosphoprotein (67) and was initially ascribed a negative regulatory function in replication which was required for stable replication of viral plasmid DNA (31). More recent findings, however, indicate that E1-M is not essential for stable plasmid replication in BPV-1 transformed cells (25). It remains possible that E1-M plays a role in the amplification of viral DNA during the vegetative phase of BPV-1 infection.

Several human papillomaviruses also express transcripts which conceptually code for E1-M-like proteins. One such transcript of HPV-11 encodes a protein containing the amino terminal quarter of the E1 ORF fused to the E2 carboxy terminal DNA binding domain (10). Expression of this gene product in tissue culture assays results in repression of both papillomavirus transcription and DNA replication (10).

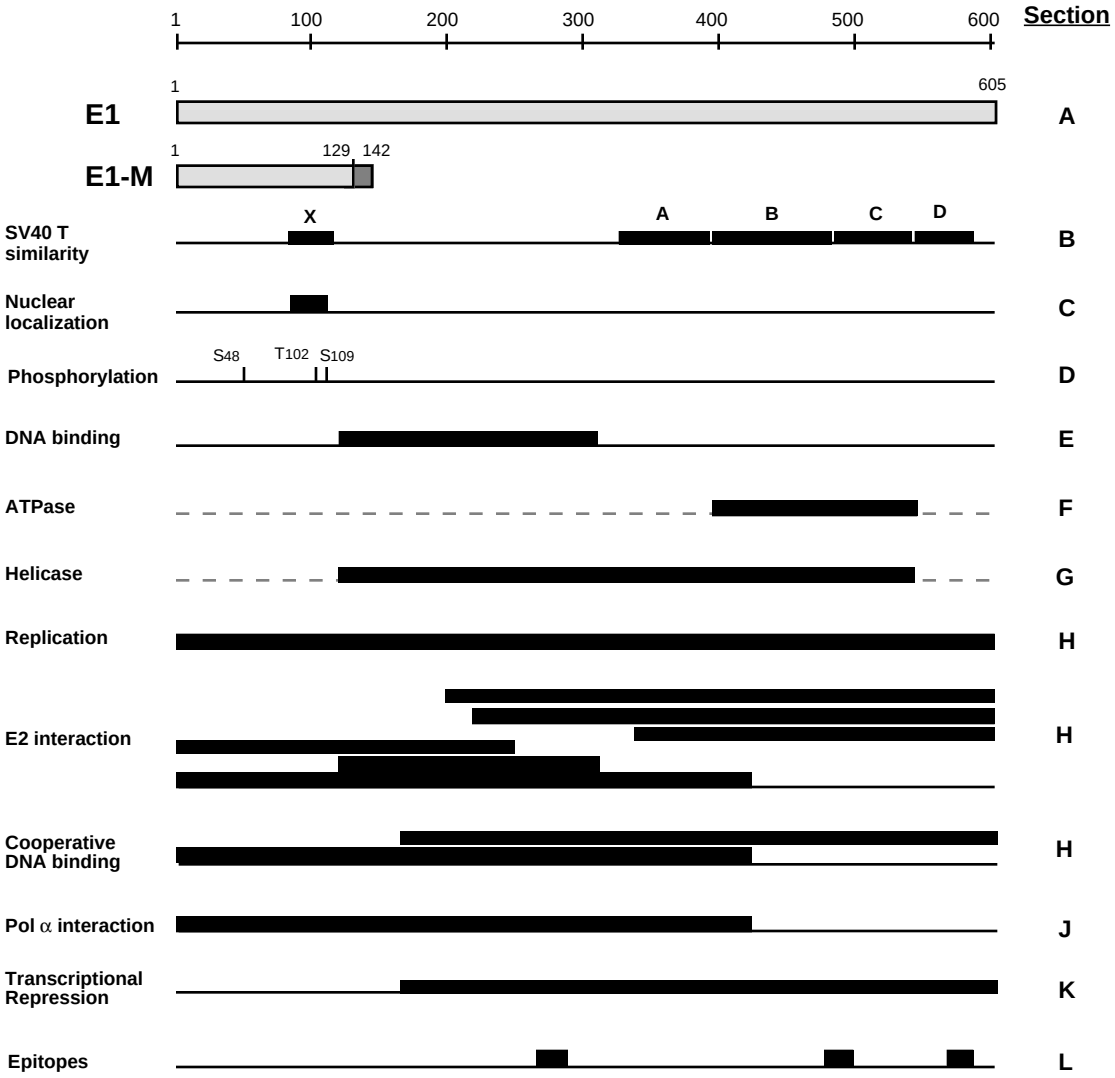


Figure 1. The structure of the two BPV-1 E1 proteins are shown at the top of the figure. Below, the functions that have been mapped to different regions of the proteins are indicated. Refer to the section indicated to the right for more details. Solid bars indicate regions implicated in a function (e.g., ATPase, helicase) but do not imply that those regions have been shown to be sufficient.

However, the involvement of E1-M-like gene products in the bovine or human papillomavirus life cycle has not been established. Finally, the five (typically) NH_2 -terminal residues of E1 form a fusion protein with the E4 gene product as described by Doorbar and Myers in Part III, pp. 65–66, of the August 1996 release of this compendium.

B. Structural similarity to papovavirus large T antigens

The full length E1 proteins of both bovine and human papillomaviruses share sequence similarities with the large T antigens of simian virus 40 (SV40) and polyomavirus (PyV) (12). Since SV40 and PyV large T proteins are viral DNA replication initiators, these observations suggest similar functions for the E1 proteins in the initiation of papillomavirus DNA replication. There is a wealth of information concerning the structure and function of SV40 large T (see reference (15) for review), and comparisons between E1 and large T have been helpful in analyzing E1 function. Five regions of BPV-1 E1 have

similarity to the large T proteins and these have been designated X, A, B, C, and D by Clertant and Seif (Figure 1B) (12). The most significant similarities occur in regions B and C corresponding to sites involved in the ATPase and nucleotide-binding activities, respectively, of SV40 T antigen (12). A parvovirus noncapsid protein also displays significant similarity across the B and C regions of E1 (Appendix A). For predictions of E1 secondary structure, see Appendix B. In the sections below, activities associated with specific regions of E1 will be discussed and, where appropriate, comparisons to SV40 T antigen will be made.

C. E1 localization

Consistent with a primary role in viral DNA replication, the E1 proteins are found in the nucleus of infected cells (5, 52, 63, 71). Sequences within the first 223 amino acids are necessary for directing this cellular localization (63); this segment of BPV-1 E1 contains the above mentioned region "X" of similarity between E1 and SV40 T antigen (12), which includes nuclear localization signal (NLS) in large T antigen (15). Mutational analysis of region "X" of BPV-1 E1 has defined a bipart NLS located between amino acids 84 and 108 which contains two short stretches of basic amino acids: KRKVLGSSQNSSGSEASETPVKRRK (29). A small region of E1 (aa 84–166) containing this NLS is both necessary and sufficient to direct nuclear localization of fusion proteins (74).

D. Phosphorylation

The E1 protein of BPV exists in a phosphorylated form in cells (5, 52, 63). Since nonphosphorylated E1 expressed and purified from bacteria is active for BPV DNA replication in vitro, phosphorylation is not absolutely required for its activities directly related to replication (46). However, it is likely that phosphorylation is involved in regulating the activities of E1 *in vivo*. Although the residues of E1 which are phosphorylated *in vivo* are not completely mapped, phosphoamino acid analysis indicates that probably only one threonine residue is phosphorylated with the rest of the detectable phosphorylations being on serine residues (29). A combination of in vitro phosphorylation studies, mutational analyses, and analyses of E1 phosphorylation and activity in cells is beginning to shed light on the role of phosphorylation of specific residues of E1. In vitro studies have shown that E1 serves as a substrate for phosphorylation by a number of cellular kinases including p34cdc2, protein kinase A (PKA), protein kinase C (PKC), casein kinase I (CKI), casein kinase II (CKII), and DNA dependent protein kinase (29, 39, 79, 80). A small subset of these phosphorylations have been mapped to specific residues of E1 (Figure 1D). The cell cycle regulated kinase p34cdc2 phosphorylates threonine 102 in vitro and this residue is a phosphoacceptor in vivo, but the effect of phosphorylation at this site is not entirely clear (29). Thr-102 is located within the extended NLS of E1 (see above), yet mutation of thr-102 to isoleucine had no effect on nuclear localization or DNA replication in the context of the intact viral genome (29). In contrast, another report indicated that a thr-102 to alanine mutation resulted in a replication-deficient genome (32). Serine 109 is also a phosphoacceptor in vivo, and is a target for phosphorylation in vitro by both PKA and PKC (80). In this case phosphorylation appears to be negatively regulating: mutation of ser-109 to alanine results in a BPV genome that replicates more efficiently than wild-type *in vivo* in transient assays while mutation of ser-109 to glutamic acid (which may functionally mimic a constitutively phosphorylated serine residue) results in a genome that replicates less efficiently than wild-type (80). It has been proposed that phosphorylation at ser-109 may affect E1 DNA binding activity (80). Thus, E1 may potentially be subject to both positive and negative regulation by phosphorylation *in vivo*. Further research will be necessary in order to understand in detail the modulation of E1 activity by phosphorylation.

