Proteomics and Sequence Analysis for HPV type 16 E7 Protein Expressed in Cervical Cancer

Upasna Srivastava[1], Satendra Singh[1], Gurmit Singh[2] and George Thomas[2]

[1]Department of Computational Biology & Bioinformatics*,
[2]Department of Computer Science & Information Technology, Jacob School of Biotechnology and Bioengineering, SHIATS, Allahabad - 211007, India.
Department of Computational Biology & Bioinformatics, Sam Higginbottom Institute of Agriculture, Technology & Sciences Deemed to be University, Allahabad - 211007, India.

Email: princy149@rediffmail.com

ABSTRACT

Proteomics is a dynamic property closely related to the conformational mechanisms of protein structure in its physiological environment. To understand and control the function of target proteins, it becomes increasingly important to develop methods and tools for predicting collective motions at the molecular level. In this paper various Bioinformatics tools are used for the primary and secondary structure analysis for HPV Type 16 E7 protein expressed in cervical cancer. ProtPram computed parameters include molecular weight (11.02 KD), theoretical pI (4.20), Protein Isoelectric Point as pH 3.97and Aliphatic index as 78.57 and Grand average of hydropathicity (GRAVY): -0.405. PSORT Signal score as (-10.24), cytoplasmic protein score (0.650), mitochondrial matrix space (0.100), lysosome lumen score (0.100), Endoplasmic reticulum (0.00) and Possible cleavage site was (61).GOR-4 predict secondary structure of HPV type 16 E7 protein in row as H=helix, S= extended 'or' Beta and C=coil and give the probable value for each amino acid. I-Mutant2.0 is used gives prediction of protein stability changes upon mutation. T-COFFEE was employed to compare the sequences which predict the key residues responsible for catalytic activity of the enzyme, amino acid sequence of HPV TYPE 16 E7 proteins.

Keywords: HPV, ProtPram, PSORT, GOR-4, T-COFFEE, Proteomics

INTRODUCTION

Human Papillomaviruses (HPVs) form a large group (>80 types) of epitheliotropic double-stranded DNA tumour viruses. A subset of HPVs, referred to as 'high-risk' HPVs, have been associated with the development of human anogenital tumours such as cervical cancers in women and penile cancers in men [1]. HPV-16 and HPV-18 are the most frequently identified HPV types found in genital tumours [2] and are present in 40–60% and 10–20% of cervical carcinomas [3, 4, 5] and in 20–50% and 3–15%

of penile cancers, respectively [6, 7, 8]. Prevalence of HPV infection in men may be similar to that observed among women. The natural evolution of genital HPV infection in men, however, is less well understood. Up to 70% of male partners of women with cervical HPV infections and intraepithelial Neoplasia are diagnosed with subclinical HPV infections or HPV-associated diseases [9, 10]. The incidence of HPV-associated penile cancer, however, is much lower, suggesting that additional events and /or cofactors are involved in the malignant evolution of HPV infection in men. For this, the Protein identification is a basic task for proteomics research. Usually, the proteins/ peptides need to be separated according to their physicochemical properties before the identification [11]. The common used physicochemical properties include isoelectric point, hydrophobicity and mass weight. So, the calculation of these physicochemical properties is very important and could be used for the design of proteomic experiments. Further analyses of physicochemical properties for the identified proteins might be helpful to discover the experimental bias, which would be important for improving the experiments. The analyses could also be used to discover the different distribution caused by some biological factors, for example the different isoelectric point and hydrophobicity in different sub-cellular localization [12, 13]. HPV belong to the Papovaviridae family. They consist of 72-capsomere capsid containing the viral genome. Capsomers are composed of two structural proteins: the 57 KD late proteins L1, which accounts for 80% of the viral particle, and the 43-53 KD Minor capsid protein L2 [14].

MATERIALS AND METHODS

HPV 16 E 7 Protein sequence retrieval and analysis

Progress in sequencing the genome of HPV 16 provides considerable recourses that may be used for the *In-Silico* analysis of its Proteome. Numerous databases are now available which contain both sequence and functional information. Most of these are accessible information for use on local servers. The complete genome sequence of HPV type 16 in the form of the nucleotide and amino acid were retrieved from FTP server of NCBI-GenBank (http://www.ncbi.nlm. nih.gov/GenBank/) shown in Figure [1].

The 98 amino acid long sequence of HPV type 16 E7 protein was obtained from the Protein sequence database of NCBI (NCBI Gene Id: 1489079 and NCBI Protein id: 9627105) and blasted against Protein Data Bank (PDB) entries to find similar sequences. Parameter values for BLAST 2.2.26 search were set as default. It was available at (http://blast.ncbi.nlm.nih.gov/Blast). It gives gene position from 561 to 857 in whole genome sequence of HPV 16 with taxonomic ID (333760), Gene Bank Accession (AY089955) and Protein Accession (NP_041326). The genomic data retrieved from FTP server was extensively used for this study.

