

Research article

Prediction of Human proteins having interactions with HPV16 proteins using virus and host sequence motifs.

Rabbiah Manzoor Malik^{1,2*}, Sahar Fazal², Syed Touqeer Abbas¹, Aamer Bhatti², Mukhtar Ullah³, Kamal M. A.

1. Wah Medical College, Wah Cantt.
2. Capital University of Science and Technology, Islamabad, Pakistan.
3. FAST University of Emerging Sciences and Technology, Islamabad, Pakistan.

Corresponding author's information:

*rabbiahmanzoor@gmail.com

Citation: Malik RM.; Fazal S.; Abbas ST., Bhatti A. Prediction of Human proteins having interactions with HPV16 proteins using virus and host sequence motifs . *Pakistan Journal of Biochemistry and Biotechnology*, 2021, 2 (2), 107-126.

<https://doi.org/10.52700/pjbb.v2i2.54>

Received: 01-09-2021

Accepted: 04-12-2021

Published: 31-12-2021

Abstract: Background: Human Papillomavirus (HPV) infection has been found to be the major cause of cancer of cervical region, in females. Genome of HPV codes for 6 functional proteins E1, E2, E4, E5, E6 and E7. These proteins play different roles in development of HPV infection and its progression towards cervical cancer. The interactions of HPV proteins with human DNA and proteins occurs in the presence of short linear peptide motifs on these proteins, have similar sequence to those found on proteins in human cells. **Methods:** After identification of human motifs in HPV proteins, by use of ELM resource, their counter domains were found from PROSITE. The proteins of human proteome containing these counter domains were predicted as the proteins having possibility of interactions with HPV proteins. **Results:** we predicted 9468 human proteins for having interactions with HPV proteins. Our predicted proteins were enriched with the host proteins having possibility of being interacted by HPV proteins. 10% of our predicted proteins were already reported to be affected by one or more HPV proteins. The list of predicated proteins can be utilized to find out the connectivity between the virus HPV and human host. It can also be used to determine the pathways involved in pathogenesis of HPV leading towards the cervical cancer. **Conclusion:** The list of predicated proteins can be utilized to find out the connectivity between the virus HPV and human host. It can also be used to determine the pathways involved in pathogenesis of HPV leading towards the cervical cancer.

Keywords: HPV Proteins, HPV16, cervical cancer, Protein interactions, motifs, domains

1. Introduction

Proteins are the structural organizations of life and every procedure in the cell is directed by mind boggling communications of proteins [1, 2]. These macromolecules are built

from 20 different subunits called amino acids. Such sub-atomic combinations are being turned out as an imposing development unit for nature and, despite the fact that being made out of just these couple of fixings, proteins are gifted by an amazing practical assorted quality over the span of advancement for all targets and purposes each procedure in the cell [1, 3, 4]. Proteome of viral particles play a major role in causing infection. Due to smaller genome and a smaller number of proteins expressed, viruses use cellular machinery of host to complete its life cycle. Interaction of viral proteins with that of host cells are the key players of malignant transformation of viral particles.

Viral attack mechanisms can be understood in more comprehensive way by analyzing Virus-host Protein-protein interactions (VHPPIs). Host metabolic mechanisms and pathways are modulated or disrupted by viruses with the help of VHPPIs. Most of these PPIs operate on the basis of mimicry: a viral protein mimics a host protein and interacts with that host protein's binding partners. The similarity of structure or sequence of protein makes this mimicry possible [5]. These interactions are usually transient, easily formed, and disrupted, yet specific. Many of these transient interactions involve the binding of a protein domain to a short stretch (3–10) of amino acid residues, which can be characterized by a sequence pattern, i.e., a eukaryotic linear motif (ELM) *a short linear motif (SLiM)* [6]. Our present study is focused on the prediction of ELMs or SLiMs involved in this mechanism of VHPPIs. Prediction of ELM or SLiM-mediated VHPPI, could help in elucidating viral carcinogenesis mechanisms, designing vaccines and antiviral drugs, by the researchers and drug developers. [7–10].

An ELM or a SLiM is a small region of a chain of protein, ranging in length from 3 to 12 amino acid residues. They have been found to be involved in marking of proteolytic cleavage sites, controlling of assembling mechanisms of protein molecules, tagging of localization of proteins localization and recruiting of enzyme [11, 12]. These are compact structures and contribute in low-affinity protein interactions [13, 14]. SLiMs or ELMs, present in proteins of eukaryotes, are curated in the database ELM [15, 16].

Mutations, insertions or deletions can evolve or change SLiMs rapidly in the viral proteins [17]. The newly evolved ELMs or SLiMs can alter the VHPPI networks by generating new valuable PPIs that ultimately can change the cell cycle [18], create new protein multiplexes by mediating new conformational changes in proteins [19]. More insights into the protein interactions reveal that VHPPI are stabilized and enriched by ELMs and SLiMs domains binding interfaces [20].

Host cell cycle is hijacked, in order to execute the viral cycle, by viruses through mimicking the host ELMs or SLiMs and using the VHPPIs [21]. This strategy has been shown by researchers in a study revealing the interactions of oncoprotein E6 of human

papilloma virus (HPV) with 14-3-3 proteins and hDlg protein through the ELMs and SLiMS interacting with the PDZ domain, found in these human proteins [22].

Presence of more than 30000 proteins in humans is a hurdle in determination of VHPPIs, experimentally. Very few proteins of viral proteome have been modelled three dimensionally and used in studies for discovering VHPPIs. These reasons made the idea behind our present study for predicting the human proteins which could be interacted by HPV proteins, on the basis of VHPPIs mediated by viral motifs identical to human motifs and domains of human proteins. An inexpensive substitute to experimental prediction of VHPPIs or guidance to an experimental strategy can be achieved by a bioinformatic method to predict VHPPIs mediated by SLiMS.