E. DNA binding

An important function of viral initiator proteins is to mark the region of replication initiation within the viral genome through sequence-specific DNA binding. The BPV-1 E1 protein has been shown to possess DNA binding activity specific for sequences located within the viral replication origin (59–61, 68, 71, 73, 75). The binding site for E1 is contained within an 18-bp inverted repeat element located in the origin (23, 24, 41). Although E1 shares no sequence homology with the SV40 T antigen

DNA binding domain, similarities in structural organization predict that the amino acids of E1 involved in DNA binding would reside in the amino-terminal half of the protein (12, 15). Deletion mapping studies have confirmed this speculation, identifying amino acids 162–284 as being the most critical for DNA binding function while the smallest peptide that retains origin-specific DNA binding activity consists of amino acids 121–311 (Figure 1E) (28, 54, 68). Mutation at sites within this region inactivate DNA binding by full length E1 (28, 68).

F. ATPase activity

The papillomavirus E1 proteins share the greatest similarity with the regions of T antigen involved in nucleotide binding and ATPase activity (Figure 1B, regions B and C) (12). Indeed, BPV E1 was initially shown to possess ATP binding activity as measured by crosslinking with periodate-oxidized ATP (52, 63) and ATPase activity was subsequently demonstrated for E1 (35, 53, 60). Several point mutations in homology regions B and C (P434S, K439E, D479A, H507L) inactivate the abilities of BPV E1 to bind and hydrolyze ATP and to support DNA replication, confirming the importance of this region of E1 for these functions (35, 48a, 63). The consensus sequence for the so-called P-loop (phosphate-loop) for ATP binding is GXXXXGK(T/S); in BPV-1, this sequence is GPPNTGKS (48a). The solid bar in Figure 1 indicates the region of E1 implicated in ATP binding and ATPase activity but does not imply that other regions may not also be involved.

G. Helicase activity

The structural similarities between E1 and T antigen combined with the demonstration of E1 ATPase activity strongly suggested that E1, like T antigen, would possess helicase activity, that is the ability to unwind duplex DNA. Using a substrate consisting of a short oligonucleotide annealed to a single stranded DNA circle, E1 was shown to displace the oligonucleotide from the circular DNA in a reaction that requires $MgCl_2$ and the energy of nucleotide triphosphate hydrolysis (60, 76). E1 also unwound double stranded supercoiled DNA containing the BPV-1 replication origin and, at higher E1 concentrations, unwound origin-minus DNA in a reaction that additionally requires a single stranded DNA binding protein (SSB) (18, 60, 76). Since helicase activity requires binding of E1 to the DNA substrate, the region of E1 required for helicase activity is predicted to extend from the DNA binding domain (see above) beginning around amino acid 121 through the ATPase/nucleotide binding region to approximately aa 530 (Figure 1G).

H. Replication

In order to directly assess the requirements for viral gene products in replication of viral DNA in vivo, transient transfection assays have been utilized in which expression vectors for individual viral proteins are cotransfected with a viral origin-containing plasmid. Replication of the origin plasmid is monitored by Southern blot analysis. Using this type of assay, it was found that replication of BPV-1 DNA in vivo requires both the E1 and E2 gene products (69–71). These findings were subsequently extended to include several human papillomavirus types (9, 14, 17, 19, 65). The E2 protein is a sequence-specific DNA binding transcriptional regulator of viral gene expression (for reviews see (38) and the chapter on E2 Proteins in this compendium). Several lines of evidence indicate that E2 plays an auxiliary role in stimulating replication and that E1 is functionally equivalent to SV40 T antigen as the initiator of viral DNA replication. When BPV-1 E1 alone is expressed to high levels in cells, a low level of origin-specific replication is detectable (21). Replication has also been noted when the E1 genes of human papillomavirus types 1a and 16 were expressed in the absence of E2 (19, 50). In fact, a modified origin from HPV-1a containing multiple E1 binding sites supported replication in the presence of E1 alone to levels comparable to replication of the wild-type origin in the presence of both E1 and E2 (20). Thus, E1 proteins appear capable of recruiting all of the cellular factors required for viral DNA replication *in vivo*. Nonetheless, E2 is highly stimulatory for replication of both BPV and HPVs and is required for maximal levels of replication *in vivo* (19, 50, 70).

The origin of replication of BPV-1 contains three elements essential for activity: an 18-bp palindromic sequence which serves as a binding site for E1, a 12-bp sequence which is a binding site for

E2, and an AT-rich sequence element (23, 71, 75). The roles of E1 and E2 in replication and their interaction with the origin have been further explored using in vitro replication systems. In vitro, E1 is the only viral protein required for the initiation of DNA replication (6, 40, 46, 60, 75, 76). However, at low levels of the E1 protein, the presence of E2 stimulates E1-dependent replication, indicating a direct role for E2 in regulating viral DNA replication (75). E1 and E2 interact to form a tight complex in solution (5, 32, 42). This interaction results in cooperative binding of E1 and E2 to the origin and stimulation of origin unwinding and replication initiation by E1 (18, 34, 42, 56, 59, 61, 75).

A peculiar difference between E1 and SV40 T antigen with regard to *in vitro* DNA replication is the unique ability of E1 to replicate DNAs which do not contain the viral replication origin (6, 7, 60, 76). This ability to initiate DNA replication in a sequence-independent manner may be related to the tendency of E1 to bind DNA with only a low degree of sequence specificity (56). In the presence of E2, however, the binding specificity of E1 for the origin of replication is greatly enhanced and replication proceeds only on origin-containing templates (7, 56). In one study, stimulation of origin-specific replication by E2 required cooperative binding of E1 and E2 to an origin template containing binding sites for both proteins (56). However, in another study using highly active preparations of E1, E2 did not have a highly stimulatory effect on origin-specific replication and instead acted only to suppress nonspecific replication by E1 (7). This activity of E2 did not require an E2 binding site in the origin (7). Nonetheless, it is clear that E2 does act as a "specificity factor" in directing E1 to the papillomavirus origin and that this activity would be critical in vivo where replication of the papillomavirus genome must proceed efficiently in the midst of a vast excess of cellular DNA. The replication of HPVs in vitro has not been studied as extensively as for BPV, but in vivo results point to some interesting differences in how various subgroups of HPVs achieve their replication specificity. For example, HPV-1a is similar to BPV-1 in that the minimal *in vivo* replication origin contains binding sites for both E1 and E2 (20, 69). In contrast, the E1 and E2 proteins of HPV-18 and HPV-11 can utilize an origin containing only two adjacent E2 binding sites (without an E1 binding site) to initiate replication *in vivo* (30, 66). In the latter case E1 must be able to initiate replication nonspecifically from vector DNA when tethered there through interaction with E2. Accordingly, there is only one location in the genomes of HPVs 11 and 18 and other genital HPVs in which there are two adjacent E2 binding sites: the replication origin (30, 66). Thus, the evolution of papillomaviruses appears to have conserved the requirement for E2 in replication of the genome and at least part of this requirement is as a specificity factor for directing E1 to the viral origin of replication.

The interaction between E1 and E2 is an essential component of the ability of E2 to stimulate E1-dependent replication. This interaction appears to involve at least two regions of E1 which can independently bind the N-terminal activation domain of the E2 protein. With BPV, several groups have noted an interaction between the E2 protein and different E1 peptides which overlap near the DNA binding domain of E1 at amino acids 121–222 (Figure 1H) (3, 28, 68). This interaction is reported to be cold sensitive such that interaction occurs at 37° C but does not occur at 4° C. Other groups have detected an E2 interaction with a more C-terminal region of E1 using peptides which overlap at amino acids 340–605 (32, 44, 54). This interaction is not cold sensitive. Because some of the E1 peptides used in the above studies contain portions of both regions of E1 and because of the different assay conditions employed, it is difficult to assess the relative strength and importance of the two interacting regions of E1. Mutations in either of the two regions of E1 which disrupt their respective interactions with E2 result in E1 proteins defective for replication *in vivo* (28, 32). It will be interesting to determine if these mutants are still active for replication under conditions where E2 is not required (e.g., *in vitro*). Nonetheless, it appears that both interactions between E1 and E2 are required for activation of replication *in vivo*. Additionally, although the studies mentioned above have implicated the activation domain of E2 as being involved in binding to E1, a recent study has indicated that the C-terminal DNA binding domain of E2 can also independently interact with E1, though this interaction does not appear to be required for replication (4).

An important aspect of the E1-E2 interaction is that it results in cooperative binding of the two proteins to origin DNA (18, 34, 56, 59, 61, 75). In BPV, the minimal origin is sufficient for cooperative DNA binding and contains binding sites for both proteins (34, 59, 69). The combination of protein-protein and protein DNA interactions can explain the cooperative DNA binding by the two proteins.

Since E1 also possesses a nonspecific DNA binding activity, E2 dramatically increases the specificity of E1 for origin DNA in the presence of nonspecific competitor DNA (56). In fact, cooperative binding of E1 and E2 to origin DNA is a necessary component of the stimulation of replication by E2 (56). The region of E1 required for cooperative DNA binding with E2 likely includes its DNA binding domain and the E2-interaction regions (Figure 1H). A peptide containing E1 amino acids 1–423 retains some ability to cooperatively bind DNA with E2 (68). However, the ability to cooperatively bind DNA as efficiently as wild-type requires the entire C-terminus of E1 including amino acids 162–605 (54). Therefore, E1 utilizes multiple interactions with E2 to form a protein complex with enhanced affinity and specificity for origin DNA. The specific complexes formed on DNA with E1 and E2 are discussed in more detail below in the section "Origin complexes."

Studies with several human papillomaviruses, though less detailed, have generally paralleled the studies with BPV-1. In HPV-16, the C-terminal 229 amino acids (aa 421–649) of E1 are capable of association with the transactivation domain of E2, but a larger domain including amino acids 144–649 is necessary for efficient complex formation with E2 at room temperature (22, 48, 62, 77). In HPV-31b, the N-terminal 268 amino acids of E1 is sufficient for complex formation with E2 (17), but a larger region is required for stimulation of E1 binding by E2 (16, 17). In HPV-33, a cold sensitive interaction occurs between E2 and E1 amino acids 312–450 while a more stable interaction requires the C-terminus from amino acids 312–644 (45).