Primary Structure Analysis for HPV Type 16 E7 Protein

Physiochemical properties such as molecular weight, theoretical pI, total number of negatively (Asp+Glu) and positively (Arg+Lys) charged residues, extinction coefficients, instability index, aliphatic index and grand average of hydropathicity

(GRAVY) of the mature protein were computed using Expasy's ProtPram Proteomics server (http://www.expasy.org/tools/protparam.html) [15]**.** Peptide Cutter was used for the cleavage of the chosen enzymes and chemicals mapped onto the HPV 16 E7 protein sequence and it was available at (http://us.expasy.org/tools/peptidecutter/). The sub cellular localization was predicted by Psort II (http://psort.hgc.jp/cgi-bin/runpsort.pl) [16]**.** NetGlyc 1.0 Server predicted an N-Glycosylation site using artificial neural networks. It is available at http://www.cbs.dtu.dk/services/NetNGlyc/ [17].

Secondary structure Analysis for HPV type 16 E7 Protein

The protein topology prediction was done by using SOPMA (Self Optimized Prediction Method with Alignment) **[18]** and percentage of amino acids was determined by GOR-4 tool The GOR V server is freely accessible to public users and private institutions at http://gor.bb.iastate.edu/ [19].

Multiple Sequence Alignment

In order to know the key residues responsible for catalytic activity of the enzyme and to identify the conserved region of amino acid sequences of HPV TYPE 16 E7 protein was compared with amino acid sequences of known selected crystallographic structures. T-Coffee was used for multiple sequence alignment which is the combination of local and global information. Such combination was probably necessary for computing high-quality alignments [20]**.** It was available from both http://www.tcoffee.org and its main mirror http://tcoffee.crg.cat.

Phylogenetic Analysis of Hpv 16 E7

In this study, the Phylogenetic tree was constructed based on the alignment of HPV 16 E 7 amino-acid sequence with the selected 24 HPVs, using both parsimony and distance matrix analyses. The Phylogenetic tree was generated with help of Phylip programs (http://www.atgc-montpellier.fr/phyml/) based on nearest-neighbour joining method. In This BLOSUM-62 was selected as protein weighted matrix for performing multiple sequence alignment. After this corresponding Phylip files were then subjected to Phylogenetic tree construction showing boot strap values at nodes as all the strains are closely related to each other and there are strong sequence similarities among themselves.

Prediction of families and domain in HPV 16 E7 and Human TMEM 50 A

The set of paralogous proteins in human Papilloma virus type 16 E7 was predicted by using the Pfam search. The Pfam is a large collection of protein domain families. In which each Domain found as functional region of proteins and family was also represented by multiple sequence alignment and Hidden Markov models (HMMs). The paralogous protein dataset was submitted at Pfam server which predicted the protein families at the default parameter. These are available at (http://pfam.sanger.ac.uk/) [21]**.** Pfam also generates higher-level groupings of related families, known as ***clans***. A clan is a collection of Pfam-A entries which are related by similarity of sequence, structure or profile-HMM. Beside this for prediction of

conserved domain in both protein sequences of HPV 16 E 7 and Human TMEM 50 A conserved domain retrieval tool (CDART) was used. It is available at http://www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi.

Virtual screening for the HPV 16 E 7 Proteins

Virtual screening was carried out for Human Papilloma Virus 16 E 7 against the NCI subset two molecules retrieved from ZINC database. The virtual screening was performed by the FINDSITE^COMB *(cssb.biology.gatech.edu/FINDSITE-COMB), were total 74,378 molecules from the NCI diversity subset two (*http://zinc.docking.org/index.html*) were screened for the HPV 16 E7 protein.*

RESULTS AND DISSCUSION

HPV 16 E 7 Protein sequence retrieval and analysis

The complete genome sequence of the human Papilloma virus obtained from FTP server of NCBI represents various HPVs strains out of which HPV 16 E 7 is expressed in cervical carcinomas diseases. In proteome analysis of HPV 16 E7 ProtPram gives molecular weight (11.02 kD), theoretical pI (4.20), Protein Isoelectric Point as pH 3.97 Total number of negatively charged residues (Asp + Glu) was 19 , Total number of Positively charged residues (Arg + Lys) was five ,Ext. coefficient was 6335, The estimated half-life is: 30 hours (Mammalian reticulocytes, *in vitro*)>20 hours (yeast, *in vivo*) ,Total number of atoms found as 1499 and Instability index second was computed to be 63.00. This classified the protein HPV type 16 E7 as stable and it also measured aliphatic index as 78.57 and Grand average of hydropathicity (GRAVY) as 0.405. In this Peptide Cutter Predicted potential protease cleavage sites and Enzymes in HPV Type 16 E 7 protein sequence are shown in Table 1. Asp-N endopeptidase + N-terminal Glu, Pepsin and Thermolysin enzymes show more position of cleavage sites while Proline-endopeptidase gives only one cleavage site at position 98 of amino acid sequence of HPV 16 E7 protein [22].