Computationally, ELMs or SLiMs are characterized by regular expressions. A subsequence of a protein matching the regular expression represents an instance of an ELM or a SLiM. For example, if an ELM or a SLiM is represented as R.[RK]R by the expression, it could have several possible instances like in Ebola virus it would be expressed as RVRRE [23] and will be expressed as RKRRF in Human respiratory syncytial virus A2 [24]. The algorithm we used for prediction of SLiMs and human proteins is based on mapping of viral protein sequences on human ELMs, and finding occurrences of ELMs or SLiMs in viral proteins, already present in host proteins. There should be some criteria to filter viral instances of ELMs or SLiMs, in order to the probability of predicting real interactions. The possibility of a SLiM or ELM from virus increases to interact with host protein, if it is conserved in the small genome of virus. In a study by Evans et al. Common SLiMs or ELMs between humans and HIV were found to be significantly conserved in proteome of HIV [25].

HPV genome codes for two types of proteins, early genes code for 6 proteins, E1, E2, E4, E5, E6 and E7, while late genes code for L1 and L2.

1. Materials and Methods:

Table 1: Amino Acid sequences of HPV proteins.

Protein	ID No.	Amino Acids Sequence/ Number
E1	P03114	MADPAGTNGEETGTCNGWFYVEAVVEKKTGDAISDDENENDSDTGEDLVDFIV NDNDYLTQAETETAHALFTAQEAKQHRDAVQVLKRKYLVSPLSDISGCVDNNIS PRLKAICIEKQSRAAKRRLFESDSGYGNTEVETQQMLQVEGRHETETPCSQYSG GSGGGCSQYSSGSGGEGVSRHTICQTPLTNILNVLKTSNAKAAMLAKFKELYGV SFSELVRPFKSNKSTCCDWCIAAFGLTPSIADSIKTLQYCLYLHIQSLACSWG M VLLLVRKYKCGKNRETIEKLLSKLLCVSPMCMMEPPKLRSTAAALYWKYKTGISNI SEVYGDTPEWIQRQTVLQHSFNDCTFELSQM VQWAYDNDIVDDSEIAYKYAQL

		ADTNSNASAFKLSNSQAKIVKDCATMCRHYKRAEKKQMSMSQWIKYRCDRVD DGGDWKQIVMFLRYQGVEFMSFLTALKRFLQGIPKKNCILLYGAANTGKSLFGM SLMKFLQGSVICFVNSKSHFWLQPLADAKIGMLDDATVPCWNYIDDNLRLNALD GNLVSMDEVKHRPLVQLKCPPLLITSNINAGTDSRWPYLHNRLVVFTFPNEFPFDE NGNPVYELNDKNWKSFFSRTWSRSLHEDEDKENDGDSLPTFKCVSGQNTNTL (649)
E2	P03120	METLCQRLNVCQDKILTHYENDSTDLRDHIDYWKHMRLCAIYYKAREMGFKH INHQVVPTLAVSKNKALQAIQLTLETIYNSQYSNEKWTLQDVSLEVYLTAPTG CIKKHGYTVEVQFDGDICNTMHYTNWTHIYICEEASVTVVEGQVDYYGLYVHE GIRTYFVQFKDDAEKYSKNKVWEVHAGGQVILCPTS VFSSNEVSSPEIIRQHLAN HPAATHTKAVALGTEETQTTIQRPRSEPDTGNPCHTTKLLHRDSVDSAPILTA FN SSHKGRINCNSNTTPIVHLKGDANTLKCLRYRFKHKCTLYTAVSSTWHWTGHN VKHKS AIVTLTYDSEWQRDQFLSQVKIPKTITVSTGFMSI (365)
E4	P06922	MADPAAATKYPLLKLLGSTWPTTPRPIPKPSWAPKKHRRSSDQDQSQTPE TP ATPLSCCTETQWTVLQSSLHLTAHTKDGLTVIVTLHP (92)
E5	P06927	MTNLDTASTLLACFLLCFCVLLCVCLLIRPLLLSVSTYTS LIILVLLWITAASA FR CFIVYIIFVYIPLFLIH THARFLIT (83)
E6	P03126	MHQKRTAMFQDPQERPRKLPQLCTELQTTIHDIILECVYCKQQLLRREVDFAFR DLCIVYRDGNPYAVCDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCI NCQKPLCPEEKQRHLDDKQRFHNIRGRWTGRCMSSCRSSRTRRETQL (158)
E7	P03129	MHGDTPTLHEYMLDLQPETTDLYCYEQLNDSSEEEDEIDGPAGQAEPDRAHYNI VTFCKCDSTLRLCVQSTHVDIRTLEDLLMGTLGIVCPICSQKP (98)

2.1. Amino acid Sequences of HPV proteins

The amino acids sequences of HPV proteins were downloaded from UniProtKB (version 2018) [26], as shown in Table 1.

2.2. Virus protein ELM annotation and conservation

FASTA sequence of E1, E2, E4, E5, E6 and E7 were annotated with sequences of motifs of human by using server ELMs using the ELM resource [27], using default settings except selecting human for the species field.

2.3. Human Protein ELM and CD Annotation

The ELM resource lists Counter Domains (CDs) or proteins known to interact with ELMs. For each ELM conserved on HPV proteins, we found the appropriate CDs. and mapped them to PROSITE domains. When the ELM resource listed a set of interacting proteins instead of CDs, we assumed that all proteins had a common unknown CD, and annotated them with that. We constructed a list of CDs and interacting proteins for each HPV protein conserved ELM.