I. Origin complexes

The conversion of E1 from a site-specific DNA binding protein to a helicase which presumably travels in front of the replication fork is poorly understood. Several groups have focused on this problem by analyzing the protein complexes that occur on origin DNA in the presence of E1 and E2 or in the presence of E1 alone. Since E2 is required for efficient replication of DNA in vivo, it is likely that the initial protein-DNA complexes contain both E1 and E2. When both proteins are present in vitro, a complex containing one dimer of E2 and probably one monomer of E1 binds to the adjacent E1 and E2 binding sites within the minimal origin and protects them from nuclease and chemical cleavage (18, 41, 56–58, 68). In the absence of E2, E1 is capable of a footprint that extends to include the AT-rich element located adjacent to the E1 binding site as well as additional protections within the E1 binding site (18, 57, 58). This extended footprint correlates with an increase in the number of E1 molecules present within the E1-origin complex (33, 57, 58). The homo-oligomeric E1 complex on DNA is thought to be the active species for replication initiation and estimates of the number of E1 molecules present in this complex have ranged from 3 to 10–15 (7, 11, 33, 57) Based on these data and comparisons to SV40 DNA replication, the following bi-phasic model for initiation of papillomavirus replication at a minimal origin has been proposed: i) one dimer of E2 and one monomer of E1 cooperatively bind the origin of replication; ii) additional monomers of E1 assemble onto the origin with the concomitant loss of E2. The larger oligomeric E1 complex would possess the activities responsible for unwinding the origin and attracting the cellular replication machinery to initiate DNA synthesis.

J. Interaction with cellular proteins

Consistent with the similarities between E1 and SV40 T as replication initiator proteins and helicases, the BPV-1 E1 protein has been shown to interact with the cellular polymerase α-primase complex (polα-primase) (6, 47). Primase-helicase associations are evolutionarily conserved and are presumed to be critical for initiation of replication (36). Thus, in relying heavily on the host cell to provide the replication machinery, the papillomaviruses and the polyomaviruses utilize their respective initiator helicases, E1 and large T antigen, to attract the cellular polα-primase to the viral replication origin. A region of the BPV-1 E1 protein including the amino terminal 424 amino acids was shown to interact specifically with the p180 (catalytic) subunit of the four subunit polymerase α-primase complex (47). This association is different from that of both SV40 large T antigen which interacts more strongly with the p70 subunit (13) an d polyoma large T antigen which targets the small p48 primase subunit (8). Additionally, the interaction between E1 and polα-primase did not exhibit the species-specificity observed with SV40 (primate) and polyoma (murine) large T antigens (6, 47). This result is consistent

with the earlier findings that transient papillomaviral replication occurred promiscuously in both primate and murine cell lines (9, 71).

Two other cellular proteins have been identified as interacting with HPV-16 E1. hUBC9, the human counterpart of *S. cerevisiae* UBC9, is a ubiquitin conjugating enzyme which may participate in cell cycle progression. One HPV-16 E1 mutant, S330R, was defective for hUBC9 interaction and DNA replication, suggesting that the association may be necessary for the replication function of E1 (78). Another E1-interacting protein, 16E1-BP, has no known function but does contain an ATPase motif (78).

K. Transcriptional repression / transformation / transactivation

BPV-1 genomes with mutations in the E1 gene showed increased transforming capacity and increased transcription from the major early promoter, P89 (26, 55). This data implied that E1 may be a transcriptional repressor involved in the negative regulation of viral oncogenes. Consistent with this idea, E1 repressed E2-stimulated transcription from the major early promoter (27, 51). This repression by E1 correlated with the ability of an E1-E2 complex to bind to the replication origin, located just upstream of P89, but did not depend on replication (51). The C-terminal 450 amino acids of E1 are sufficient for this repression (37). The use of additional E1 and E2 mutants in the context of the BPV-1 genome has indicated that E1 and E2 together suppress transformation by the E5, E6 and E7 oncogenes (72, 81). This theme is also apparent for the high-risk human papillomavirus type 16, in which disruption of either the E1 or E2 genes results in deregulation of viral gene expression and increased capacity for immortalization of primary human keratinocytes (49). Taken together, these results strongly implicate the involvement of E1 in the negative regulation of oncogene expression during the viral life cycle. In contrast to negative transcriptional effects, several researchers have observed stimulation of transcription by E1 alone or by E1 in conjunction with E2 in certain circumstances (27, 43, 48, 64). However, a target for transcriptional activation by E1 has not been identified.

By regulating the expression of viral oncogenes, E1 has indirect effects on altering the host cell growth program. However, a more direct effect of E1 on the host cell has been demonstrated recently. Cells expressing E1 in the absence of other viral gene products exhibit perturbations of the cell cycle, resulting in decreased duration of the G1 phase and increased S and G2 phase durations (1, 2). These cell cycle effects are coincident with changes in histone H1 kinase activity and the abundance and timing of appearance of cyclin D1 (2). Thus, in addition to viral oncoproteins, E1 may participate directly in altering the host cell growth program which is an essential component of the papillomaviral life cycle.

L. E1 Antigenicity.

In general, the E1 protein is not highly immunogenic. In a study of 54 HPV-8 reactive sera from patients without apparent skin disease, 12 displayed anti-E1 reactivity, and the majority of these sera were not anti-L1 (61a). Dillner (14a) reports two linear epitopes in the C-terminal region of HPV-16 E1, SKSHFWLQPLADAKIGMLDD and FDENGNPVYELNDKNWKS. Of 30 patients with HPV-16-associated cervical neoplasia, the majority of sera recognized the first of these epitopes. In a subsequent study (14b), IgG reactivity against an epitope in the middle of HPV-16 E1, LSKLLCVSPMCM–MIEPPKLR, was significantly correlated with cervical cancer.

References

[1] Belyavskyi, M., J. Miller and V. Wilson. 1994. The bovine papillomavirus E1 protein alters the host cell cycle and growth properties. *Virology*. **204**:132–143.

[2] Belyavskyi, M., M. Westerman, L. Dimichele and V. G. Wilson. 1996. Perturbation of the host cell cycle and DNA replication by the bovine papillomavirus replication protein E1. *Virology*. **219**:206–219.

[3] Benson, J. D. and P. M. Howley. 1995. Amino-terminal domains of the Bovine Papillomavirus Type 1 E1 and E2 proteins participate in complex formation. *J. Virol*. **69**:4364–4372.

[4] Berg, M. and A. Stenlund. 1997. Functional interactions between papillomavirus E1 and E2 proteins. *J. Virol*. **71**:3853–3863.

[5] Blitz, I. L. and L. A. Laimins. 1991. The 68-kilodalton E1 protein of bovine papillomavirus is a DNA-binding phosphoprotein which associates with the E2 transcriptional activator *in vitro*. *J. Virol*. **65**:649–656.

[6] Bonne-Andrea, C., S. Santucci, P. Clertant and F. Tillier. 1995. Bovine papillomavirus E1 binds specifically DNA polymerase a but not replication protein A. *J. Virol*. **69**:2341–2350.

[7] Bonne-Andrea, C., F. Tillier, G. D. McShan, V. G. Wilson and P. Clertant. 1997. Bovine papillomavirus type 1 DNA replication: the transcriptional activator E2 acts as a specificity factor. *J. Virol*. **71**:6805–6815.

[8] Bruckner, A., F. Stadblauer, L. A. Guarino, A. Brunahl, C. Schneider, C. Rehfuess, C. Prives, E. Fanning and H.-P. Nasheuer. 1995. The mouse DNA polymerase α-primase subunit p48 mediates species-specific replication of polyomavirus DNA *in vitro*. *Mol. Cell. Biol*. **15**:1716–1724.

[9] Chiang, C.-M., M. Ustav, A. Stenlund, T. F. Ho, T. R. Broker and L. T. Chow. 1992. Viral E1 and E2 proteins support replication of homologous and heterologous papillomaviral origins. *Proc. Natl. Acad. Sci. USA*. **89**:5799–5803.

[10] Chiang, C. M., T. R. Broker and L. T. Chow. 1991. An E1M–E2C fusion protein encoded by human papillomavirus type 11 is a sequence-specific transcription repressor. *J. Virol*. **65**:3317–3329.

[11] Clertant, P. Personal Communication.

[12] Clertant, P. and I. Seif. 1984. A common function for polyoma virus large-T and papillomavirus E1 proteins? *Nature*. **311**:276–279.

[13] Collins, K. L., A. R. Russo, B. Y. Tseng and T. J. Kelly. 1993. The role of the 70 kDa subunit of human DNA polymerase a in DNA replication. EMBO J. **12**:4555–4566.

[14] Del Vecchio, A. M., H. Romanczuk, P. M. Howley and C. C. Baker. 1992. Transient replication of human papillomavirus DNAs. *J. Virol*. **66**:5949–5958.

[14a] Dillner, J. 1990. Mapping of linear epitopes of human papillomavirus type 16: the E1, E2, E4, E5, E6 and E7 open reading frames. *Int. J. Cancer* **46**:703–711.

[14b] Dillner, J. 1995. Antibodies against linear and conformation epitopes of human papillomavirus type 16 that independently associate with incident cervical cancer.. *Int. J. Cancer* **60**:377–382.

[15] Fanning, E. and R. Knippers. 1992. Structure and function of simian virus 40 large tumor antigen. *Annu. Rev. Biochem*. **61**:55–85.

