RESEARCH ARTICLE

Prediction of Promiscuous Epitopes in the E6 Protein of Three High Risk Human Papilloma Viruses: A Computational Approach

Subramanian Nirmala, Chinnappan Sudandiradoss*

Abstract

A najor current challenge and constraint in cervical cancer research is the development of vaccines against human papilloma virus (HPV) epitopes. Although many studies are done on epitope identification on HPVs, no computational work has been carried out for high risk forms which are considered to cause cervical cancer. Of all the high risk HPVs, HPV 16, HPV 18 and HPV 45 are responsible for 94% of cervical cancers in women worldwide. In this work, we computationally predicted the promiscuous epitopes among the E6 proteins of high risk HPVs. We identified the conserved residues, HLA class I, HLA class II and B-cell epitopes along with their corresponding secondary structure conformations. We used extremely precise bioinformatics tools like ClustalW2, MAPPP, NetMHC, EpiJen, EpiTop 1.0, ABCpred, BCpred and PSIPred for achieving this task. Our study identified specific regions 'FAFR(K)DL' followed by 'KLPD(Q)LCTEL' fragments which proved to be promiscuous epitopes present in both human leukocyte antigen (HLA) class I, class II molecules and B cells as well. These fragments also follow every suitable character to be considered as promiscuous epitopes with supporting evidences of previously reported experimental results. Thus, we conclude that these regions should be considered as the important for design of specific therapeutic vaccines for cervical cancer.

Keywords: High risk HPVs - E6 protein - computational - epitope - peptide vaccine

Asian Pac J Cancer Prev, **14 (7)**, 4167-4175

Introduction

Approximately 500,000 women worldwide develop cervical cancer and it is the most common cancer among women in underdeveloped countries (Soliman et al., 2004) with 274,000 deaths each year due to this disease. In India, it has become the leading cancer among women with an annual incidence of about 130,000 cases and 70-75,000 deaths (Agarwal et al., 2011). HPV infection is known to be one of the important causes for the development of cervical cancer in women, which also forms a major risk factor for the development of anal, penile, and vulvar cancers. Papilloma Viruses are not classified by serotype, but by genotype, and to date, approximately 151 HPV types have been identified in humans alone (Bernard et al., 2010) with a circular genome of approximately 8 kbp. On the basis of carcinogenicity, the HPVs can be divided into two subgroups as low risk (e.g. HPV 6 and 11) associated with benign genital lesions and high-risk (e.g. HPV 16, 18 and 45) associated with invasive carcinomas of the cervix (Jacob et al., 2003). HPV comprises of three functional regions, namely, a non-coding region called the long control region, involved in the regulation of both HPV transcription and replication, the coding regions for early genes (E1, E2, E4, E5, E6 and E7) which regulate the vegetative and productive phases of viral cycle and the late genes, L1 and L2, which codes for the major and minor capsid proteins, respectively (Boccardo et al., 2010). The risk factors for tumor development include persistent infection with high-risk viral types (Longworth et al., 2004) and transforming potential of these high-risk HPVs is due to viral oncoproteins, E6 and E7 (Fakhry and Gillison, 2006). The E7 induces cell proliferation, disrupting the cell cycle regulation by inactivating the pRb protein, whereas E6 blocks cell apoptosis by directing the p53 tumor suppressor protein to the proteasome. Thus, the expression of high-risk HPV E6 and E7 results in cellular proliferation, loss of cell cycle regulation, impaired cellular differentiation, increased frequency of spontaneous and mutagen-induced mutations, and chromosomal instability (Munger and Howley, 2002). According to the latest report (de Sanjose et al., 2010), HPV types 16, 18, 31, 33, 35, 45, 52, and 58 should be given priority when the cross-protective effects of current vaccines are considered. It is also expected that the next generation of cervix cancer vaccines will specifically include each of the eight HPV types. Of these the most common high-risk subtypes of HPV are 16, 18, and 45.

