

# Revealing highly conserved regions in the E6 protein among distinct human papillomavirus types using comparative analysis of multiple sequence alignments

Gabriel, JE. \*, de Figueiredo, DDLG. and de Farias, RP.

Colegiado de Ciências Biológicas, Centro de Ciências Agrárias, Universidade Federal do Vale do São Francisco – UNIVASF, CEP 56300-000, Petrolina, PE, Brazil

\*e-mail: jane.gabriel@univasf.edu.br

Received April 16, 2012 – Accepted May 11, 2012 – Distributed May 31, 2013  
(With 1 figure)

Several epidemiological and molecular studies have confirmed that cervical infection by certain human papillomaviruses (HPV) types is a precursor of the genesis of cervical neoplasia (Gravitt, 2011; Guzmán-Olea et al., 2012). It is yet well elucidated that high risk genotypes lead to the development of cervical cancer and are also associated with other mucosal anogenital, head and neck tumors. The HPV genome is approximately 8 kb in length and is divided into three regions, the non-coding long control region (LCR, ~1 kb), and the protein coding early (E, ~4 kb) and late (L, ~3 kb) regions (Ganguly and Parihar, 2009). The viral genome encodes six early (E1, E2, E4, E5, E6 and E7) and two late (L1 and L2) proteins. HPV *E6* and *E7* genes encode oncoproteins that cause transformation of the host cell, whose action is involved in maintenance of the HPV genome extrachromosomally. Since HPV immortalisation of keratinocytes *in vitro* has been identified as an important step in tumor progression *in vivo*, some studies have provided strong experimental supports for the hypothesis that the maintenance and expression of E6 and E7 proteins observed in human cervical carcinomas has pathologic significance (Hawley-Nelson et al., 1989; Ganguly and Parihar, 2009). The E6 protein in high risk HPV genotypes has been reported to prevent apoptosis by a p53-independent mechanism which involves inhibition of *bax* gene expression and degradation of Bax protein in human keratinocytes, resulting in inhibition of apoptosis and therefore cells accumulate mutations in their DNA (Magal et al., 2005). Recently, Liu et al. (2009) have suggested that the E6 protein mediates telomerase activation by a posttranscriptional mechanism, exerting potential effects on the direct modulation of cell telomerase/telomere function, and triggering a crucial role in both neoplasia and virus replication.

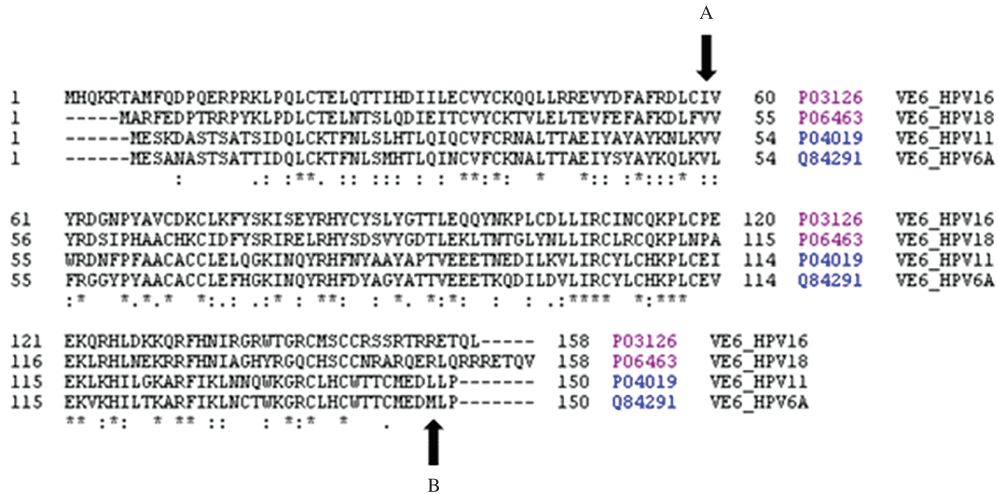
Over the last decades, computational biology tools have opened new insights concerning the characterisation and conservation of biological molecule sequences among distinct viral genomes. Using multiple sequence alignments, the purpose of the present study was to perform comparative analysis of E6 protein segments between high and low risk HPV genotypes. For these analyses, E6 protein segments present in HPV 16 and 18 (high risk and virulent HPV types) and in HPV11 and 06A (low risk and non-virulent HPV types) were selected (Ganguly and Parihar, 2009). Initially, amino acid sequences of E6 protein from these four genotypes were searched using the computational algorithm UniProtKB/Swiss-Prot that provides a high quality annotated and non-redundant protein sequence database (Lesk, 2002; Jungo et al., 2012). Such a search