2.4 Prediction of human proteins interacting with HPV proteins

The prediction of HHP, the set of human proteins that might interact with HPV proteins, was based on interactions mediated by ELMs and CDs. By using PROSITE we found the human proteins reported to be containing those CDs.

2.5. Validation using the HPV Databases

2211 human proteins have been retrieved from different databases of Human Papilloma Virus which have been reported to interact with HPV proteins. These databases include PaVE, HPV Sequence Database, HPVDB, NCBI database, VirusHost database and HPRD, which we called HHE, and used to investigate the usefulness of HHP. We restricted the human proteins interacting with HPV proteins to those belonging to the set of proteins that have PROSITE domains. The statistics in this research focused on the comparison of our predicted set HHP and the experimental dataset HHE based on the overlap between the two sets. P-values for the overlap between HHP and HHE and their various subsets were calculated using the hypergeometric test in the R Project for Statistical Computing.

3. Results

3.1. Eukaryotic Linear Motifs (ELMs)

ELMs or short linear motifs (SLiMs), are compact sites on proteins, involved in interaction, formed by short continuous sequences of amino acids. They are enriched with intrinsically flexible regions of the proteome and provide a broad range of functionality to the functional proteins. They play important roles in regulation of cellular activities, and are also clinically important, because exceptional SLiM functions have been found to be associated with many diseases and SLiM imitations are often used by pathogens to manipulate cellular machinery of their hosts [28, 29]. Table 2 shows the conserved ELMs for each HPV protein, which were obtained by annotation of amino acid sequence of HPV proteins on 133 peptide motifs (ELMs) by using the ELM Resource, which was accessed March 2018. Overall, 77 of these ELMs present in the ELM resource were conserved on HPV proteins. Some of the ELMs, like LIG_PDZ_Class_1 and LIG_Rb_LxCxE_1 [30, 31] have been experimentally proved as active binding sites for human proteins with HPV proteins E6 and E7, respectively. E1 was having the highest number (28) of ELMs, contrary to E5 having only 1 motif.

Table 2: Names of ELMs conserved on HPV proteins. Green color shows the presence of ELM on the respective HPV protein.

S. No.	ELMs	E1	E2	E4	E5	E6	E7
1.	CLV_C14_Caspase3-7						
2.	CLV_NRD_NRD_1						
3.	CLV_PCSK_FUR_1						
4.	CLV_PCSK_PC1ET2_1						
5.	CLV_PCSK_PC7_1						
6.	CLV_PCSK_SKI1_1						
7.	CLV_Separin_Metazoa						
8.	DEG_APCC_KENBOX_2						
9.	DEG_Kelch_Keap1_1						
10.	DEG_Nend_UBRbox_2						
11.	DEG-MDM2_SWIB_1						
12.	DOC_CKS1_1						
13.	DOC_CYCLIN_1						
14.	DOC_CYCLIN_RxL_1						
15.	DOC_MAPK_FxFP_2						
16.	DOC_MAPK_gen_1						
17.	DOC_MAPK_MEF2A_6						
18.	DOC_PP1_RVXF_1						
19.	DOC_PP2A_B56_1						
20.	DOC_PP2B-LxvP_1						
21.	DOC_USP7_MATH_1						
22.	DOC_USP7_UBL2_3						
23.	DOC_WW_Pin1_4						
24.	LIG_14-3-3_CanoR_1						
25.	LIG_Actin_WH2_2						
26.	LIG_APCC_ABBA_1						
27.	LIG_BIR_II_1						
28.	LIG_BIR_III_1						
29.	LIG_Clathr_ClatBox_1						
30.	LIG_eIF4E_1						
31.	LIG_FHA_1						
32.	LIG_FHA_2						
33.	LIG_LIR_Gen_1						
34.	LIG_MYND_2						
35.	LIG_NRBOX						
36.	LIG_PDZ_Class_1						
37.	LIG_Pex14_1						

38	LIG_Pex14_2						
39	LIG_PTB_Apo_2						
40	LIG_PTB_Phospho_1						
41	LIG_Rb_LxCxE_1						
42	LIG_SH2_GRB2						
43	LIG_SH2_SRC						
44	LIG_SH2_STAT3						
45	LIG_SH2_STAT5						
46	LIG_SH3_1						
47	LIG_SH3_2						
48	LIG_SH3_3						
49	LIG_SUMO_SIM_anti_1						
50	LIG_SUMO_SIM_par_1						
51	LIG_TRAF2_1						
52	LIG_TRFH_1						
53	LIG_TYR_ITIM						
54	LIG_TYR_ITSM						
55	LIG_UBA3_1						
56	MOD_CDK_SPK_2						
57	MOD_CDK_SPxK_1						
58	MOD_CDK_SPxxK_3						
59	MOD_CK1_1						
60	MOD_CK2_1						
61	MOD_GlcNHglycan						
62	MOD_GSK3_1						
63	MOD_LATS_1						
64	MOD_NEK2_1						
65	MOD_NEK2_2						
66	MOD_N-GLC_1						
67	MOD_N-GLC_2						
68	MOD_PIKK_1						
69	MOD_PKA_1						
70	MOD_PKA_2						
71	MOD_PLK						
72	MOD_ProDKin_1						
73	MOD_SUMO_rev_2						
74	TRG_ENDOCYTIC_2						
75	TRG_ER_diArg_1						
76	TRG-NLS_MonoExtC_3						
77	TRG-NLS_MonoExtC_4						
78							

ELMs of Proteins interact with their specific counter domains (CDs) present in other proteins. CDs are larger stretches of conserved sequences of amino acids on proteins. Binding of ELMs with CDs play major role in stabilizing the protein-protein interactions. 46 counter domains were found in human proteins which could interact with these ELMs, enabling HPV proteins to interact with human proteins on the basis of presence of their respective ELMs on them [Table 3]. On the basis of presence of these domains in human proteins, over all 12598 proteins were predicted to have possibility of interactions with HPV proteins. Individual number of proteins predicted for HPV proteins E1, E2, E4, E5, E6 and E7 are given in table 4. 45% of these proteins were common for having interactions with two or more HPV proteins. After considering these common proteins having occurrence of 1 time, 9468 proteins were found to have interactions with HPV proteins. Domain PKinase has been found on more than 3000 human proteins, while only 9 proteins have shown the presence of domain Adap_comp_sub. These proteins were grouped as Human Predicted Proteins (HPP), containing human proteins which might have interactions with proteins of HPV. This prediction was based on the mechanism of protein-protein interactions mediated through ELMs and their CDs. FHA and PKinase domains show interaction with ELMs of all HPV proteins.