PSORT gives prediction of Protein sorting signals and localization sites in amino acid sequence of HPV type 16 E7 proteins. It predicted Signal score as (-10.24), cytoplasmic protein score (0.650), mitochondrial matrix space (0.100), lysosome lumen score (0.100), Endoplasmic reticulum (0.00) and Possible cleavage site was (61). HPV 16 E7 protein sequence has no N-terminal signal sequence. NetGlyc 1.0 Server predict HPV16 E7 without signal peptides are unlikely to be exposed to the N-glycosylation machinery and thus may not be glycosylated (*in vivo*) even though they contain potential motifs (at Threshold=0.5). Peptide Mass displaying peptides with a mass bigger than 500 Dalton. Using monoisotopic masses of the occurring amino acid residues and giving peptide masses shown as in Table 2.

SOPMA predicted the percentages of amino acid composition as shown in Table 3. It predicted the presence of alpha helix 28.57%; extended strand 21.43% and random coils 50% in the secondary structure of HPV 16 E 7 protein while 310 helix, Pi helix, Beta Bridge, Beta turn and Bend regions were absent.

I-Mutant2.0 predicted changes in HPV 16 E7 protein sequence after mutation. It is a

neural-network-based web server for the automatic prediction of protein stability changes upon single-site mutations. Proline was found as constant amino acid in wild type protein, which was converted into other different types of proteins after a single site mutation using DDG as energy score. The whole mutation was carried out at pH 7 and 25 degree Celsius temperature Table 4.

GOR-4predicts secondary structure of HPV type 16 E7 protein in row as H=helix, S=extended 'or' Beta and C=coil and give the probable value for each amino acid. In this helix was shown in red colour while extended sheets and coils were shown in blue and black colour Figure 2.

Table 1: Predicted potential protease cleavage sites and Enzymes in a HPV Type 16 E 7 protein sequence by Peptide Cutter.

S.No.	Name of enzymes	No. of cleavage	Position of cleavage site
1.	Arg-C proteinase	3	49, 66, 77
2.	Asp-N endopeptidase	10	3 ,13, 20 ,29 ,3,5 38, 47 61, 74,80
3.	Asp-N endopeptidase + N-terminal Glu	19	3 ,9 ,13, 17, 20, 25, 29 32, 33 ,34, 35 36, 38 45, 47, 61 ,74 ,79, 80
4.	CNBr	3	1 ,12 ,84
5.	Glut amyl endopeptidase	9	10, 18 ,26 ,33, 34 ,35 37, 46, 80
6.	Lys C	2	60,97
7.	Lys N	2	59,96
8.	NTCB (2-nitro-5-thiocyano-benzoic acid)	7	23, 57, 58, 60, 67, 90 93
9.	Pepsin (pH1.3)	26	8, 10, 12, 13 ,14, 21 ,22 22, 23, 24, 25 27, 28 52, 56, 57, 64, 65, 67 78 ,81 82, 83 ,86 87
10.	Proline-endopeptidase	1	98
11.	Staphylococcal peptidase I	7	10, 18 ,26 ,33 ,37 ,46 80
12.	Thermolysin	21	7, 11, 12, 27 ,41, 44, 49 53, 54, 5664, 66, 68 73, 78, 82, 83 ,86 ,88 89 92
13.	Trypsin	4	4, 49, 60, 66, 77

Table 2: Peptide Mass predict Chain Protein E7 at positions 1 - 98 [Theoretical pI: 4.20 / Mw (average mass): 11022.32 / Mw (monoisotopic mass): 11015.02] with peptide mass and peptide sequence.

Mass	Position	Peptide Sequence
5597.3737	1-49	MHGDTPTLHEYMLDLQPETT DLYCYEQLNDSSEEEDEIDG PAGQAEPDR
2231.1484	78-98	TLEDLLMGTLGIVCPICSQK P
1298.6020	50-60	AHYNIVTFCCK
1270.6572	67-77	LCVQSTHVDIR
694.3188	61-66	CDSTLR

Table 3: Secondary structure elements as predicted by SOPMA for HPV type 16 E7 Protein.

S.No.	Secondary structural element	Percentage of structural elements
1.	Alpha helix (Hh)	28.57%
2.	310 helix (Gg)	0.00%
3.	Pi helix (Ii)	0.00%
4.	Beta bridge (Bb)	0.00%
5.	Extended strand (Ee)	21.43%
6.	Beta turn (Tt)	0.00%
7.	Bend region (Ss)	0.00%
8.	Random coil (Cc)	50.00%
9.	Ambiguous states (?)	0.00%

Table 4: I-Mutant v2.0 Prediction of Protein Stability Changes upon Mutations for HPV 16 E7 protein.