resulted in the identification of amino acid sequences corresponding to E6 protein previously sequenced and deposited in this database with access numbers: P03126 for HPV16, P06463 for HPV18, P04019 for HPV11 and Q84291 for HPV06A.

The multiple alignments of the amino acid residues in the E6 viral protein among distinct HPV genotypes were generated using the UniProtKB/Swiss Prot database tools (Figure 1), providing relevant statistical parameters, such as similarity and identity. The comparative analysis of the E6 viral protein showed a 56.3% similarity between the HPV16 and 18 types. Additionally, the amino acid alignments between HPV11 and 06A genotypes revealed similarity value corresponding to 81.3%. Multiple alignments of the segment of interest between the four HPV types resulted in a 22.7% similarity, identifying similar and identical positions corresponding to 48 and 37 residues, respectively, in a total amino acid amount of 158 (Figure 1).

Using the basic local alignment search tools (BLAST), further analysis provided additional statistical parameters, such as expectation values (e-values) and score bits. E-values in blast results represent the probability of the alignment has occurred by chance, whereas the value of score bits depends on the size of alignment, the number of matches/mismatches/gaps and matrix used for the comparison of sequences and is normalised by means of variable statistics (Lesk, 2002; Healy, 2007). Blast analysis generated significant hits with E-values corresponding to  $1 e^{-61}$ , maximum identity of 57% and 178 score bits between HPV16 and 18 types, whereas significant alignments were observed between HPV11 and 06A with E-values corresponding to  $1 e^{-95}$ , maximum identity of 81% and 264 score bits (data not shown). According to Lesk (2002), E-values close to zero indicate that the match of E6 proteins is significant in short sequences and therefore, the findings presented in this study demonstrate accentuated similarities in the alignments of the amino acid residues in the E6 protein between high and low risk HPV types.

Based on the substitutions that really do or do not occur in real protein sequences, the ability of a protein to tolerate those replacements is related to the chemical properties of the amino acids in the pair. As observed in Figure 1 (arrows), replacements of amino acid residues with identical chemical properties (for example, of valine (V) for isoleucine (I) between HPV18 and 16, of methionine (M) for leucine (L) between HPV11 and 06A) might indicate conservative changes throughout E6 viral protein. Such conservative changes could represent replacements of both amino acids with identical chemical proprieties (for



**Figure 1.** Multiple alignments of the amino acid residues in the E6 protein sequences between high and low risk HPV types. Columns indicating identical amino acids in the same position in all types analysed are exclusively signalled by an asterisk (\*), whereas alignments showing the presence of similar amino acids are represented by other symbols (: or .). Arrows A and B indicate replacements of amino acids resulting in conservative changes.

example, the substitutions evaluated in this study occurred among hydrophobic, nonpolar amino acids). Interestingly, high conservative changes were identified in blast analysis between HPV11 and 06A (91%) throughout sequence of E6 protein, as well as accentuated conservation of 71% were also detected between HPV16 and 18 types. In biology computational analysis, similar positions indicate that an amino acid may be substituted for another with similar chemical properties, which is more likely to observe, or for one with different chemical characteristics. In the regions where an amino acid has no correspondence in the other in the same position, it would be interpreted as the occurrence of an insertion or deletion of one residue in the peptide sequence.