*Bioinformatics Division, School of Biosciences and Technology, VIT University, Vellore, India *For correspondence: csudandiradoss@vit.ac.in*

Immunotherapeutic approaches based on prophylactic and therapeutic vaccine development have made tremendous efforts to prevent cervical cancer that plague mankind (Waldmann, 2003). Though the prophylactic vaccines prevent infection by induction of neutralizing antibodies production, they have their limitations like cost, coverage of HPV types, (Ma et al., 2010) no safety, no immediate effect and no long lasting protection (Gnanamony et al., 2007). Merck's Gardasil, a quadrivalent vaccine against HPV 6, 11, 16, and 18 and GlaxoSmithKline's Cervarix, a bivalent vaccine protective against HPV 16 and 18 are two prophylactic vaccines administered to prevent the HPV entry into the host (Tumban et al., 2011). Nevertheless, it is estimated that these two preventative vaccines will provide protection against cervical cancer caused by HPV 16 and 18 only (Govan, 2008). Also, administration of these vaccines in developing countries like India has proved severe side effects and even deaths in some cases which ultimately lead a complete ban for the usage of these vaccines. But the therapeutic vaccine on the other hand, which is considered to be safe, stable and easily producible (Wu et al., 2010) is the need of the hour as considerable population is already suffering from HPV infection worldwide. HPV E6 and E7 proteins though do not elicit strong immune response; are targeted by the scientific community (Morrow et al, 2013) as they play a significant role in the malignant transformation of HPV associated lesions (Ma et al., 2010) and form the ideal choice for the therapeutic vaccines.

Different therapeutic strategies have been developed which includes vector-based, peptide-based, protein based, DNA based, cell based and combinational approaches (Ma et al., 2010). Here we have focused on the prediction of promiscuous epitopes for peptide based therapeutic vaccination against subset of antigens which elicits the immune response derived from high risk HPVs (De Groot, 2004). The possible mechanism why some women do not eradicate HPV is due to defect in their cell-mediated immune response (Mark et al., 2009). However due to continuous effect on HPV research, well documents reports are available mentioning the significance of cell-mediated immune responses in eradicating the HPV infections (Agnieszka et al, 2012). Thus, the HPV immunotherapy has largely focused on cervical cancer because of the poor prognosis for patients with advanced disease (Smith, 2004). The in-depth understanding of antigen recognition at the molecular level has become a boon for the development of rationally designed peptide vaccines. The basic idea behind the peptide vaccines lies in the production of the identified immunodominant B-cell and T-cell epitopes to induce specific immune responses (Patronov and Doytchinova, 2013). HPV immunotherapy with peptide based vaccine has created a new dimension in cancer treatment, as these vaccine candidates can be easily identified by the human cells. But Human Leukocyte Antigens (HLA) both class I and class II molecules being highly polymorphic in nature, experimental approaches in identifying the T-cell epitopes within these antigens, are not applicable against all different HLA alleles. Moreover as a result of this polymorphism, there is an increased chance of identification of several epitopes for every pathogen to be recognized by individuals within the population (Walshe, 2009).

The field of immunoinformatics which deals with the mapping of potential B-cell and T-cell epitopes (Tomar and De, 2010) has thus made the scientists to look forward in search of promiscuous epitopes which not only minimizes the number of experiments but also enables a systematic identification of candidate epitopes (Bian et al., 2003) from which experimental testing could be made easier. Also peptide based vaccines reduces the possibility of provoking any reaction against self- antigens, thereby proving to be a safer vaccine by inhibiting the stimulation of auto immunity (Sirskyj et al., 2011). Epitopes are regions present on the antigens that are easily recognized by the antibodies (Yao et al., 2013) and thus epitope-based techniques allow the accurate and precise characterization of Immune responses (Peters et al., 2005). Moreover, the main aim of epitope prediction is to design a molecule that can mimic the structure and function of a genuine epitope and replace it in medical diagnostics, therapeutics and also in vaccine design (Pingping et al., 2011). In addition to all these, the most important and prior advantage of epitope based vaccines is that it eliminates the possibility of using the whole deadly viral proteins (Shehzadi et al., 2011). Predictions of reliable epitopes are vital for rational vaccine design which identifies the potential targets and in-turn paves way for immunotherapeutic cancer treatment (Iurescia et al., 2012). The immunogenic peptides presented by HLA class I molecules are mostly endogenous proteins. These antigenic peptides are generated by the proteasome and transported by the transporter associated with antigen processing (TAP) protein (Ackerman and Cresswell, 2004). The TAP protein enables the viral peptide entry into the rough endoplasmic reticulum so that these peptides could be made available to the complex HLA class I molecules (Engelhard, 1994). On the other hand, the peptides presented by HLA class II molecules are mainly derived from exogenous antigens. Thus, epitopes are useful in assessing dominant patterns of immune recognition and precisely tracking immune responses following natural exposure to pathogens or in response to vaccination (Gallagher and Man, 2007). The peptide based vaccine also has an added advantage of increased safety, the opportunity to rationally engineer epitopes for increased potency and the ability to focus immune responses on conserved epitopes (Lin et al., 2013; Oyarzun et al., 2013). Thus, the aim of our study was to identify peptide(s) that are commonly recognized and restricted by both the T-cell and B-cell epitopes for the E6 protein of the three high risk HPVs (HPV 16, HPV 18, HPV 45) so that one common peptide based vaccine can be developed against these viral genomes. Therefore, in this work we dealt with epitope predictions of E6 protein of three high risk HPVs namely HPV 16, HPV 18 and HPV 45, as they are responsible for 94% of cervical adenocarcinomas.