[16] Frattini, M. G. and L. A. Laimins. 1994. Binding of the human papillomavirus E1 origin-recognition protein is regulated through complex formation with the E2 enhancer-binding protein. *Proc. Natl. Acad. Sci. USA*. **91**:12398–12402.

[17] Frattini, M. G. and L. A. Laimins. 1994. The role of the E1 and E2 proteins in the replication of human papillomavirus type 31b. *Virology*. **204**:799–804.

[18] Gillette, T. G., M. Lusky and J. A. Borowiec. 1994. Induction of structural changes in the bovine papillomavirus type 1 origin of replication by the viral E1 and E2 proteins. *Proc. Natl. Acad. Sci. USA*. **91**:8846–8850.

[19] Gopalakrishnan, V. and S. A. Khan. 1994. E1 protein of human papillomavirus type 1a is sufficient for initiation of viral DNA replication. *Proc. Natl. Acad. Sci. USA*. **91**:9597–9601.

[20] Gopalakrishnan, V., S. Walker and S. A. Khan. 1995. Stimulation of human papillomavirus type 1a DNA replication by a multimerized AT-rich palindromic sequence. *Virology*. **214**:301–306.

[21] Grossel, M. J., F. Sverdrup, D. E. Breiding and E. J. Androphy. 1996. Transcriptional activation function is not required for stimulation of DNA replication by bovine papillomavirus type 1 E2. *J. Virol*. **70**:7264–7269.

[22] Hibma, M. H., K. Raj, S. J. Ely, M. Stanely and L. Crawford. 1995. The interaction between human papillomavirus type 16 E1 and E2 proteins is blocked by an antibody to the N-terminal region of E2. *Eur. J. Biochem*. **229**:517–525.

[23] Holt, S. E., G. Schuller and V. G. Wilson. 1994. DNA binding specificity of the bovine papillomavirus E1 protein is determined by sequences contained within an 18-base-pair inverted repeat element at the origin of replication. *J. Virol*. **68**:1094–1102.

[24] Holt, S. E. and V. G. Wilson. 1995. Mutational analysis of the 18-base-pair inverted repeat element at the bovine papillomavirus origin of replication: identification of critical sequences for E1 binding and *in vivo* replication. *J. Virol*. **69**:6525–6532.

[25] Hubert, W. G. and P. F. Lambert. 1993. The 23-kilodalton E1 phosphoprotein of bovine papillomavirus type 1 is nonessential for stable plasmid replication in murine C127 cells. *J. Virol*. **67**:2932–2937.

[26] Lambert, P. F. and P. M. Howley. 1988. Bovine papillomavirus type 1 E1 replication-defective mutants are altered in their transcriptional regulation. *J. Virol*. **62**:4009–4015.

[27] Le Moal, M. A., M. Yaniv and F. Thierry. 1994. The bovine papillomavirus type 1 (BPV1) replication protein E1 modulates transcriptional activation by interacting with BPV1 E2. *J. Virol*. **68**:1085–1093.

[28] Leng, X., J. H. Ludes-Meyers and V. G. Wilson. 1997. Isolation of an amino-terminal region of bovine papillomavirus type 1 E1 protein that retains origin binding and E2 interaction capacity. *J. Virol*. **71**:848–852.

[29] Lentz, M. R., D. Pak, I. Mohr and M. R. Botchan. 1993. The E1 replication protein of bovine papillomavirus type 1 contains an extended nuclear localization signal that includes a p34cdc2 phosphorylation site. *J. Virol*. **67**:1414–1423.

[30] Lu, J. Z., Y. Sun, R. C. Rose, W. Bonnez and D. J. McCance. 1993. Two E2 binding sites (E2BS) alone or one E2BS plus an A/T-rich region are minimal requirements for the replication of the human papillomavirus type 11 origin. *J. Virol*. **67**:7131–7139.

[31] Lusky, M. and M. R. Botchan. 1986. A bovine papillomavirus type 1-encoded modulator function is dispensable for transient viral replication but is required for the establishment of the stable plasmid state. *J. Virol*. **60**:729–742.

[32] Lusky, M. and E. Fontane. 1991. Formation of the complex of bovine papillomavirus E1 and E2 proteins is modulated by E2 phosphorylation and depends upon sequences within the carboxyl terminus of E1. *Proc. Natl. Acad. Sci. USA*. **88**:6363–6367.

[33] Lusky, M., J. Hurwitz and Y.-S. Seo. 1994. The bovine papillomavirus E2 protein modulates the assembly of but is not stably maintained in a replicationn-competent multimeric E1-replicaiton origin complex. *Proc. Natl. Acad. Sci. USA*. **91**:8895–8899.

[34] Lusky, M., J. Hurwitz and S. Y.-S. 1993. Cooperative assembly of the bovine papilloma virus E1 and E2 proteins on the replication origin requires an inact E2 binding site. *J. Biol. Chem*. **268**:15795–15803.

[35] MacPherson, P., L. Thorner, L. M. Parker and M. Botchan. 1994. The bovine papillomavirus E1 protein has ATPase activity essential to viral DNA replication and efficient transformation of cells. *Virology*. **204**:403–408.

[36] Marians, K. J. 1992. Prokaryotic DNA replication. Annu. Rev. Biochem. **61**:673–719.

[37] McBride, A. A. and M. C. Ferran. 1996. Repression BPV-1 transcription by the E1 protein is mediated by the E1 C-terminal domain, abstr. 158, p. 137. In Abstracts of the 15th International Papillomavirus Workshop. Brisbane, Australia.

[38] McBride, A. A., H. Romanczuk and P. M. Howley. 1991. The papillomavirus E2 regulatory proteins. *J. Biol. Chem*. **266**:18411–18414.

[39] McShan, G. D. and V. G. Wilson. 1997. Casein kinase II phosphorylates bovine papillomavirus type 1 E1 *in vitro* at a conserved motif. *J. Gen. Virol*. **78**:171–177.

[40] Melendy, T., J. Sedman and A. Stenlund. 1995. Cellular factors required for papillomavirus DNA replication. *J. Virol*. **69**:7857–7867.

[41] Mendoza, R., L. Gandhi and M. R. Botchan. 1995. E1 recognition sequences in the bovine papillomavirus type 1 origin of DNA replication: interaction between half sites of the inverted repeats. *J. Virol*. **69**:3789–3798.

[42] Mohr, I. J., R. Clark, S. Sun, E. J. Androphy, P. MacPherson and M. R. Botchan. 1990. Tageting the E1 replication protein to the papillomavirus origin of replication by complex formation with the E2 transactivator. *Science*. **250**:1694–1699.

[43] Monini, P., L. de Lellis, P. Borgatti, M. Hassan-Omran and E. Cassai. 1993. Activation of eukaryotic transcriptional promoters by the bovine papillomavirus E1-replication factor. *Intervirology*. **36**:245–254.

[44] Moscufo, N. and E. J. Androphy. Unpublished observations.

[45] Muller, F. and M. Sapp. 1996. Domains of the E1 protein of human papillomavirus type 33 involved in binding to the E2 protein. *Virology*. **219**:247–256.

[46] Muller, F., Y. S. Seo and J. Hurwitz. 1994. Replicationof bovine papillomavirus type 1 origin-containing DNA in crude extracts and with purified proteins. *J. Biol. Chem*. **269**:17086–17094.

[47] Park, P., W. Copeland, L. Yang, T. Wang, M. R. Botchan and I. J. Mohr. 1994. The cellular DNA polymerase α-primase is required for papillomavirus DNA replication and associates with the viral E1 helicase. *Proc. Natl. Acad. Sci. USA*. **91**:8700–8704.

[48] Piccini, A., A. Storey, P. Massimi and L. Banks. 1995. Mutations in the human papillomavirus type 16 E2 protein identify multiple regions in binding to E1. *J. Gen. Virol*. **76**:2909–2913.

[48a] Raj, K., and Stanley, M. A. 1995 The ATP-binding and ATPase activities of human papillomavirus type 16 E1 are significantly weakened by the absence of prolines in its ATP-binding domain. *J. Gen. Virol*. **76**:2949–2956.

[49] Romanczuk, H. and P. M. Howley. 1992. Disruption of either the E1 or the E2 regulatory gene of human papillomavirus type 16 increases viral immortalization capacity. *Proc. Natl. Acad. Sci. USA*. **89**:3159–3163.

[50] Sakai, H., T. Yasugi, J. D. Benson, J. J. Dowhanick and P. M. Howley. 1996. Targeted mutagenesis of the human papillomavirus type 16 E2 transactivation domain reveals separable transcriptional activation and DNA replication functions. *J. Virol*. **70**:1602–1611.

[51] Sandler, A. B., S. B. Vande Pol and B. A. Spalholz. 1993. Repression of bovine papillomavirus type 1 transcription by the E1 replication protein. *J. Virol*. **67**:5079–5087.

[52] Santucci, S., E. Androphy, C. Bonne-Andrea and P. Clertant. 1990. Proteins encoded by the bovine papillomavirus E1 open reading frame: expression in heterologous systems and in virally transformed cells. *J. Virol*. **64**:6027–6039.

[53] Santucci, S., C. Bonne-Andrea and P. Clertant. 1995. Bovine papillomavirus ATPase activity does not depend on binf=ding to DNA nor to viral E2 protein. *J. Gen. Virol*. **7640**:1129–1140.

[54] Sarafi, T. R. and A. A. McBride. 1995. Domains of the BPV-1 E1 replication protein required for origin-specific DNA binding and interaction with the E2 transactivator. *Virology*. **211**:385–396.