Table 3: Counter Domains of ELMs found on HPV proteins, The blue color indicates the possibility of interaction of respective HPV protein with CD.

S. No.	CDs	E1	E2	E4	E5	E6	E7
1.	14-3-3						
2.	Actin						
3.	Adap_comp_sub						
4.	Arm						
5.	Armadillo type fold						
6.	Atg 8						
7.	BIR						
8.	Branch						
9.	CKS						
10.	Clathrin_propel						
11.	Cyclin_N						
12.	FHA						
13.	Hormone_recep						
14.	ICAP-1_int_bdg						

15	IF4E						
16	IRS						
17	Kelch_1						
18	MATH						
19	Metallophos						
20	EF_hand_7						
21	PDZ						
22	Peptidase_M16						
23	Peptidase_S8						
24	Peptidase_C14						
25	Peptidase_C50						
26	PI3_PI4_kinase						
27	PID						
28	Pkinase						
29	Pkinase_C						
30	PNP_UDP_1						
31	PTB						
32	Pex14_N						
33	Rad60-SLD						
34	RB_B						
35	SH2						
36	SH3_1						
37	SH3_9						
38	STT3						
39	SWIB						
40	TRF						
41	UBA3						
42	UQ_con						
43	USP7_ICP0_bdg						
44	WD40						
45	WW						
46	Zf_MYND						
47	Zf-UBR						

3. 2. Prediction of human proteins having interactions with HPV proteins

This work has been verified by finding significant overlap between predicted proteins and validated proteins, which are experimentally proven for interacting with HPV proteins. All the experimentally reported proteins having interactions with HPV proteins were grouped into a Human Experimented Proteins (HEP). For every HPV protein, significance of the overlap has been evaluated between human proteins and HPV proteins. In Table 4. The HPP column shows the number of human proteins which have been predicted in this work for having interaction with individual HPV proteins, while HEP are the number of humans proteins which have been verified experimentally as proteins targeted by HPV in the PaVE, HPV Sequence Database, HPVDB, NCBI database, VirusHost database and HPRD. The lists of HEP proteins is provided in Appendix A-7. The column of Overlap gives the counts of the number of proteins present in both sets. P-values for this overlap between validated and predicted proteins were calculated by using a hypergeometric test in R. p-value gives the significance of results. P-value<0.01 gives statistically highly significant results.

Table 4 : Overlap between HEP and HPP with p-values for all individual HPV proteins.

HPV protein	HPP	HEP	Overlap	p-value
E1	4006	853	150	1.74E-02
E2	395	534	151	1.38E-02
E4	1230	437	51	4.19E-02
E5	280	790	30	1.42E-19
E6	3657	688	222	1.27E-11
E7	3390	877	204	1.63E-02

E2 becomes functional after some post translational modifications. These include phosphorylation, acetylation and Sumoylation. Phosphorylation has been observed majorly at amino acid 253, 298, and 301, while minor sites of phosphorylation are 36, 76, 235 and 286 [32].

4.2.2.1. Motifs on E2:

9 ELMs have been found involving these amino acids (Table 5), which experience phosphorylation in order to maintain an interaction with proteins as well as to perform their functions. This validates the role of these specific amino acids in phosphorylation of E2.

Table 5: ELMs found in the phosphorylated regions of E2.

ELMs	Amino acids Sequence/ Position
LIG_FHA_1	252-258 (HTTKLLH)
MOD_PKA_2	258-264 (HRDSVDS)
DOC_MAPK_MEF2A_6	68-77 (KALQAIQL)
MOD_PIKK_1	231-237 (TEETQTT)
DOC_WW_Pin1_4	283-288 (SNTTPI)
LIG_FHA_1	283-289 (SNTTPIV), 284-290 (NTTPIVH)
MOD_CK1_1	283-289 (SNTTPIV)
MOD_N-GLC_1	283-288 (SNTTPI)
MOD_ProDKin_1	283-289 (SNTTPIV)

ELMs, LIG_FHA_1 and DOC_MAPK_gen_1 have been found to be associated with the Lysine residues present at 111 and 112 position, among which the Lysine at 111 position is most highly conserved among the HPV types. This Lysine plays a major role in Acetylation of E2 by p300, CBP, and pCAF. This motif interacts with FHA containing domains. Another protein transcription factor 19 also contains FHA domain and could play a role in enhancing capability of E2 during propagation of viral particle in the host cells. That has different lengths and amino acid sequences.

E4 has been reported to interact with cytokeratin [33] and different kinases including Cyclin B/Cdk1 [34], Cyclin a/Cdk2, Cyclin E/Cdk2 [35], p42 MAPK, protein kinase C[36], Protein kinase A [37], and SR protein kinase 1. Many E4 proteins are characterized by an N-terminal leucine-rich LLXLL motif. Initially it was found that if these residues are deleted from E4 of HPV1 and HPV16, it results in the inability of these mutated proteins to localize to the keratin based cytoskeleton. Latterly, in a research it was found that some other regions of E4 protein also play vital role for its interaction with cytokeratin. These include residues 61 to 92 of C-terminal region, in which the most conserved and important are residues 84 to 92 (LTVIVTLHP). Further analysis showed that in particular valines 86 and 88 [38].