Several studies have showed a major role of the E6 protein in the induction and maintenance of cellular transformation, acting mainly as an oncoprotein by stimulating the destruction of many host cell key regulatory proteins, culminating in the cervical cancer symptoms (Liu et al., 2009). Consistent evidence suggests that the genome of papillomaviruses is static, and changes in its sequence by mutation or recombination are very rare events (Gravitt, 2011; Guzmán-Olea et al., 2012). Different genera of HPVs share less than 60% similarity in the nucleotide sequence of the major capsid of the protein L1 ORF (Leto et al., 2011). According to Leto et al. (2011), different virus species within the same genus share about 60% to 70% similarity. Thus, the findings presented in this study, demonstrating the presence of highly conserved regions with significant similarity of the amino acid residues in the E6 protein among distinct HPV types, might reveal stronger evidence regarding the molecular characterisation and evolutionary conservation of viral genomes using comparative analysis of multiple sequence alignments.

**References**

GANGULY, N. and PARIHAR, SP., 2009. Human papillomavirus E6 and E7 oncoproteins as risk factors for tumorigenesis. *Journal of Biosciences*, vol. 34, p. 113-123. PMID:19430123. <http://dx.doi.org/10.1007/s12038-009-0013-7>

GRAVITT, PE., 2011. The known unknowns of HPV natural history. *Journal of Clinical Investigation*, vol. 121, p. 4593-4599. PMID:22133884 PMCid:3225991. <http://dx.doi.org/10.1172/JCI57149>

GUZMÁN-OLEA, E., BERMÚDEZ-MORALES, VH., PERALTA-ZARAGOZA, O., TORRES-POVEDA, K. and MADRID-MARINA, V., 2012. Molecular mechanism and potential targets for blocking HPV-induced lesion development. *Journal of Oncology*, vol. 2012, p. 1-11. PMID:22220169 PMCid:3246776. <http://dx.doi.org/10.1155/2012/278312>

HAWLEY-NELSON, P., VOUSDEN, KH., HUBBERT, NL., LOWY, DR. and SCHILLER, JT., 1989. HPV16 E6 and E7 proteins cooperate to immortalize human foreskin keratinocytes. *The EMBO Journal*, vol. 8, p. 3905-3910. PMID:2555178 PMCid:402081.

HEALY, MD., 2007. Using BLAST for performing sequence alignment. *Current Protocols in Human Genetics*, vol. 52, p. 6.8.1-6.8.23.

JUNGO, F., BOUGUELERET, L., XENARIOS, I. and POUX, S., 2012. The UniProtKB/Swiss-Prot Tox-Prot program: a central hub of integrated venom protein data. *Toxicon*, vol. 60, p. 230-243. PMID:22465017. <http://dx.doi.org/10.1016/j.toxicon.2012.03.010>

LESK, AM., 2002. *Introduction to Bioinformatics*. London: Oxford University Press. 283 p.

LETO, MGP., SANTOS JUNIOR, GF., PORRO, AM. and TOMIMORI, J., 2011. Human papillomavirus infection: etiopathogenesis, molecular biology and clinical manifestations. *Anais Brasileiros de Dermatologia*, vol. 86, p. 306-317. PMID:21603814. <http://dx.doi.org/10.1590/S0365-05962011000200014>

LIU, X., DAKIC, A., ZHANG, Y., DAI, Y., CHEN, R. and SCHLEGEL, R., 2009. HPV E6 protein interacts physically and functionally with the cellular telomerase complex. *Proceedings of the National Academy of Sciences USA*, vol. 106, p. 18780-18785. PMID:19843693 PMCid:2773972. <http://dx.doi.org/10.1073/pnas.0906357106>

MAGAL, SS., JACKMAN, A., ISH-SHALOM, S., BOTZER, LE., GONEN, P., SCHLEGEL, R. and SHERMAN, L., 2005. Downregulation of Bax mRNA expression and protein stability by the E6 protein of human papillomavirus 16. *Journal of General Virology*, vol. 86, p. 611-621. PMID:15722521. <http://dx.doi.org/10.1099/vir.0.80453-0>