[55] Schiller, J. T., E. Kleiner, E. J. Androphy, D. R. Lowy and H. Pfister. 1989. Identification of bovine papillomavirus mutants with increased transforming and transcriptional activity. *J. Virol*. **63**:1775–1782.

[56] Sedman, J. and A. Stenlund. 1995. Co-operative interaction between the initiator E1 and the transcriptional activator E2 is required for replicator specific DNA replication of bovine papillomavirus *in vivo* and *in vitro*. *EMBO J*. **14**:6218–6228.

[57] Sedman, J. and A. Stenlund. 1996. The initiator protein E1 binds to the bovine papillomavirus origin of replication as a trimeric ring-like structure. *EMBO J*. **15**:5085–5092.

[58] Sedman, T., J. Sedman and A. Stenlund. 1997. Binding of the E1 and E2 proteins to the origin of replication of bovine papillomavirus. *J. Virol*. **71**:2887–2896.

[59] Seo, Y.-S., F. Muller, M. Lusky, E. Gibbs, H.-Y. Kim, B. Phillips and J. Hurwitz. 1993. Bovine papillomavirus (BPV)-encoded E2 protein enhances binding of E1 protein to the BPV replication origin. *Proc. Natl. Acad. Sci. USA*. **90**:2865–2869.

[60] Seo, Y.-S., F. Muller, M. Lusky and J. Hurwitz. 1993. Bovine papillomavirus (BPV)-encoded E1 protein contains multiple activities required for BPV DNA replication. *Proc. Natl. Acad. Sci. USA*. **90**:702–706.

[61] Spalholz, B. A., A. A. McBride, T. Sarafi and J. Quintero. 1993. Binding of bovine papillomavirus E1 to the origin is not sufficient for DNA replication. *Virology*. **193**:201–212.

[61a] Steger, G., Olszewsky, M., Stockfleth, E. and Pfister, H. 1990. Prevalence of antibodies to human papillomavirus type 8 in human sera. *J. Virol*. **64**:4399–4406.

[62] Storey, A., A. Piccini, P. Massimi, V. Bouvard and L. Banks. 1995. Mutations in the human papillomavirus type 16 E2 protein identify a region of the protein involved in binding to E1 protein. *J. Gen Virol*. **76**:819–826.

[63] Sun, S., L. Thorner, M. Lentz, P. MacPherson and M. Botchan. 1990. Identification of a 68-kilodalton nuclear ATP-binding phosphoprotein encoded by bovoni papillomavirus type 1. *J. Virol*. **64**:5093–5105.

[64] Sverdrup, F. Unpublished observations.

[65] Sverdrup, F. and S. Khan. 1994. Replication of human papillomavirus (HPV) DNAs supported by the HPV type 18 E1 and E2 proteins. *J. Virol*. **68**:505–509.

[66] Sverdrup, F. and S. A. Khan. 1995. Two E2 binding sites alone are sufficient to function as the minimal origin of replication of human papillomavirus (HPV) type 18 DNA. *J. Virol*. **69**:1319–1323.

[67] Thorner, L., N. Bucay, J. Choe and M. Botchan. 1988. The product of the bovine papillomavirus type 1 modulator gene (M) is a phosphoprotein. *J. Virol*. **62**:2474–2482.

[68] Thorner, L. K., D. A. Lim and M. R. Botchan. 1993. DNA-binding domain of bovine papillomavirus type 1 E1 helicase: structural and functional aspects. *J. Virol*. **67**:6000–6014.

[69] Ustav, E., M. Ustav, P. Szymanski and A. Stenlund. 1993. The bovine papillomavirus origin of replication requires a binding site for the E2 transcriptional activator. *Proc. Natl. Acad. Sci. USA*. **90**:898–902.

[70] Ustav, M. and A. Stenlund. 1991. Transient replication of BPV-1 requires two viral polypeptides encoded by the E1 and E2 open reading frames. *EMBO J*. **10**:449–457.

[71] Ustav, M., E. Ustav, P. Szymanski and A. Stenlund. 1991. Identification of the origin of replication of bovine papillomavirus and characterization of the viral origin recognition factor E1. *EMBO J*. **10**:4321–4329.

[72] Vande Pol, S. B. and P. M. Howley. 1995. Negative regulation of the bovine papillomavirus E5, E6, and E7 oncogenes by the viral E1 and E2 genes. *J. Virol*. **69**:395–402.

[73] Wilson, V. G. and J. Ludes-Meyers. 1991. A bovine papillomavirus E1-related protein binds specifically to bovine papillomavirus DNA. *J. Virol*. **65**:5314–5322.

[74] Xiao, X. L. and V. G. Wilson. 1994. Genetically defined nuclear localization signal sequence of bovine papillomavirus E1 protein is necessary and sufficient for the nuclear localization of E1-beta-galactosidase fusion proteins. *J. Gen. Virol*. **75**:2463–2467.

[75] Yang, L., R. Li, I. J. Mohr, R. Clark and M. R. Botchan. 1991. Activation of BPV-1 replication *in vitro* by the transcription factor E2. *Nature*. **353**:628–632.

[76] Yang, L., I. Mohr, E. Fouts, D. A. Lim, M. Nohaile and M. Botchan. 1993. The E1 protein of bovine papillomavirus 1 is an ATP-dependent DNA helicase. *Proc. Natl. Acad. Sci. USA*. **90**:5086–5090.

[77] Yasugi, T., J. D. Benson, H. Sakai, M. Vidal and P. M. Howley. 1997. Mapping and characterization of the interaction domains of human papillomavirus type 16 E1 and E2 proteins. *J. Virol*. **71**:891–899.

[78] Yasugi, T., M. Vidal, H. Sakai, P. M. Howley and J. D. Benson. 1997. Two classes of human papillomavirus type 16 E1 mutants suggest pleiotropic conformational constraints affecting E1 multimerization, E2 interaction, and interaction with cellular proteins. *J. Virol*. **71**:5942–5951.

[79] Zanardi, T. A., M. A. St-Laurent, C. M. Stanley and M. R. Lentz. Phosphorylation of bovine papillomavirus E1 protein and its role in viral DNA replication, abstr. in 14th International Papillomavirus Conference. 1995. Quebec City, Canada:

[80] Zanardi, T. A., C. M. Stanley, B. M. Saville, S. M. Spacek and M. R. Lentz. 1997. Modulation of bovine papillomavirus DNA replication by phosphorylation of the viral E1 protein. *Virology*. **228**:1–10.

[81] Zemlo, T. R., B. Lohrbach and P. F. Lambert. 1994. Role of transcriptional repressors in transformation by bovine papillomavirus type 1. *J. Virol*. **68**:6787–6793.

Appendix A. E1 BLOCKS

Six BLOCKS were recognized by the MOTIF algorithm, the third and fifth of which were also recognized by the Gibbs Sampler algorithm. The first two MOTIF BLOCKS overlap the DNA binding region of E1. The third BLOCK is at the center of the protein where it is presumably involved in several of the principal functions – helicase, replication activity, etc. BLOCKS four through six overlap the ATPase region and are most similar to SV40 large T. A BLAST search using the MOTIF Cobbler sequence (aka HPV31, see below) yielded similarities with very significant scores and probabilities for other E1 protein sequences, with the weakest matches being to the bovine and elk PV E1s. The strongest similarity to a non-E1 protein was to the NS1 protein of the B19 parvovirus (Poisson probability 0.0043), not to SV40 large T (Poisson probability 0.050), but the score was fairly low; this similarity corresponds to the end of BLOCK four and BLOCK 5, E1 regions with ATPase activity. This similarity and the ATPase activity of the parvovirus NS1 have been previously noted by Astell et al., *J. Gen. Virol*. **68**:885–893,1987.