In this study these findings were validated on the basis of presence of specific motifs in E4 in these particular regions (Table 6), both involving in interactions of E4 with cytokeratin. 9 different motifs were found in E4 which interact with their respective domains operating through the specific kinases. These specific kinases include MAP kinases, kinases of NEK and PIKK family.

Table 6: ELMs found in N-terminal and C terminal regions of E4 protein involved in interaction with Cytokeratins.

ELMs	Amino acids Sequence/ Position
DOC_MAPK_MEF2A_6	81-90 (KDGLTVIVTL)
LIG_FHA_1	83-89 (GLTVIVT)
LIG_LIR_Gen_1	64-70 (ETQWTVL), 63-70 (TQWTVL)
LIG_SUMO_SIM_anti_2	82-89 (DGLTVIVT)
LIG_SUMO_SIM_par_1	84-89 (LTVIVT)
MOD_CK1_1	60-66 (SCCTETQ)
MOD_GSK3_1	65-72 (TQWTVLQS), 70-77 (LQSSLHLT), 82-89 (DGLTVIVT)
MOD_NEK2_1	70-75 (LQSSLH), 74-79 (LHLTAH)
MOD_PIKK_1	62-68 (CTETQWT)
LIG_eIF4E_1	10-16 (YPLLKLL)

10 motifs have been found in E4 which interact with Protein kinases domain, thus facilitating the interaction of E4 with different kinases to phosphorylate or interact with keratins (Table 7).

Table 7: ELMs found in E4 with their CDs

S.No.	ELMs	CDs	Amino acids Sequence/ Position
1.	CLV_NRD_NRD_1	Peptidase_M16	40-42 (RRL)
2.	DOC_CKS1_1	CKS	21-26 (PTTPPR), 52-57 (PETPAT), 55-60 (PATPLS)
3.	DOC_MAPK_MEF2A_6	Pkinase	81-90 (KDGLTVIVTL)
4.	MOD_CDK_SPxK_1	Pkinase	20-26 (WPTTPPR)
5.	MOD_CK1_1	Pkinase	60-66 (SCCTETQ)
6.	MOD_GSK3_1	Pkinase	65-72 (TQWTVLQS), 70-77 (LQSSLHLT), 82-89 (DGLTVIVT)
7.	MOD_NEK2_1	Pkinase	70-75 (LQSSLH), 74-79 (LHLTAH)
8.	MOD_PKA_1	Pkinase	40-46 (RRLSSDQ)
9.	MOD_PKA_2	Pkinase	40-46 (RRLSSDQ)
10.	MOD_ProDKin_1	Pkinase	20-26 (WPTTPPR), 29-35 (PKPSPWA), 48-54 (QSQTPE), 5 1-57 (TPETPAT), 54-60 (TPATPLS)
11.	DOC_PP2A_B56_1	Armadillo type fold	59-65 (LSCCTET)
12.	DOC_WW_Pin1_4	WW	20-25 (WPTTPP), 29-34 (PKPSPW), 48-53 (QSQTPE), 5 1-56 (TPETPA), 54-59 (TPATPL)
13.	LIG_WW_3	WW	23-27 (TPPRP)
14.	LIG_BIR_II_1	BIR	1-5 (MADPA)
15.	LIG_BIR_III_1	BIR	1-5 (MADPA)

16	LIG_eIF4E_1	IF4E	10-16 (YPLLKLL)
17	LIG_FHA_1	FHA	83-89 (GLTVIVT)
18	LIG_LIR_Gen_1	Atg8	64-70 (ETQWTVL), 63-70 (TQWTVL)
19	LIG_SH3_1	SH3_1, SH3_9	30-36 (KPSPWAP)
20	LIG_SH3_2	SH3_1, SH3_9	21-26 (PTTPPR), 33-36 (KPSPWAP)
21	LIG_SH3_3	SH3_1, SH3_9	18-24 (STWPTTP), 21-27 (PTTPPRP), 30-36 (KPSPWAP), 49-55 (SQTPETP), 52-58 (PETPATP)
22	LIG_SUMO_SIM_anti_1	Rad60-SLD	82-89 (DGLTVIVT)
23	LIG_SUMO_SIM_par_1	Rad60-SLD	84-89 (LTVIVT)
24	MOD_LATS_1	Pkinase_C	39-45 (HRRLSSD)
25	MOD_PIKK_1	PI3_PI4_kinase	62-68 (CTETQWT)
26	TRG_ENDOCYTIC_2	Adap_comp_sub	10-13 (YPLL)
27	TRG_ER_diArg_1	WD40	40-42 (RRL)
28	TRG-NLS_MonoExtC_3	Arm	36-41 (PKKHRR)
29	TRG-NLS_MonoExtC_4	Arm	36-41 (PKKHRR)

In our work as mentioned earlier that the ELM source only maps and finds those motifs which are present on the surface of a protein. Motif matches which are buried in inner folded cores of domains of globular proteins are not reliable candidates. When we mapped E5 amino acid sequence on ELMs. We found only one significant motif on the surface region. But by applying the structural filter we got able to view the motifs lying in hidden regions. As in case of E5 due to involvement of its helices in interactions these motifs could play a major role. We found 22 motifs on E5 (Table 8)