Materials and Methods

Dataset

The E6 protein sequences for the three high risk HPVs,

namely, HPV 16, HPV 18 and HPV 45 were retrieved using Uniprot database [The Universal protein resource (Uniprot) in 2010] whose accession numbers are P03126 (Seedorf et al., 1985), P06463 (Cole and Danos, 1987) and P21735 (Delius and Hofmann, 1994) respectively. All the E6 proteins taken for our study are the basic nuclear and cytoplasmic protein of about 18 KDa with 158 amino acid residues each (Androphyl et al., 1987). These sequences were used for the epitopic prediction.

Sequence alignment

We computed the conservation (Zhang and Niu, 2010) of amino acid residues in each E6 protein sequence of the three selected genomes using the ClustalW2 multiple alignment program (Larkin et al., 2007; Sudandiradoss et al., 2008). This is because the epitopic regions that are highly conserved are less prone to antigen escape and viral mutation (Shehzadi et al., 2011). The alignment file was then used to estimate the evolutionary distance among the HPV 16, HPV 18 and HPV 45 strains. We followed Unweighted Pair of Group Method with Arithmetic Mean (UPGMA) (Khan et al., 2008) which is embedded in MEGA4 (Koichiro et al., 2007) program for calculating the evolutionary distance. This program evaluates the evolutionary distance by following the Maximum Composite Likelihood (MCL) feature (Khan et al., 2008) and provides the evolutionary distance for all pairs of sequences simultaneously. Also, by following this MCL feature, we reduced the errors obtained by the Independent Estimation (IE) (Whelan et al., 2001) approach considerably and reported the phylogenics with more accuracy (Koichiro et al., 2007).

HLA class I epitope prediction

We used three different bioinformatics tools namely MAPPP (Hakenberg et al., 2003), NetMHC (Lundegaard et al., 2008) and EpiJen (Doytchinova et al., 2006) for predicting the epitopes at different levels of HLA I processing. Each tool has its specialization for predicting the epitopes at each levels of HLA class I molecules dispensation. MAPPP is the MHC-I Antigenic Peptide Processing Prediction tool for identifying the potential antigenic epitopes presented on the cell surface by Human Leukocyte Antigen class I molecules to CD8+ T-lymphocytes. This method combines proteasome cleavage prediction with TAP transport. FragPredict is the part of MAPPP package that deals with the proteasome cleavage prediction. FragPredict consists of two algorithms. The first algorithm uses a statistical analysis of cleavage enhancing and inhibiting amino acid motifs to predict potential proteasome cleavage sites (Holzhutter et al., 1999). The second algorithm, which uses the results of the first algorithm as an input and predicts which fragments are most likely to be generated. When the score is 1 as an output, it indicates the cleavage prediction is perfect and that particular peptide can be considered to be a promiscuous epitope. Similarly, we used NetMHC tool for predicting CTL epitopes which are potential candidates for designing peptide vaccines for various diseases. This tool follows artificial neural networks and position-specific scoring matrices (PSSM) (Nielsen et al., 2004). Artificial neural network predictions are given as actual IC_{50} values (Nielsen et al., 2003) whereas PSSM predictions are given as a log-odds likelihood scores. We also used the EpiJen in which a source protein is passed through four steps: proteasome cleavage, TAP transport, MHC binding and epitope selection. At each stage, different proportions of non epitopes are eliminated. The final set of peptides represents no more than 5% of the whole protein sequence and will contain the true epitope.