Position	Amino acid in wild type protein	New amino acid after mutation	DDG: DG (New Protein)-DG (Wild Type) in Kcal/mol	pH	Temperature in Celsius degrees
49	P	V	-0.65	7	25
49	P	L	-0.68	7	25
49	P	I	-0.21	7	25
49	P	M	-0.66	7	25
49	P	F	-0.06	7	25
49	P	W	-1.25	7	25
49	P	Y	-1.11	7	25
49	P	G	-1.34	7	25
49	P	A	-0.76	7	25
49	P	S	-1.41	7	25
49	P	T	-1.62	7	25
49	P	C	-1.11	7	25
49	P	H	-1.45	7	25
49	P	R	-0.48	7	25
49	P	K	-0.66	7	25
49	P	Q	-1.47	7	25
49	P	E	-0.84	7	25
49	P	N	-1.90	7	25
49	P	D	-1.89	7	25

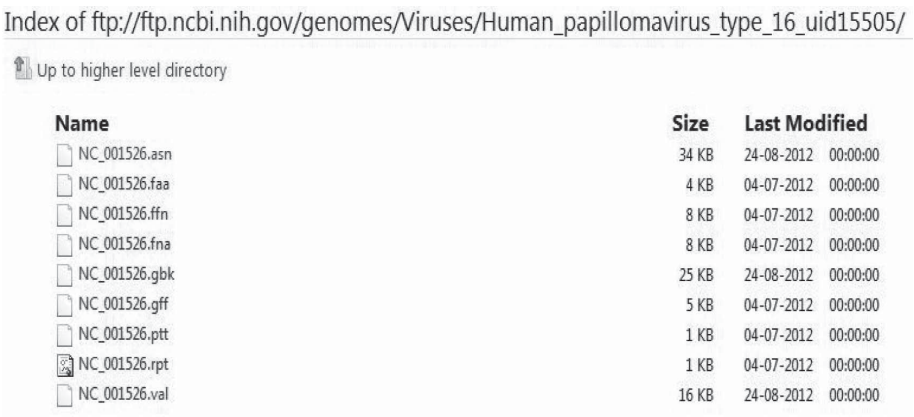

Fig. 1: FTP Directory links for Retrieval of sequence data in FASTA format

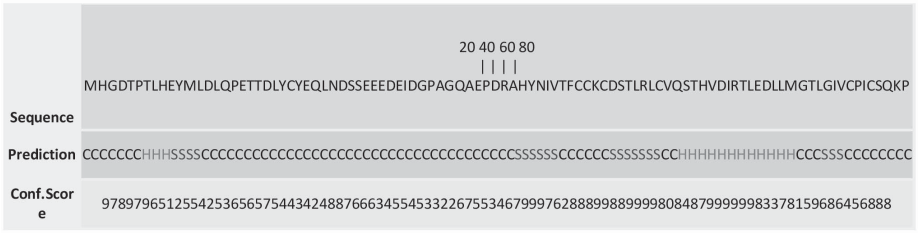

Fig. 2: Predicted secondary structure of HPV type 16 E7 protein in row as H=helix, S= extended 'or' Beta and C=coil by GOR-4.

T-COFFEE analysis

T-Coffee multiple sequence alignment of HPV type 16 E7 protein sequence with other related crystallographic protein sequences depicted a high degree of conservation among the sequences. Hence, the compared sequences varied in length but essentially conserved. The key catalytic residues have been highlighted as shown in Figure 3. The red region shows more conserved region between the sequences, while blue or green show bad and orange or yellow give average conserved regions. The list of mutants which are selected for multiple sequence alignment and Phylogenetic analysis are shown as in Table 5. Where in VE7_HPV16, Q8JNA0_9PAPI, C8YJL3_9PAPI, E0YDP5_9PAPI, HPV 101, Q8BE75_9PAPI Protein, Q77GV7_HPV16 Protein E7 and Q77AI1_HPV16 Protein E7 show the same protein sequence length Figure 3.

Table 5: List of selected HPV strains for multiple alignments and Phylogenetic Analysis

S.N.	Accession No.	Protein name	Sequence length	E-value
1.	P03129	VE7_HPV16	98	4.0×10-69
2.	P27230	VE7_HPV35	99	1e-38
3.	P36831	VE7_HPV52	99	2e-20
4.	Q8JNA0	Q8JNA0_9PAPI	98	7e-20
5.	P36828	VE7_HPV34	97	7e-20
6.	P22161	VE7_RHPY1	113	7e-19
7.	P36826	VE7_HPV30	105	5e-18
8.	Q02271	VE7_HPV13	101	4e-17
9.	Q81998	VE7_HPV72	100	1e-12
10.	P54668	VE7_HPV68	110	8e-13
11.	P36824	VE7_HPV26	104	8e-10
12.	P50780	VE7_HPV22	100	3e-07
13.	Q07858|	VE7_HPV63	88	4e-07
14.	C8YJL3	C8YJL3_9PAPI	98	5e-04
15.	P06932	VE7_HPV05	103	3e-06
16.	E0YDP5	E0YDP5_9PAPI	98	9e-06
17.	Q1AHS1	Q1AHS1_9PAPI(HPV 101)	98	9e-06
18.	G3DRD8	G3DRD8_9PAPIProteinE7 (canine Papilloma virus 8)	104	4e-05
19.	Q8BE75	Q8BE75_9PAPI Protein E7 (Zetapapillomavirus1)	98	3e-04
20.	Q80928	VE7_HPV50	93	5e-04
21.	G1CR69	BPV3 Protein E7	90	0.010
22.	H6UYR3	H6UYR3_9PAPI Protein E7	116	0.18
23.	Q77GV7	Q77GV7_HPV16 Protein E7	98	4.0×10-69
24.	Q77AI1	Q77AI1_HPV16 Protein(Fragment)	98	4.0×10-69
25.	Q778I3	Q778I3_HPV16 Protein E7 (Fragment)	98	4.0×10-69