Table 8: ELMs found in alpha helices of E5

ELM	Amino Acid Position/Sequence	CDs
CLV_PCSK_SKI1_1	30-34	Peptidase S8
DEG_APCC_DBOX_1	29-37	WD40
DOC_CYCLIN_RxL_1	27-38	Cyclin_N
DOC_MAPK_JIP1_4	30-36	Pkinase
DOC_MAPK_MEF2A_6	58-66	Pkinase
DOC_SPAK_OSRI_1	79-83	OSRI_C
LIG_14-3-3_CterR_2	79-83	14-3-3
LIG_eIF4E_1	68-74	eIF4e
LIG_FHA_1	7-13	FHA

	38-44	
LIG_LIR_Gen_1	37-42 38-42	Atg8
LIG_NRBOX	44-50	Hormone recep
LIG_Pex14_2	15-19	SH3_1 Px14_N
LIG_SH2_PTP2	68-71	SH2
LIG_SH2_STAT5	39-42 63-66 68-71	SH2
LIG_SH3_3	64-70	SH3
LIG_SUMO_SIM_anti_2	40-47 41-47	Rad60_SLD
LIG_SUMO_SIM_par_1	30-35 32-38	Rad60_SLD
MOD_CK1_1	35-41 37-43	Pkinase
MOD_GSK3_1	3-10 34-41	Pkinase
MOD_NEK2_1	32-37 34-39 49-54 73-78	Pkinase
MOD_Plk_4	7-13 38-44	PI3_PI4 kinase
TRG_ENDOCYTIC_2	39-42 63-66 68-71	Adap_comp_sub

Table 9: Proteins having validated interactions with HPV16 E5 (Reported in literature and taken from VirusHost database).

EGF Receptor	P38 MAP kinase
Erk1/2	Phospholipase C-gamma 1
AKT	ErB4
Cyclooxygenase-2	c-jun, c-fos
VEGF	EVER1 AND EVER2
c-Cbl	ZNt-1
16K vacuolar-ATPase subunit	Bap 31
MHC class I antigens	Calnexin
MHC class II antigens	Calpactin 1
CD1d	Karyopherin beta3
Connexin 43	Endothelin-1 receptor
Fas	Bcl-2
Cyclin dependent kinase inhibitor p21	Bax

HPV E5 protein has the ability to modulate the activity of a large number of proteins (Table 9), despite of having smaller size, and hydrophobic nature. It lacks the highly soluble and globular domains which play main role in mediation of protein-protein interactions. Preferably, an alternative mechanism is adopted by the E5 in order to engage their protein targets. An interface for stabilizing protein-protein interactions is provided by the hydrophobic region of the E5. In the hydrophobic environment of membrane, the amide of main chain and carboxyl groups of the transmembrane domains get attached with each another through hydrogen bonds thus minimizing the membrane insertion energetic cost, resulting the creation of an α -helix. Thus, despite of their smaller size and infrequent composition, the E5 proteins accept an energetically promising, well-formed assembly in cell membranes. In these arrangements the amino acid R groups are outwardly displayed on the surface of the helical axis. By this the well exposed side chains can easily interact with other molecules in a sequence specific way, within the cell membrane. With the help of these interactions, E5 can form homo-oligomers, or with other cellular proteins spanning transmembrane segments, ultimately creating protein complexes [39].

The cellular functions of an organism rely on signaling networks, which are stabilized majorly by protein kinases. In eukaryotes, the mitogen-activated protein kinases (MAPKs) are among the most important and most conserved members of this family of protein kinases. Studies have shown that one of the mechanisms by which 16HPVE5 stimulates EGFR signaling pathways, is based on its mechanism of affecting MAPK. In our study we found two motifs DOC_MAPK_JIP1_4 and DOC_MAPK_MEF2A_6, which interact with

proteins through the protein kinases domains based on MAP kinases. This validates the possibility of E5 to interact through MAPK. E5 has also been reported to interact with p38, ERK1/2, c-jun amino-terminal kinase, which are members of MAPK family. Our study also validates these interactions to occur through these motifs. Additionally, 4 motifs were also found which interact with protein kinases domains. These motifs could also play role in regulating other activities of E5 involving protein kinases. Oncogenesis has been shown to be promoted by abnormal synthesis and functioning of proteins of the PI3K-Akt pathway. The proteins affected by E5 also include Akt [39]. This study has predicted the motif MOD_Plk_4 interacting with the domains undergoing phosphorylation by PI3 and PI4. Same is the mechanism adopted by Akt proteins to operate the signaling pathways. In inflammatory process, Arachidonic acid is converted to thromboxane and prostaglandin by a key enzyme Cyclooxygenase (COX). COX-1 and COX-2 are two isoforms of COX. Chronic inflammation interposed by COX-2 plays some role in cancer progression and carcinogenesis. It is caused by various factors including bacterial infections and chemical irritants. MAPK activity is increased by Prostaglandin E2, that is the product of COX 2. As this mechanism of oncogenicity is associated with viral infection that is HPV. The E5 of HPV through its MAPK motifs could possibly increase the MAPK activity thus promoting the infection towards cervical cancer. E5 interacts with many cyclin dependent kinase inhibitors including Cyclin dependent kinase inhibitor including P38 MAP kinase, p27, Phospholipase C-gamma 1, ZNt-1, ErbB4, c-fos, c-jun, , EVER1, EVER2, Bap 31, Calnexin, Calpactin 1, Karyopherin beta3 and Endothelin-1 receptor. E5 has a motif DOC_CYCLIN_RxL_1 at position 27-38. Here we predict the involvement of this motif in stabilizing interaction of E5 with these Cyclin dependent kinase inhibitors as all of these have CYCLIN_N domain to interact, with which E5 could also interact through motif DOC_CYCLIN_RxL_1 [39].