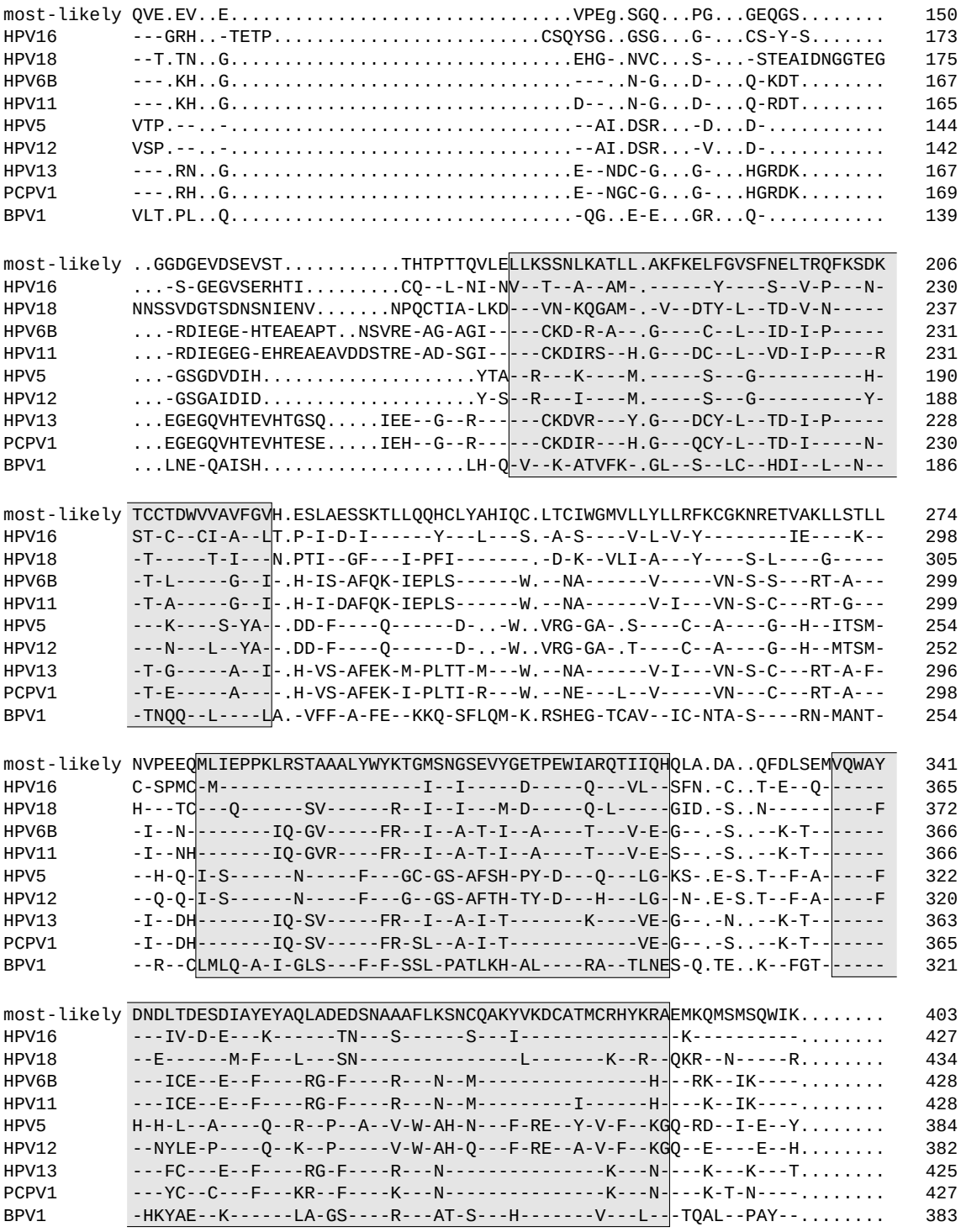
The nucleotide subsequence TAAAACGAAAGT at the 5′ end of the E1 coding sequence is reported to be totally conserved (Campione-Piccardo et al., *Virus Genes* **5**:349–357, 1991). We have found this motif to be perfectly conserved in all human papillomavirus sequences with exception of HPV42 (TAAAACGAAAGC) and HPV48 (TAAAAAGAAAGT). It varies considerably in animal PVs, though the peptide motif appears to be conserved (LKRK at approximately positions 80–83 below).

```
most-likely MAD.............P..KGT..DG......EG.GCSGWFLV.EA.ICS..DkTGDtISEDEDEE..ED     40
HPV16       ---.............-..A-..N-E.......T--N---Y-.--.VVE..K----A--D--N-ND.S-     43
HPV18       ---.............-..E-..--........T--N---Y-.Q-.-VD..K----V--D----NA.T-     42
HPV6B       ---.............D..S-..EN........S--T---M-.--.-VQ..HP--TQ--D-----V.--     42
HPV11       ---.............D..S-..EN........S--T---M-.--.-VE..HT--TQ-----E--V.--     42
HPV5        -T-.............-NS--S..TS......KE.-FGD-C-L.--.D----.V.E.NDLGQLF-..R-     39
HPV12       ---.............S..--S..TS......KE.-L-D-CIL.--.E----.L.E.NDFEQLF-..Q-     37
HPV13       --E.............D..T--..NN........T---------.--.VVE..RT--QQ--D----TV.--     42
PCPV1       ---.............N..T--..-N......K-T---------.--.-VD..R---EE--D----TV.--     42
BPV1        --N.............D..--SNW-S......GL.----.YL-T.--.E-ESDKENEEPGAGV-LSVE.S-     44

most-likely T........GSDLSDFIDDD..ADI.EQG...N.AQELFHQQEAEADEEAVQALKRKYLGSPKAV.....     90
HPV16       -.........-E--V---VN-..N-YLT-AE.TET-HA--TA---KQHRD---V----------..L.SDIS     99
HPV18       -.........---MV----TQ..GTFC--AE.LET--A---A--VHN-AQVLHV----FA-GST.ENSPLG    100
HPV6B       S.........-Y-MV---.--..SN-.THNS.LE.--A--NR---DTHYAT--D---------Y.-.SPIN     96
HPV11       S.........-Y-MV---.---..RH-.T-NS.VE.--A--NR---D-HYAT--D---------Y.-.SPIS     96
HPV5        -.........D--I--LL.--..TEL.-------.SL-------C-QS--QL-K------.---..K.....     85
HPV12       -.........D--V--LL.-N..GEL.-------.SL-------C-QS--QL-I------.--..K.....     83
HPV13       S.........-L-MV---.--..RP-.THNS.VE.--A-LNE---D-HYA---D---------Y.-.SPLG     96
PCPV1       S.........-L-MV---.--..RC-.THNS.LE.--A-LNE---D-HYA---D---------Y.-.SPLG     96
BPV1        RY.......D-QDE--V.-N..-SV.F------.HL-V-QAL-KK-G--QILN----V---SQ.N.S...     94

most-likely ......ENQLSPRLQAISL...SPQKK......AKRRLFE..E.QDSGYGNT......EVEA...ED...    130
HPV16       ...GCVD-NI----K--CIE..KQSRA...............S.E---------.....---T...QQM.L    145
HPV18       E.RLEVDTE------E---N..-G---.............T..I.S-----CS.....------...TQI..    147
HPV6B       TIAEAV-SEI----D--K-T..RQP--.......V-----QTR-LT-----YS.....----G.TGT...    148
HPV11       NVANAV-SEI----D--K-T..TQP--.......V------TR-LT-----YS.....-------.AT...    146
HPV5        ....AVA.------ES---...---Q-.......S-----A..-.----LEL-L.....NN--...-...    127
HPV12       ....AVA.------ES---...---Q-.......S-----A..-.----LELSL.....NN--...-...    125
HPV13       HVEQSVDCDI----D--K-S..RNS--.......V-----QSR-IT-----YS.....--------T...    146
PCPV1       HIEQSV-CDI----N--Q-S..RKP--.......V-----QSR-IT-----H-E....V-------AT...    148
BPV1        ....SGS.EA-ETP....VK..RRKSG.............A.................-N--...NR...    123
```