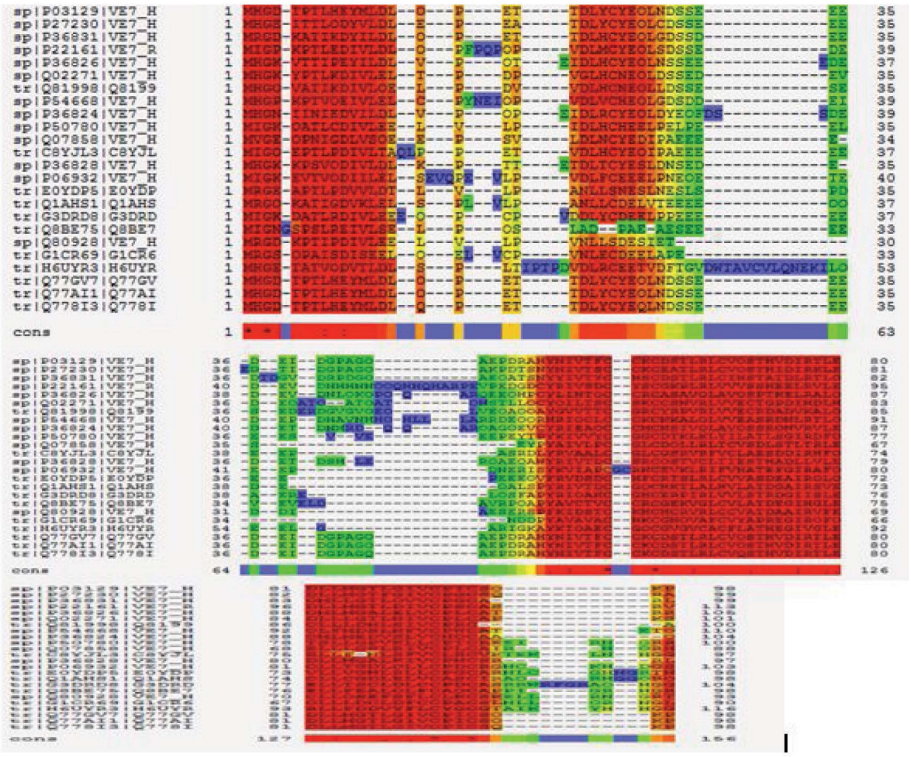

Fig. 3: T-Coffee Ver. 9.03 Multiple sequence alignment of the HPV Type 16 E 7 Protein with respect to another selected HPVs strain protein sequences.

Phylogenetic Tree analysis of HPV 16 E7s

In the Phylogenetic tree topology obtained on E7 proteins analysis of selected 25 HPVs strains reveals some sort of divergence among different strains multiple sequence alignment results among different strains suggest some variations in amino acids Figure 4. Besides this, it also suggests that some conserved residues among divergent lineages of the proteins may not be a random process but instead create mechanisms which lead to specific carcinomas. The phylogeny tree shows bootstrap values in which bootstrapping can be considered a two-step process comprising the generation of (many) new data sets from the original set and the computation of a number that gives the proportion of times that a particular branch (e.g., a taxon) appears in the tree. Each of the newly created data sets has the same number of total positions as the original data set, but some positions are duplicated or triplicated.

Predicted families, Domains and Motifs in Human Papilloma Virus type 16 E7

The HPV 16 E7 polypeptide is biologically active and possesses at least two functional domains; the first induces cellular DNA synthesis in quiescent rodent cells and the second Trans activates the adenovirus E1A-inducible early E2 promoter and binds