HLA class II epitope prediction

HLA class II epitope prediction is a critical immuno-informatic problem within vaccine design. We used EpiTop 1.0 (Dimitrov et al., 2010) for this prediction written in PHP, HTML, and integrating the MySQL database environment. It deals with ligands binding to a set of similar proteins. In a traditional QSAR analysis (Hellberg et al., 1987), the X matrix of descriptors only includes chemical information from ligands. The proteochemometric X matrix (Lapinsh et al., 2001) contains information from both proteins and ligands. Proteochemometrics is specifically designed to solve QSAR tasks where a set of ligands bind to a set of related proteins. A single proteochemometrics model potentially predicts peptide binding to a whole group of HLA class II proteins.

Linear B-cell epitope prediction

Similar to HLA class I and II epitopes, B-cell epitopes also play a vital role in the development of peptide vaccine (Larkin et al., 2007). Hence, prediction of B-cell epitopes is important to many immunodetection and immunotherapeutic applications as they elicit humoral immune response. B cell epitopes are classified into two groups as: *i*) the linear or continuous epitopes and; *ii*) conformational or discontinuous epitopes. We were interested in continuous epitopes because, these are considered to be highly potential for vaccines development (Chandra et al., 2012). We used ABCpred (Saha and Raghava, 2006) and BCpreds (Yasser et al., 2008) tools for predicting linear epitopes. ABCpred applies artificial neural network (Nielsen et al., 2003) for predicting continuous epitopes in protein sequences. The predicted B-cell epitopes are ranked according to their score obtained by trained recurrent neural network (Kavitha et al., 2013). A default threshold value of 0.51 and amino acid sequence length of 16 was set for the epitope prediction. Likewise, BCpreds also predicts linear B-cell epitopes using Support Vector Machine (SVM) classifiers (Alessia et al., 2006) trained on the homology reduced data sets of B-cell epitopes. This method predicts with 86.7% accuracy for a linear epitope length of 20 amino acid residues. Hence a default value of 20 amino acid length was taken for the prediction. Greater the score, higher the reliability for a peptide sequence to be considered as an epitope.

Epitope conformation by secondary structure prediction

According to the rule of thumb, epitopes that are present either in the coil or helix regions of the secondary structure of the protein are confirmed to be the promiscuous epitopes. Hence, PSIPred tool (Jones, 1999)

was used to confirm whether the predicted peptides by the previously mentioned tools are the promiscuous epitopes. PSIPred follow neural network which is used to predict protein secondary structure based on the position specific scoring matrices generated by PSI-BLAST (Altschul et al., 1997). This method is divided into three stages as follows: *i*) Generation of a sequence profile using PSI-BLAST; *ii*) prediction of initial secondary structure and; *iii*) finally the filtering of the predicted structure. PSIPRED has 3 filtering options which filter out low complexity regions, likely transmembrane segments and coiled-coil regions.

Results

The multiple sequence alignment for the E6 proteins of high risk HPV 16, HPV 18 and HPV 45 was done using Clustal W2 (Figure 1). The E6 protein sequences have the following conserved regions KLPD(Q)LCTEL, FAFK(R) DL, CVYCK, LLIRC, CQKPL and RRETQV(L) among the high risk HPV genomes. Further, we also performed the phylogenetic analysis using MEGA4 (Figure 2). It was observed that HPV 45 is nearest neighbor of HPV 18 with an evolutionary distance of 0.10 whereas HPV 16 is the farthest neighbor estranged by distance 0.25. It is very much evident that though HPV 16 is a distant neighbor for HPV 18 and HPV 45 still six regions are highly conserved among the three genomes. Of the six highly conserved sequences, the two sequences namely KLPD(Q)LCTEL and FAFK(R)DL have only a single amino acid residue change in their fourth position. For instance, the fourth position of conserved sequence KLPD(Q)LCTEL contains glutamine in HPV 16 E6 protein whereas aspartic acid is present in HPV 18 and HPV 45 E6 proteins. Similarly, the FAFK(R)DL conserved sequence also varies at the fourth position with Arginine in HPV 16 and Lysine in HPV 18 and HPV 45 E6 proteins. For easy convenience, we have shown the fourth residue of HPV 16 E6 protein of both the conserved sequences KLPD(Q)LCTEL and FAFK(R)DL in parenthesis. Having predicted the highly conserved regions among the three high risk HPVs, we further analyzed the proteasomal cleavage and epitope prediction for both T-cell (HLA class I and HLA class II) and B-cell in order to confirm whether any of these sequences could act as promiscuous epitopes.