E1 Proteins


most-likely QVE.EV..E.....VPEg.SGQ...PG...GEQGS..... 150
HPV16 ---GRH..-TETP.....CSQYSG..GSG...G-...CS-Y-S..... 173
HPV18 --T.TN..G.....EHG-.NVC...S-...-STEAIDNGGTEG 175
HPV6B ---.KH..G.....---..N-G...D-...Q-KDT..... 167
HPV11 ---.KH..G.....D-..N-G...D-...Q-RDT..... 165
HPV5 VTP.--..-.....--AI.DSR...-D...D-..... 144
HPV12 VSP.--..-.....--AI.DSR...-V...D-..... 142
HPV13 ---.RN..G.....E--NDC-G...G-...HGRDK..... 167
PCPV1 ---.RH..G.....E--NGC-G...G-...HGRDK..... 169
BPV1 VLT.PL..Q.....-QG..E-E...GR...Q-..... 139
most-likely ..GGDGEVDSEVST.....THTPTTQVLELLKSSNLKATLL.AKFKELFGVSFNELTRQFKSDK 206
HPV16 ...-S-GEGVSERHTI.....CQ--L-NI-NV--T--A--AM-.------Y----S--V-P---N- 230
HPV18 NNSSVDGTSDNSNIENV.....NPQCTIA-LKD---VN-KQGAM-.-V--DTY-L--TD-V-N----- 237
HPV6B ...-RDIEGE-HTEAEAPT..NSVRE-AG-AGI-----CKD-R-A--.G----C--L--ID-I-P----- 231
HPV11 ...-RDIEGEG-EHREAEAVDDSTRE-AD-SGI-----CKDIRS--H.G---DC--L--VD-I-P----R 231
HPV5 ...-GSGDVDIH.....YTA--R---K----M.-----S---G----------H- 190
HPV12 ...-GSGAIDID.....Y-S--R---I----M.-----S---G----------Y- 188
HPV13 ...EGEGQVHTEVHTGSQ.....IEE--G--R-----CKDVR---Y.G---DCY-L--TD-I-P----- 228
PCPV1 ...EGEGQVHTEVHTESE.....IEH--G--R-----CKDIR---H.G---QCY-L--TD-I-----N- 230
BPV1 ...LNE-QAISH.....LH-Q-V--K-ATVFK-.GL--S--LC--HDI--L--N-- 186
most-likely TCCTDWVVAVFGVH.ESLAESSKTLLQQHCLYAHIQC.LTCIWGMVLLYLLRFKCGKNRETVAKLLSTLL 274
HPV16 ST-C--CI-A--LT.P-I-D-I------Y---L---S.-A-S----V-L-V-Y--------IE----K-- 298
HPV18 -T-----T-I---N.PTI--GF---I-PFI-------.-D-K--VLI-A---Y----S-L----G----- 305
HPV6B -T-L-----G--I-.H-IS-AFQK-IEPLS------W.--NA------V-----VN-S-S---RT-A--- 299
HPV11 -T-A-----G--I-.H-I-DAFQK-IEPLS------W.--NA------V-I---VN-S-C---RT-G--- 299
HPV5 ---K----S-YA--.DD-F----Q------D-..-W..VRG-GA-.S----C--A----G--H--ITSM- 254
HPV12 ---N---L--YA--.DD-F----Q------D-..-W..VRG-GA-.T----C--A----G--H--MTSM- 252
HPV13 -T-G-----A--I-.H-VS-AFEK-M-PLTT-M---W.--NA------V-I---VN-S-C---RT-A-F- 296
PCPV1 -T-E-----A---.H-VS-AFEK-I-PLTI-R---W.--NE---L--V-----VN---C---RT-A--- 298
BPV1 -TNQQ--L----LA.-VFF-A-FE--KKQ-SFLQM-K.RSHEG-TCAV--IC-NTA-S----RN-MANT- 254
most-likely NVPEEQMLIEPPKLRSTAAALYWYKTGMSNGSEVYGETPEWIARQTIIQHQLA.DA..QFDLSEMVQWAY 341
HPV16 C-SPMC-M-----I--I-----D-----Q---VL--SFN.-C..T-E--Q-----
HPV18 H---TC---Q------SV------R--I--I---M-D-----Q-L-----GID.-S..N----------F 372
HPV6B -I--N--------IQ-GV-----FR--I--A-T-I--A----T---V-E-G--.-S..--K-T--------
HPV11 -I--NH-------IQ-GVR----FR--I--A-T-I--A----T---V-E-S--.-S..--K-T--------
HPV5 --H-Q-I-S------N-----F---GC-GS-AFSH-PY-D---Q---LG-KS-.E-S.T--F-A-----F
HPV12 --Q-Q-I-S------N-----F---G--GS-AFTH-TY-D---H---LG--N-.E-S.T--F-A-----F
HPV13 -I--DH-------IQ-SV-----FR--I--A-I-T-------K----VE-G--.-N..--K-T--------
PCPV1 -I--DH-------IQ-SV-----FR-SL--A-I-T------------VE-G--.-S..--K-T--------
BPV1 --R--CLMLQ-A-I-GLS---F-F-SSL-PATLKH-AL----RA--TLNES-Q.TE..K--FGT-------
most-likely DNDLTDESDIAYEYAQLADEDSNAAAFLKSNCQAKYVKDCATMCRHYKRAEMKQMSMSQWIK..... 403
HPV16 ---IV-D-E---K------TN---S------S---I----------K----------..... 427
HPV18 --E------M-F---L---SN----------L-------K--R--QKR--N-----R..... 434
HPV6B ---ICE--E--F----RG-F----R---N--M----------H---RK--IK----..... 428
HPV11 ---ICE--E--F----RG-F----R---N--M---------I------H----K--IK----..... 428
HPV5 H-H-L--A----Q--R--P--A--V-W-AH-N---F-RE--Y-V-F--KGQ-RD--I-E--Y..... 384
HPV12 --NYLE-P----Q--K--P-----V-W-AH-Q---F-RE--A-V-F--KGQ--E----E--H..... 382
HPV13 ---FC---E--F----RG-F----R---N----------K---N----K---K---T..... 425
PCPV1 ---YC--C---F---KR--F----K---N----------K---N----K-T-N----..... 427
BPV1 -HKYAE--K------LA-GS----R---AT-S---H-------V---L--TQAL--PAY--..... 383

```
most-likely .HRCD..KVEGE.GDWKPIVKFLRYQG..VEFISFLAALKLFLKGTPKKNCIVIYGPPNTGKSYFCMSLI   467
HPV16       .Y---..R-DDG.----Q--M------..---M---T---R--Q-I------LL--AA-----L-G---M   491
HPV18       .F--S..-IDEG.---R---Q-----Q..I---T--G---S----------L-FC--A-------G--F-   498
HPV6B       .--GS..-I--T.-N-----Q---H-N..I---P--TKF--W-H--------A-V---D-----------   492
HPV11       .Y-GT..--DSV.-N-----Q---H-N..I---P--SK---W-H--------A-V---D----C------   492
HPV5        .TKIN..E-----.-H-SD----I---N..IN--V--T---E--HSV------L--------S---S-A----   448
HPV12       .TKIN..E-----.-H-SD--------D..-N--T----F-N--HAV--H---L--------S---S-A----   446
HPV13       .Y-SK..-I-EA.-N-----Q---H-N..I---P--SK---W-H--------A-V---D----C------   489
PCPV1       .--SK..-IDET.-N-----Q---H-N..I------SK---W-Q--------A-V---D----M------   491
BPV1        .A--K..LAT--.-S--S-LT-FN--N..I-L-T-IN----W---I-----LAFI--------ML-N---   447

most-likely KFLGGKVISFVNSKSHFWLQPLADAKIALLDDATDPCWTYIDTYLRNALDGNPVSIDRKHKALVQIKCPP   537
HPV16       ---Q-S--C------------------GM-----V---N---DN--------L--M-V--RP---L----   561
HPV18       H-IQ-A-------T-----E--T-T-V-M-----TT-----F---M--------I-------P-I-L----   568
HPV6B       S----T---H---S--------V---V-------Q---I-M---M--L-----M---------TL-----   562
HPV11       -----T---Y---C--------T---V-------Q----M---M--L-----M------R--TL-----   562
HPV5        RV-K-R-L-------Q------SEC-------V-----I-M------G---HY--L-C-YR-PT-M-F--   518
HPV12       -V-K-R-L-------Q------GES-------V-----V--------G---HF--L-C-Y--P----F--   516
HPV13       -----T---Y---S--------CN--V-------QS--V-M---M--L-----M-------S-AL-----   559
PCPV1       -----T---Y---S--------CNT-V-------HS--G-M---M--L-----M-------S-AL-----   561
BPV1        H----S-L--A-H------AS---TRA--V----HA--R-F----------Y----------A----A--   517

most-likely LLITSNINVHKDDRYK.YLHSRITVFEFPNPFPFDSNGNPVYELTDQNWKSFFERLWSRLDLSD....QE   602
HPV16       --------AGT-S-WP.---N-LV--T---E----E--------N-K------S-T----S-HE....D-   626
HPV18       I-L-T--HPA--N-WP.--E---------A----K-------IN-K---C----T------HEE...E-   634
HPV6B       --V----DIT-E-K--.---T-V-T-T--------R---A----SNT---C-----S-S--IQ-....S-   627
HPV11       --V----DIS-EEK--.-----V-T-T--------R---A----S-A---C-----S-S--IE-....S-   627
HPV5        --L-------GETN-R.---TT-KG--------MKADNT-QF-----S-----T---TQ----.....--   583
HPV12       --L-------GETN-R.------KG----H---MKPDNT-QFQ----S---------TQ----.....--   581
HPV13       --V---VDIT---K--.--Y--V-TLT--------R---A----S-A---C--T--SAS--IQ-....S-   624
PCPV1       --V----DITTEEK--.--Y--V---K--------R---A----C-A---C--A--SAS--IQ-....S-   626
BPV1        --V----D-QAE---L.-----VQT-R-EQ-CT.-ES-EQPFNI--AD-----V---G----I-E...E-   582

most-likely DE.EE.DGE..SQRTFRCVAGSNNRT..L                                          625
HPV16       -K.-N.--D..-LP--K--S-Q-TN-..-                                          649
HPV18       -A.DT.E-N..PFG--K----Q-H-P..-                                          657
HPV6B       -..--.--S..NSQA----P-TVV--..-                                          649
HPV11       -..--.--S..NSQA----P--VV--..-                                          649
HPV5        E-.G-.---.---A-Q-S-R-A-EH..-                                           606
HPV12       E-.GQ.H--.---A-Q-S-R-A-EH..I                                           604
HPV13       -..-D.--D..NSQA----P-TVV--..V                                          646
PCPV1       -..-D.--D..TSQA----P-TVV--..V                                          648
BPV1        -S.--.--D..-M---T-S-RNT-AV..D                                          605
```

block COBBLER sequence from MOTIF

HPV31 E1

madpagtdgegtgcngwfyveavidrqtgdnisedenedssdtgedmvdfidncnvynnq
aeaetaqalfhaqeaeehaeavqvlkrkyvgsplsdisscvdynisprlkaiciennskt
akrrlfelpdsgygntevetqqmvqveeqqttlscngsdgthserenetptrnilqLLKS
SNMKATLLAKFKELYGVSFNELTRQFKSDKTCCNDWVIAVYGVtgtvaegfktllqpycl
ychlqslacswgmvmlmlvrfkcaknritiekllekllcistncMLIEPPKLRSVAAALY
WYKTSMSNASFTYGETPEWIARQTIINHsfndttfdlsqmVQWAYDNDYTDESDIAYEYA
KLADEDSNAAAFLKSNSQAKYVKDCATMCRHYKRAekrqmsmgqwiksrcdkvsdeGDWK
PIVKFLRYQGVEFISFLSALKQFLHGVPKKNCLlihGPPNTGKSYFCMSLIKFLQGKVIS
FVNSKSHFWLQPLADAKIALLDDATHPCWhYIDTYLRNALDGNPVSIDRKHKAPVQIKCP
PLLITSNINVHKDDKWKYLHSRIvvftfpnpfpfdkngnpvyelsdknwksffsrtwcrl
nlheeedkendgdsfstfkcvsgqnirtl

Appendix B. Structural predictions for the E1 protein.