Conclusion:

In present study, we used a bioinformatics algorithm to explore the crosstalk among the proteins of HPV and humans. The method was based on sequence alignment of HPV proteins E1, E2, E4, E5, E6 and E7, different datasets of human proteome, previously reported interactions between HPV and human proteins, and interactions of ELMs and SLiMs with protein domains. The output we retrieved, included a list of ELMs and SLiMs having occurrences in viral proteome, and a list of human proteins that we predicted to be interacted by the HPV proteins on the basis of interactions of Eukaryotic Linear Motifs of human proteins on HPV proteins and the counter domains of these motifs in human proteome. We predicted 12961 human proteins which could have interactions with one or more of the HPV proteins E1, E2, E4, E5, E6 and E7. The list of predicated proteins can be utilized to find out the connectivity between the virus HPV and human host. It can also be

used to determine the pathways involved in pathogenesis of HPV leading towards the cervical cancer.

One third of the human proteins were found to be having Pkinase domain. The motifs DOC_CYCLIN_1, MOD_CDK_SPK_2, and MOD_CDK_SPxxK_3 on N-terminal region of E1 were predicted, involved in maintaining its interaction with Cdk2 protein, that is already reported. The motifs having counter domain of Pkinase in N-terminal region of E1, were also found which helps in maintaining interaction of E1 with Histone H1 and chaperones proteins. The motifs ELMs, LIG_FHA_1 and DOC_MAPK_gen_1 in E2 were discovered to be involved in phosphorylation interactions with human proteins.

This methodology can be applied to other viruses also to predict the host proteins affected by that viral proteins. Therefore, this approach has probable utility in the prediction and identification of human proteins approached by different viruses. The resulting protein list will be useful for selection of optimal drug therapies by therapists and discovery of new antiviral drugs. The prediction and identification of motifs on viral proteome can lead to the exploration of hot spot residues of both the viral and human proteins to define more precise drug targets at amino acid residual level. Our study examined our predictions of proteins with independent biological datasets for identifying the most promising protein interacting pairs. From this effort we hope that these predictions will be helpful in accelerating experimental efforts for defining a reliable system of HPV and human protein interactions.

a) Ethics approval and consent to participate: Not Applicable

b) Animal and Human rights: No Animal or Human used.

c) Availability of Data and materials: Available on request.

d) Consent for Publication: Not Applicable

e) Conflict of Interest: No conflict of interest

f) Acknowledgements: None

g) Funding Information: No Funding

References:

1. B.Alberts, "DNA replication and recombination". Nature, 421(6921):431-5, 2003.
2. D. Voet and J. G. Voet, "Biochemistry", John Wiley & Sons, New Jersey,3. ed., 2004.
3. M.CooperGeoffrey and R. E. Hausman, "The Cell: A Molecular Approach", Sinauer Associates, Inc., fifth ed., 2009.
4. H. F. Lodish, A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore, and D. James, Molecular Cell Biology, vol.5, W. H. Freeman, 4ed.,2008.
5. Elde NC, Malik HS. The evolutionary conundrum of pathogen mimicry. Nat Rev Microbiol. 2009; 7(11):787-97.

6. Hugo W., Sung WK., Ng SK. (2013) Discovering Interacting Domains and Motifs in Protein–Protein Interactions. In: Mamitsuka H., DeLisi C., Kanehisa M. (eds) Data Mining for Systems Biology. Methods in Molecular Biology (Methods and Protocols), vol 939. Humana Press, Totowa, NJ. https://doi.org/10.1007/978-1-62703-107-3_2
7. de Chassey B, Meyniel-Schicklin L, Aublin-Gex A, Andre P, Lotteau V. New horizons for antiviral drug discovery from virus-host protein interaction networks. *Curr Opin Virol.* 2012; 2(5):606–13.
8. Zoraghi R, Reiner NE. Protein interaction networks as starting points to identify novel antimicrobial drug targets. *Curr Opin Microbiol.* 2013; 16(5):566–72.
9. Dömling A, Mannhold R, Kubinyi H, Folkers G. Protein-Protein Interactions in Drug Discovery. In: Methods and Principles in Medicinal Chemistry. Weinheim: Wiley: 2013.
10. Ma-Lauer Y, Lei J, Hilgenfeld R, von Brunn A. Virus-host interactomes–antiviral drug discovery. *Curr Opin Virol.* 2012; 2(5):614–21.
11. Van Roey K, Uyar B, Weatheritt RJ, Dinkel H, Seiler M, Budd A, et al. Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem Rev.* 2014; 114(13):6733–78.
12. Diella F, Haslam N, Chica C, Budd A, Michael S, Brown NP, et al. Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Front Biosci J Virtual Libr.* 2008; 13:6580–603.
13. Davey NE, Van Roey K, Weatheritt RJ, Toedt G, Uyar B, Altenberg B, et al. Attributes of short linear motifs. *Mol Biosyst.* 2012; 8(1):268–81.
14. Perkins JR, Diboun I, Dessailly BH, Lees JG, Orengo C. Transient protein-protein interactions: structural, functional, and network properties. *Structure.* 2010; 18(10):1233–43.
15. Dinkel H, Michael S, Weatheritt RJ, Davey NE, Van Roey K, Altenberg B, et al. ELM–the database of eukaryotic linear motifs. *Nucleic Acids Res.* 2012; 40(Database issue):D242–251.
16. Dinkel H, Van Roey K, Michael S, Davey NE, Weatheritt RJ, Born D, et al. The eukaryotic linear motif resource ELM: 10 years and counting. *Nucleic Acids Res.* 2014; 42(Database issue):D259–266.
17. Davey NE, Cyert MS, Moses AM. Short linear motifs - ex nihilo evolution of protein regulation. *Cell Commun Signal.* 2015; 13(1):43.
18. Berlow RB, Dyson HJ, Wright PE. Functional advantages of dynamic protein disorder. *FEBS Lett.* 2015; 589(19 Pt A):2433–40.
19. Latysheva NS, Flock T, Weatheritt RJ, Chavali S, Babu MM. How do disordered regions achieve comparable functions to structured domains? *Protein Sci.* 2015; 24(6):909–22.