We pursued different tools: *i*) for the prediction antigen preprocessing by proteasome and; *ii*) for the prediction of promiscuous epitopes in HLA class I molecules since the process of these molecules occurring at different levels (Tong et al., 2006). The MAPPP tool identifies the regions KLPD(Q)LCTEL and FAFK(R) DL as peptide fragments processed by the proteasome for all the three HPVs of interest (Table 1). It is also to be noted that there are hydrophobic residues like alanine and phenylalanine in FAFK(R)DL fragment; leucine and proline in KLPD(Q)LCTEL fragment at second and third positions respectively. This makes a strong evident that these fragments could consider as promiscuous epitopes since TAP transportation always favors a hydrophobic residues at second and third positions. Thus, we narrowed down our prediction analysis on these two fragments from six fragments which were predicted in multiple sequence alignment analysis.

We retrieved the cytotoxic T cells (CTL) epitope predictions from NetMHC tool. Based on the score values, we categorized the predicted epitopes as strong binders which showed values <50, as weak binders which showed values range from 50-500 and non binders which showed >500 (Table 2). From this analysis, we also found that the fragments FAFK(R)DL and KLPD(Q)LCTEL were predominantly restricted as promiscuous epitopes in all the alleles of high risk Human Papilloma Viruses. Of these, the four alleles HLA-A0201, HLA-A0202, HLA-A0203, HLA-A0206 restricted both these fragments. Similarly, we obtained the results from EpiJen tool which predicted epitopes for 18 different alleles. Interestingly, we noted that the same two fragments namely KLPD(Q)LCTEL and FAFK(R)DL were predicted as promiscuous epitopes by this tool among the alleles for high risk HPVs (Table 3).

We used EpiTop 1.0 for predicting HLA class II epitopes. Surprisingly, almost all the alleles were shown with the epitope fragments KLPD(Q)LCTEL and FAFK(R)DL for HLA class II molecules (Table 4). Also, these two fragments fulfill most of the conditions for HLA-DR motifs reported by Sette et al. (1993).

In addition to the HLA I and HLA II molecules, prediction of immunogenic epitopes in B- cells will be a valid outcome. To achieve this task, we used ABCpred and BCpreds tool for predicting linear B-cell epitopes. The results obtained from ABCpred reveals that the epitope fragment FAFK(R)DL exist as the most prominent epitope in all the three high risk HPVs analyzed. In the case of

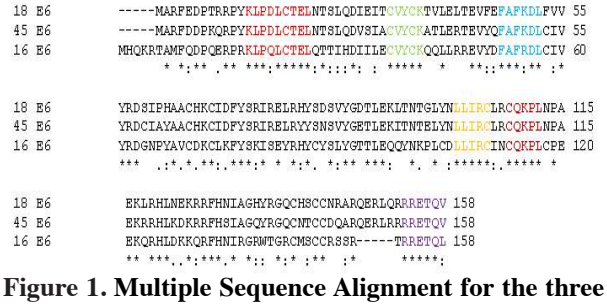

Figure 1. Multiple Sequence Alignment for the three HR-HPVs using CLUSTALW2.1

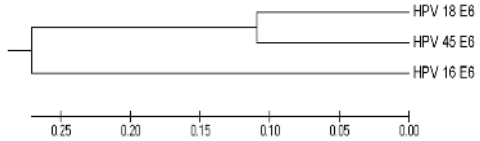

Figure 2. Phylogenetic Analysis by UPGMA Distance Matrix Method of E6 Protein of Three High Risk HPVs using MEGA4