The MOTIF cobbler sequence for E1 (HPV-31; see appendix A above) was analyzed using several different secondary structure prediction algorithms: Gibrat (Gibrat et al., J.Mol.Biol. 198:425, 1987), Levin (Levin et al., FEBS 205:303, 1986), DPM (Rost and Sander, Prot:Strct.Funct.Gen. 19:55,1994) and SOPM (Geourjon and Deleage, CABIOS 11:681, 1995). The SOPMA service (http://www.ibcp.fr/predict.html) provides these as well as a consensus prediction. Four states, helix (H), beta sheet (E), coil (C), and turn (T) are predicted, of which helices are most certain. Note that some regions of apparent structural conservation are not represented as conserved regions at the primary sequence level, namely in the form of BLOCKS (appendix A). Superimposing the six BLOCKS inferred for the E1 alignment, the first and third (involved in DNA binding) have high helical content.

```
                1       10        20        30        40        50        60
                |       |         |         |         |         |         |
E1-all_motif_co MADPAGTDGEGTGCNGWFYVEAVIDRQTGDNISEDENEDSSDTGEDMVDFIDNCNVYNNQ
Gibrat method   ECCCCCCCCCCCEEECCCEEEEEEHHCCCCCCCCCCCCCCCCCCCCCEEEEECCCCCCCCC
Levin method    HSCHTCCCTTSSCTTSEEEHHHHHHHTCCCCCCCHCCSCCCHCTHHHHHHHHHTCCCCCCTH
DPM method      CCCCCTTTTCTTTTCCEEEEHHEECCCCTCCCTCTCTTTTTTTCCCCHECEECTCTCTCCH
SOPMA predict   CCCCCCCCCCCCCCCCCEEEHHHHHCCCCCCCEECCCCCCCCCCCCHHHHHHHHCCCCCCCH
Consensus       CCCCCCCCCCCCCCTCCEEEHHHHHCCCCCCCCCCCCCCCCCCCCCCHHHHHCCCCCCCCH

                AEAETAQALFHAQEAEEHAEAVQVLKRKYVGSPLSDISSCVDYNISPRLKAICIENNSKT
                HHHHHHHHHHHHHHHHHHHHHHHHHHHHCECCCCCCCEEEEECCCCCCCHEEEEHHHCCCC
                HHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCHCHHHCCCCCCHCCEEEEECCCCHC
                HHHHHHHHHHHHHHHHHHHHHHHHEEHHCTECTCCCCECCEECCCCCCCHHHEEETTTCCC
                HHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCHHCCCCCCCCHEEEEECCCCCC
                HHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCHHCCCCCCCCHEEEEECCCCCC

                AKRRLFELPDSGYGNTEVETQQMVQVEEQQTTLSCNGSDGTHSERENETPTRNILQLLKS
                HHHHEEEECCCCCCCCEEHHHHHEHHCHCCEEEEECCCCCCEEEECCCCCHHHHHHHHHH
                HHHHEEECCTCSCTCCHHCHCHEEEECCSCCCEEECCCTCCCCCCTTCCCCCHHHEEECC
                HHHHHHHTTTTTTCCCCEHHHHHEHHHHHHCCCTTTTTCTCTCCCCCCCCCCCEEHHHTT
                CHHEEEEECCCCCCCCCCCHHHHHHHHCCCCEEEECCCCCCCCCCCCCCCCHHHEEEHCC
                HHHHEEEECCCCCCCCCECHHHHEHHCCCCCCEEECCCCCCCCCCCCCCCCHHHHHHCC

                SNMKATLLAKFKELYGVSFNELTRQFKSDKTCCNDWVIAVYGVTGTVAEGFKTLLQPYCL
                HHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCEEEEEEEECEEEHHCCEECCCCCCH
                THHHHHHHHHHTHCCCCCHHHHHHHCCTTTCCCCHHEEEHHCEETCCHHHCCHHHHHHHH
                CCHHHHHHHHHHHHECECECHHHHHHTTTTCTTCEEEEEEEEECCEHCCHHCCHCCCCE
                CCCHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCHEEEEEEEEEEEHHHHHHHHHHHH
                CCHHHHHHHHHHHHHCHCHHHHHHHCCCCCCCCCCCEEEEEEEEECEHHCCHHCHCCCCH

                YCHLQSLACSWGMVMLMLVRFKCAKNRITIEKLLEKLLCISTNCMLIEPPKLRSVAAALY
                CECCCCCCCCHHHHHEHHHHHHHHCCCHHHHHHHHHHHHHCCCCCCCCCCCHHHHHHHHH
                EECCCCCHHHHHCHHHHHHHHTTCTTCCHHHHHHHHHHHCCTCEEECCCCCCCHHHHHHE
                CECHHCHHTCECEEEEEEEHHHHHCCHEEEHHHHHHHEEEECCEHHCHCCCCHHHHHHHH
                HHHHHHHHHHHHHEEEEEEECCCCCCHHHHHHHHHHHHHHHHHHEEECCCCCHHHHHHHH
                CECCCCCHHCHHHHHEHHHHHHHCCCCHHHHHHHHHHHHHCCCEEECCCCCCHHHHHHHH

                WYKTSMSNASFTYGETPEWIARQTIINHSFNDTTFDLSQMVQWAYDNDYTDESDIAYEYA
                EEHEEECECEEEECCCCHHHHHHEEEECCCCCCCHHHHHHEEEECCCCCCCHHHHHHHHH
                CTCSCCCTTCCCHTCCCCHEHTEEEECTTCCCTCCCCCSCHHHCCCTTCCCHHHHHHHHH
                EECCCCCTCCCCCCCCCHHEHHEEEECCCCTCCCCCECHHEHHHCCTCCCCTCCHHHHHH
                HHHHHCCCCEEEECCCCHHHHHHHEECCCCCCCCCCCHHHHHHHHHCCCCCCCHHHHHHHH
                EECCCCCTCCCCECCCCHHHHHHEEECCCCCCCCCCCHCHHHHHHCCCCCCCHHHHHHHH
```

```
E1-all_motif_co KLADEDSNAAAFLKSNSQAKYVKDCATMCRHYKRAEKRQMSMGQWIKSRCDKVSDEGDWK
Gibrat method   HHHHHHHHHHHHHHHHCHHHCEECHHHHHHHCCCCHHHHHHCCCEEECCCCCCCCCCCHC
Levin method    HCCHHHHHHHHHHHHTCHHHHHHHHHHHHHHHHTSHHHHHCHHHHHHHHTHCHHCHHTTTH
DPM method      HHHTCTCCHHHHHTTCCCHCCEHCHHHHHHHHHHHHHHHHHTHCCEETTCCCCTTTTCCCC
SOPMA predict   HHHHCCCHHHHHHHCCCCHHEEHHHHHHHHHHHHHHHHCCCCHHHHHHHCCCCCCCCCCCC
Consensus       HHHHCHCHHHHHHHHCCCHHCEHCHHHHHHHHHCHHHHHCHHCCHHHHCCCCCCCCCCCC

E1-all_motif_co PIVKFLRYQGVEFISFLSALKQFLHGVPKKNCLLIHGPPNTGKSYFCMSLIKFLQGKVIS
Gibrat method   HHHHHHHCCCHHHHHHHHHHHHHHHCCCCCCEEEECCCCCCCCCEEEEHHHHHCCCEEEE
Levin method    HHHHHHTTCCHHHHHHHHHHHHHHTTCCSCCEEEEECCCCCCCCSEEEEEEECHSHCCEE
DPM method      CEEEEEEECCEHEEEEEHHHHHHHCCTTTCCCEECCTTTTCTCCCEEEEEEHEEECEEEE
SOPMA predict   CEEEHHHHCCCCEHHHHHHHHHHHHHCCCCCCEEEECCCCCCCCCHEEEEHHHHHHCCCEEE
Consensus       CHHHHHH-CCHHHHHHHHHHHHHHHCCCCCCEEEECCCCCCCCCCCEEEHHHHHHCCCEEE

E1-all_motif_co FVNSKSHFWLQPLADAKIALLDDATHPCWHYIDTYLRNALDGNPVSIDRKHKAPVQIKCP
Gibrat method   EEECCHHHHHHHHHHHHHHHHHHHCCCCEEECCHEEECCCCCCCHHHHHHCCCCCHHCCC
Levin method    EECSSSCCCCCCCHHHHHEEETHSSCCCCCCEHHHHCHTCTTCCCEECCCTCCCCEESCC
DPM method      EETTCCCCHHCCHHHHHHHHHCCCCCCCCCCECCCHHCCTTCTCCCCCHCCCHCECCCCC
SOPMA predict   EECCCCHEEHHHHHHHHHEEECCCCCCCCEEEEEEHHHCCCCCCCEEECCCCCCCCCCCC
Consensus       EECCCCCCCHHCCHHHHHHHHHCCCCCCCCCCECHEHHCCCCCCCCCEECCCCCCCCCCCC

E1-all_motif_co PLLITSNINVHKDDKWKYLHSRIVVFTFPNPFPFDKNGNPVYELSDKNWKSFFSRTWCRL
Gibrat method   CCCECCCHHHCHHHHHHHHHCHEEEEECCCCCCCCCCCCCEHEHHCCHHEEEECHHHHEH
Levin method    CCCCCTTHCCCCTCTHCHHTSSEEECSCCSSCCCCTTTCCEEECCTCCHCHHCCHCCCTC
DPM method      CCCECECECETTTTCHCCEEEEEEEEECCCCCCCTTTCTCCCCTTTCCCCCEECEEEEHH
SOPMA predict   CEEEEEEEECCCCCCCHHHHHCEEEEEECCCCCCCCCCCCCEEEECCCCCCHEEHHHHEEE
Consensus       CCCECECHCCCCTCCHCHHHCCEEEEECCCCCCCCCCCCCEEE-CCCCCCHEECHHHEEH

E1-all_motif_co NLHEEEDKENDGDSFSTFKCVSGQNIRTL
Gibrat method   CCCCHCHHHCCCCCCEEEEEECCCCCEEE
Levin method    CCCCHHCTTCSCCCHCHEEECCSSHHCEE
DPM method      CHHHHHTCTTTTTCCTEECEECTCCECCC
SOPMA predict   EEEHCCCCCCCCCCCCEEEEEECCCCCCEH
Consensus       CCCCHCCCTCCCCCCEEEEEECCCCCCEE
```