20. Halehalli RR, Nagarajaram HA. Molecular principles of human virus protein-protein interactions. *Bioinformatics*. 2015; 31(7):1025–33.
21. Kadaveru K, Vyas J, Schiller MR. Viral infection and human disease—insights from minimotifs. *Front Biosci*. 2008; 13:6455–471.
22. Ganti K, Broniarczyk J, Manoubi W, Massimi P, Mittal S, Pim D, et al. The human Papillomavirus E6 PDZ binding motif: from life cycle to malignancy. *Viruses*. 2015; 7(7):3530–51.
23. Volchkova VA, Klenk HD, Volchkov VE. Delta-peptide is the carboxy-terminal cleavage fragment of the nonstructural small glycoprotein sGP of Ebola virus. *Virology*. 1999; 265(1):164–71.
24. Sugrue RJ, Brown C, Brown G, Aitken J, McL Rixon HW. Furin cleavage of the respiratory syncytial virus fusion protein is not a requirement for its transport to the surface of virus-infected cells. *J Gen Virol*. 2001; 82(Pt 6):1375–86.
25. Evans P, Dampier W, Ungar L, Tozeren A. Prediction of HIV-1 virus-host protein interactions using virus and host sequence motifs. *BMC Med Genomics*. 2009; 2:27.
26. S. Pundir, M. J. Martin, C. O'Donovan C. **UniProt Protein Knowledgebase**. [Methods Mol. Biol. 1558:41-55, 2017](#).
27. M. Gouw, S. Michael, H. Sámano-Sánchez, M. Kumar, A. Zeke, B. Lang, B. Bely, L. B Chemes, N. E. Davey, Z. Deng, F. Diella, C. Gürth, A. Huber, S. Kleinsorg, L. S. Schlegel, N. Palopoli, K. V. Roey, B. Altenberg, A. Reményi, H. Dinkel, T. J. Gibson. The eukaryotic linear motif resource – 2018 update, *Nucleic Acids Research*, Volume 46, Issue D1, 4 January 2018, Pages D428–D434, <https://doi.org/10.1093/nar/gkx1077>
28. N. E. Davey, G. Travé, T. J. Gibson. “How viruses hijack cell regulation. Trends”. *Biochem Sci.*;36(3):159-169, 2011.
29. B. Uyar, R. J. Weatheritt, H. Dinkel, N. E. Davey, T. J. Gibson. “Proteome-wide analysis of human disease mutations in short linear motifs: neglected players in cancer?” *Mol Biosyst.*;10(10):2626-2642, 2014.
30. Z. Songyang, A. S. Fanning, C. Fu, J. Xu, S. M. Marfatia, A. H. Chishti, A. Crompton, A. C. Chan, J. M. Anderson, L. C. Cantley. “Recognition of unique carboxyl-terminal motifs by distinct PDZ domains.” *Science*; 275:73–77, 1997.
31. A. R. Atilgan, S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, I. Bahar. “Anisotropy of fluctuation dynamics of proteins with an elastic network model.” *Biophys J*; 80:505–515, 2001.
32. A. A. McBride. “The papillomavirus E2 proteins.” *Virology*.;445(1-2):57-79, 2013.
33. J. Doorbar. “The E4 protein; structure, function and patterns of expression.” [Virology Volume 445, Issues 1-2](#), October, Pages 80-98, 2013.

-
34. C. E. Davy, D. J. Jackson, K. Raj, W. L. Peh, S. A. Southern, P. Das, R. Sorathia, P. Laskey, K. Middleton, T. Nakahara, Q. Wang, P. J. Masterson, P. F. Lambert, S. Cuthill, J. B. A. Millar, J. Doorbar. "Human Papillomavirus Type 16 E1^{E4}-Induced G₂ Arrest Is Associated with Cytoplasmic Retention of Active Cdk1/Cyclin B1 Complexes." *Journal of Virology* Mar, 79 (7) 3998-4011, 2005.
35. Q. Ding, L. Li, & P. Whyte. "Human papillomavirus 18 E1^{E4} protein interacts with cyclin A/CDK 2 through an RXL motif." *Mol Cell Biochem* 373, 29–40, 2013..
36. G. L. Knight, A. G. Pugh, E. Y. Ian Bell, R. Wilson, C. A. Moody, L. A. Laimins, S. Roberts. "A cyclin-binding motif in human papillomavirus type 18 (HPV18) E1^{E4} is necessary for association with CDK-cyclin complexes and G₂/M cell cycle arrest of keratinocytes, but is not required for differentiation-dependent viral genome amplification or L1 capsid protein expression." *Virology*, 412 (1), pp. 196-210, 2011.
37. Q. Wang, A. Kennedy, P. Das, P. B. McIntosh, S. A. Howell, E. R. Isaacson, S. A. Hinz, C. Davy, J. Doorbar. "Phosphorylation of the Human Papillomavirus Type 16 E1^{E4} Protein at T57 by ERK Triggers a Structural Change That Enhances Keratin Binding and Protein Stability." *Journal of Virology* Mar, 83 (8) 3668-3683, 2009.
38. S. Roberts, S. R. Kingsbury, K. Stoeber, G. L. Knight, P. H. Gallimore, G. H. Williams. "Identification of an Arginine-Rich Motif in Human Papillomavirus Type 1 E1^{E4} Protein Necessary for E4-Mediated Inhibition of Cellular DNA Synthesis In Vitro and in Cells." *Journal of Virology* Aug, 82 (18) 9056-9064, 2008.
-