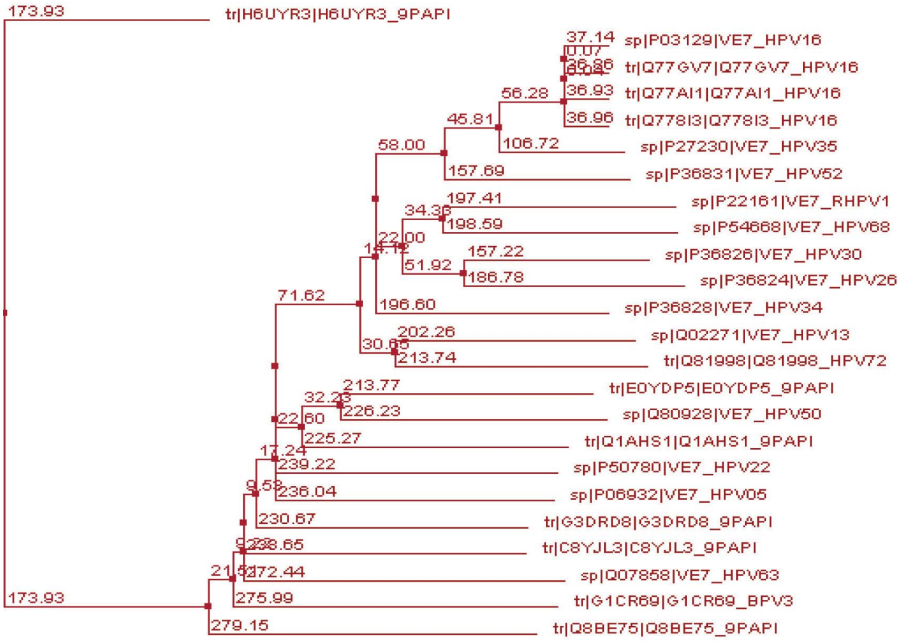

Fig. 4: Phylogenetic tree of Human Papilloma Virus 16 E 7 constructed from alignment of 25 HPVs strains using Phylip program showing boot straps values.

Zinc. Further, each domain is autonomous and can function on separate peptides. DNA synthesis induction activity maps within the N-terminal portion of the molecule, which contains sequences related to adenovirus E1A conserved domains 1 and 2 required for cell transformation and binding of the retinoblastoma gene product. Trans-Activation and Zn-binding activities map within the C-terminal portion of the molecule, a region which contains Cys-X-X-Cys motifs. The predicted Pfam families of human Papilloma Virus 16 E 7 are shown in Table 6 with their molecular functions.

Besides this, in human Papilloma virus 16 E 7 two different motifs were also identified and both are represented by Pfam id E 7 and Adeno_E1A shown in Table 7. In which Adeno_E1 is a family of adenovirus early E1A proteins. The E1A protein is 32 KD it can however be cleaved to yield the 28 KD protein. The E1A protein is responsible for the transcriptional activation of the early genes within the viral genome at the start of the infection process as well as some cellular genes [23]. In this motif position of E 7 early protein was predicted in between 3 to 95 with 5.9 e-30 E-value while Early E1A protein shows motif position between 8 to 41 with 0.0059 E-value. After motif prediction, Conserved Domain Architecture Retrieval Tool was used for identification of conserved domains in HPV 16 E7. It predicted mainly four super families of human Papilloma virus as in the form of their clans form which are as shows in Table 8.

Table 6: Predicted Pfam families in HPV16 E7 protein

S.No.	Protein	GO no.	Reference DB	Reference family	E-value	Description
1.	sp|P03129| VE7_HPV16	GO:0003677	PFAM	PF00527	2.8e-29	DNA binding
2.	sp|P03129| VE7_HPV16	GO:0003700	PFAM	PF00527	2.8e-29	transcription factor activity
3.	sp|P03129| VE7_HPV16	GO:0006355	PFAM	PF00527	2.8e-29	Regulation of transcription, DNA-dependent
4.	sp|P03129| VE7_HPV16	GO:0005622	PFAM	PF00527	2.8e-29	intracellular

Table 7: Predicted motif in HPV 16 E 7 sequence

S.No	Pfam ID	Description	Motif Position	Predicted motif	E-value
1.	E 7	E7 Protein, Early protein	3…95	GDTPTLHEYMLDLQPETTDLYC YEQLNDSSEEEDEIDGPAGQAEP DRAHYNIVTFCCKCDSTLRLCV QSTHVDIRTLEDLLMGTLGIVCPICS	5.9e-30
2.	Adeno_E1A	Early E1A protein	8…41	LHEYMLDLQPETTDL YCYEQLNDSSEEEDEIDGP	0.0059

Table 8: Predicted conserved Domains in HPV 16 E7 using Conserved Domain Architecture Retrieval Tool.

S.No	Super Family	Clans	Proteins
1.	Early protein	Cl02891	E7 Protein
2.	Y2_Tnp Super family	Cl04874	Putative Transposes
3.	PHA02778 Super family	Cl14423	Major capsid L Protein
4.	V MSA Super family	Cl02933	Major surface antigen

Virtual screening for validated model structure of HPV 16 E 7 Protein

Virtual screening was carried out for Human Papilloma virus against the NCI subset two of the ZINC database. The top ten molecules having minimum energy were screened out as the possible inhibitor of the HPV 16 E 7 protein Table 9. All selected molecules have very less energy score. In these molecules ZINC17108357 has minimum energy score and highest number of hydrogen bond donor and H-bond acceptor so it was found as the possible inhibitor lead molecule against HPV 16 E 7 protein. Although ZINC07191846, ZINC20911247, ZINC06574930 and ZINC04350123 have no H-bond donor so they represent a poor interaction with protein molecules.

Table 9: Identification of Molecules against HPV 16 E7 Protein from ZINC database after virtual Screening.

S.No.	ZINC ID of the screened molecules	Energy score Kcal/Mol	No. of H-bond donor	No. of H-bond Acceptor	Molecularl Weightg/mo	X log P	Polar desolvation Kcal/Mol	Apolar desolvation Kcal/Mol
1.	ZINC17108357	0.771468	6	9	340.35	2.92	-12.31	-11.19
2.	ZINC08764110	0.759846	3	5	295.342	2.89	-10.98	-0.49
3.	ZINC07191846	0.738258	0	3	277.323	3.29	-14.95	3.23
4.	ZINC04744390	0.736614	2	5	312.387	2.93	-10.66	-4.18
5.	ZINC20911247	0.734044	0	5	305.374	1.70	-14.68	5.68
6.	ZINC06574930	0.728408	0	4	281.311	3.18	-7.45	0.1
7.	ZINC05004073	0.723281	4	5	224.278	1.33	-9.58	-9.26
8.	ZINC04350123	0.723009	0	1	370.15	4.76	-3.56	-2.22
9.	ZINC01746223	0.721276	1	2	230.331	3.49	-29.82	-0.3
10.	ZINC03881190	0.719124	2	3	210.236	3.04	-10.95	4.21

CONCLUSION

The aim of this present work was to describe, proteomics of HPV Type 16 E7 virus protein by using bioinformatics tools. In proteome analysis of HPV 16 E7 ProtPram gives molecular weight (11.02 kD), theoretical pI (4.20), Protein Isoelectric Point as pH 3.97, Total number of negatively charged residues (Asp + Glu) was 19, Total number of Positively charged residues (Arg + Lys) was five, Ext. coefficient was 6335, The estimated half-life is: 30 hours, Total number of atoms found as 1499 and Instability index second was Computed To be 63.00. This classified the protein HPV type 16 E7 as unstable and it also measured aliphatic index as 78.57 and Grand average of hydropathicity (GRAVY) as 0.405. Psort predict Signal score in HPV16 E7 as (-10.24), cytoplasmic protein score (0.650), mitochondrial matrix space (0.100), lysosome lumen score (0.100), Endoplasmic reticulum (0.00) and Possible cleavage site was (61). GOR-4predicts secondary structure of HPV type 16 E7 protein in row as H=helix, S=extended 'or' Beta and C=coil and gave the probable value for each amino acid as alpha helix 28.57%, Extended strand 21.43% and random coils are found as 50% inside the protein sequence. T-Coffee multiple sequence alignment of HPV type 16 E7 protein sequence with another related crystallographic protein sequences depicted a high degree of conservation among the sequences. In this study, the Phylogenetic tree was constructed based on the alignment of HPV 16 E 7 amino-acid sequence with the selected 24 HPVs, using both parsimony and distance matrix analyses. The Pfam database search has identified significant number of paralogous proteins which were categorized in Pfam families, domains, repeats and clans. Such identification is helpful for functional annotation of HPV 16 E 7 proteins.

ACKNOWLEDGEMENT

The authors are deeply grateful to Sam Higginbotom Institute of Agriculture, Technology and Sciences for providing the necessary guidance, facilities and support required for completing the research work.

REFERENCES

1. Zur Hausen, H. 1991. Human Papillomaviruses in the pathogenesis of anogenital Cancer. *Virology.* **184**: 9-13.
2. Wagatsuma, M., Hashimoto, K., Matsukura, T. 1990. Analysis of integrated human papillomavirus type 16 DNA in cervical cancers: amplification of viral sequences together with cellular flanking sequences. *Journal of Virology* **64**, 813-821.
3. Riou, G., Favre, M., Jeannel, D., Bourthis, J., Le Doussal, V., Orth, G. 1990. Association between poor prognosis in early-stage invasive cervical carcinomas and non-detection of HPV DNA. *Lancet* **335**: 1171-1174.
4. Van den Brule, A. J., Walboomers, J. M., Du Maine, M., Kenemans, P., Meijer, C. J. L. M. 1991. Difference in prevalence of human papillomavirus genotypes in cyto-morphologically normal cervical smears is associated with a history of cervical intraepithelial Neoplasia. *International Journal of Cancer* **48**, 404-408.
5. Higgins, G. D., Phillips, G. E., Smith, L. A., Uzelin, D. M., Burrel, C. J. 1992**.** High prevalence of human papillomavirus transcripts in all grades of cervical intraepithelial glandular Neoplasia. *Cancer.* **70**: 136-146.
6. McCone, D. J., Kalache, A., Ashdown, K. X., Andrade, L., Menzes, F., Smith, P. Doll, R. 1986. Human papillomavirus type 16 and 18 in carcinomas of the penis from Brazil. *International Journal of Cancer.* **37**:55-59.

7. Tornesello, M. L., Buonaguro, F. M., Meglio, A., Buonaguro, L., Beth-Giraldo, E. Giraldo, G. 1997. Sequence variations and viral genomic state of human papillomavirus type 16 in penile carcinomas from Ugandan patients. *Journal of General Virology.* **78**: 2199-2208.
8. Wiener, J. S., Effert, P. J., Humphrey, P. A., Yu, L., Liu, E. T., Walther, P. J. 1992. Prevalence of human papillomavirus types 16 and 18 in squamous-cell carcinoma of the penis: a retrospective analysis of primary and metastatic lesions by differential polymerase chain reaction. *International Journal of Cancer.* **50**: 694-701.
9. Barrasso, R., de Brux, J., Croissant, O., Orth, G. 1987. High prevalence of papillomavirus-associated penile intraepithelial neoplasia in sexual partners of women with cervical intraepithelial neoplasia. *New England Journal of Medicine.* **317**:916-923.
10. Scheffner, M., Huibregtse, J. M., Vierstra, R. D., Howley, P. M. 1993. The HPV-16 E6 and E6AP complex functions as a Ubiquitin protein lipase in the ubiquitination of p53.*Cell.* **75**: 495-505.
11. Ly, L., Wasinger, V. C. 2011. Protein and peptide fractionation, enrichment and depletion: tools for the complex proteome. *Proteomics*. **11**:513-34.
12. Chan, P. and Warwicker, J. 2009. Evidence for the adaptation of protein pH-dependence to sub cellular pH
13. Zhang, T., Chou, K. C. 2006. Prediction of protein sub cellular location using hydrophobic patterns of amino acid sequence. *Computational Biology Chem.* **30**:367-71.
14. Baker, C.C., Phelps, W.C., Lindgren, V., Braun, M.J, Gonda, M.A., Howley, P. M. 1987. Structural and transcriptional analysis of human papillomavirus type 16 sequences in cervical carcinoma cell lines. *Journal of Virology*. **61**:962-967.
15. Wilkins, M. R., Lindsborg, I., Gasteiger, E., Bairoch, A., Sanchez, J.C., Hochstrasser, D.F., Appel, R. D. 1997. ***Detailed peptide characterisation using PEPTIDEMASS - a World-Wide Web accessible tool.*** *Electrophoresis.* **18**:403-408.
16. Gasteiger, E., Hoagland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A. 2005. ***Protein Identification and Analysis Tools on the ExPASy Server.*** *The Proteomics Protocols Handbook, Humana Press*.
17. Nakai, K., Horton, P. 1999. PSORT: a program for detecting sorting signals in proteins and predicting their sub cellular localization. *Trends Biochemical Sciences.* 24:34-36
18. Gupta, R., Jung, E., Brunak, S. 2004. Prediction of N-glycosylation sites in human proteins. *IN Preparation*.
19. Geourjon, C., Deleage, G. 1995. SOPMA: Significant improvement in protein secondary structure prediction by prediction from alignments and joint prediction. *CABIOS*. **11**:681-684.
20. Sen, T. Z., Jernigan, R.L., Garnier, J., Kloczkowski, A. 2005. GOR V server for protein secondary structure prediction. *Bioinformatics* **21**:2787-2788.
21. Notredame, C., Higgins, D. G., Heringa, J. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology.* **302**:205-17.
22. Capriotti, E., Fariselli, P., Casadio, R. 2005. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Resources* **33**: 306-310.
23. Altschul, S. F., Thomas, L., Madden, A. A. S., Jinghui, Z., Zheng, Z., Webb, M., David, J. L. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Resources* **25**:3389-3402.
24. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., Bairoch, A. 2005. Protein Identification and Analysis Tools on the ExPASy Server. In: John M. Walker (Ed.), the Proteomics Protocols Handbook, Humana Press, 571-607.
25. Nakai, K., Horton, P. 1999. PSORT: a program for detecting sorting signals in proteins and predicting their sub cellular localization. *Trends Biochemical Science.* **24**:34-36.

26. Gupta, E. J., Brunak, S. 2004. Prediction of N-glycosylation sites in human proteins. *Pacific Symposium on Bio computing*. **7**:310-322.
27. Geourjon, C., Deleage, G. 1995. SOPMA: Significant improvement in protein secondary structure prediction by prediction from alignments and joint prediction. *CABIOS*. **11:**681-684.
28. Garnier, J.F., Gibrat, B., Robson, B. 1996. GOR secondary structure prediction method Version IV. *Methods in Enzymology.* **266:** 540-553.
29. McClure, M. A., Vasi, T. K. F. 1994**.** Comparative analysis of multiple protein-sequence alignment methods. *Molecular Biology Evolution*. **11**:571-592.
30. Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunesekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E. L., Eddy, S. R., Bateman, A. 2010. The Pfam Protein families' database. *Nucleic Acid Research Database Issue.* **38**: 211-222.
31. Srivastava, U., Singh, G. and Thomas, G. 2013. In-silico structural analysis and Protein disorder prediction: Implications for structural proteomics of human Papilloma virus type-16 E7 Protein. *International Journal of Biotechnology and Bioengineering Research*. **4**:63-72.
32. Webster, L. C., Zhang, K., Chance, B., Ayene, I., Culp, J. S., Huang, W. J., Wu, F. Y. and Ricciardi, R. P. 1991. Conversion of the E1A Cys4 zinc finger to a non-functional His2, Cys2 zinc finger by a single point mutation. *Protocol of National Academic Sciences USA*. **88**: 9989-9993.