Table 1. Prediction of Proteasomal Cleavage of High Risk HPV E6 Proteins by MAPPP Tool

Protein	Process	Peptide	Starting Position	Cleavage Probability
HPV 16 E6	Proteasome	KLPDLCTEL	18	0.999
	TAP	FAFRDLCIV	52	1.000
HPV 18 E6	Proteasome	KLPDLCTEL	13	0.996
	TAP	FAFRDLCIV	47	0.935
HPV 45 E6	Proteasome	KLPDLCTEL	13	0.999
	TAP	FAFRDLCIV	47	1.000

*TAP - Transporter associated with antigen processing

Table 2. Names of Pomiscuous Epitopes and their HLA Class I alleles Predicted by NetMHC and EpiJen Tools

Predicted Epitope	Prediction of HLA class I molecules by NetMHC			Prediction of HLA class I molecules by EpiJen		
	HPV 16 E6	HPV 18 E6	HPV 45 E6	HPV 16 E6	HPV 18 E6	HPV 45 E6
KLPD(Q)LCTEL	HLA-A0201	HLA-A0201	HLA-A0201	HLA-A0201	HLA-A0201	HLA-A0201
	HLA-A0202	HLA-A0202	HLA-A0202	HLA-A0202	HLA-A0202	HLA-A0202
	HLA-A0203	HLA-A0203	HLA-A0203	HLA-A0203	HLA-A0203	HLA-A0203
	HLA-A0206	HLA-A0206	HLA-A0206	HLA-A0206	HLA-A0206	HLA-A0206
	HLA-A0211	HLA-A0211	HLA-A0211			
	HLA-A0216	HLA-A0212	HLA-A0212			
	HLA-A0250	HLA-A0216	HLA-A0216			
	HLA-A3201	HLA-A0219	HLA-A0219			
		HLA-A0250	HLA-A0250			
		HLA-A3201	HLA-A3201			
FAFK(R)DLCIV	HLA-A0201	HLA-A0201	HLA-A0201	HLA-A0201	HLA-A0201	HLA-A0201
	HLA-A0202	HLA-A0202	HLA-A0202	HLA-A0202	HLA-A0202	HLA-A0202
	HLA-A0203	HLA-A0203	HLA-A0203	HLA-A0203	HLA-A0203	HLA-A0203
	HLA-A0206	HLA-A0206	HLA-A0206	HLA-A0206	HLA-A0206	HLA-A0206
	HLA-A0211	HLA-A0211	HLA-A0211	HLA-A6801	HLA-A6801	HLA-A6801
	HLA-A0212	HLA-A0212	HLA-A0212	HLA-A6802	HLA-A6802	HLA-A6802
	HLA-A0216	HLA-A0216	HLA-A0216	HLA-B3501	HLA-B3501	HLA-B3501
	HLA-A0250	HLA-A0219	HLA-A0250	HLA-B51	HLA-B51	HLA-B51
	HLA-A6802	HLA-A0250	HLA-A6802	HLA-B53	HLA-B53	HLA-B53
	HLA-A6901	HLA-A6802	HLA-A6901			
	HLA-B1503	HLA-A6901	HLA-B1503			
	HLA-B5401	HLA-B1503	HLA-B5401			
		HLA-B4601				
		HLA-B5401				

Table 3. Names of Promiscuous Epitopes and their HLA Class II Alleles Predicted by EpiTop

Allele	HPV 16 E6			HPV 18 E6			HPV 45 E6		
	Position	Peptide	Log ($1/IC_{50}$)	Position	Peptide	Log ($1/IC_{50}$)	Position	Peptide	Log ($1/IC_{50}$)
DRB1*0101	18	KLPQLCTEL	9.146	47	FAFKDLFVV	9.361	47	FAFKDLCIV	9.38
	52	FAFRDLCIV	8.326	13	KLPDLCTEL	9.163	13	KLPDLCTEL	9.163
DRB1*0301	18	KLPQLCTEL	5.738	13	KLPDLCTEL	5.91	13	KLPDLCTEL	5.91
	52	FAFRDLCIV	6.467						
DRB1*0401	52	FAFRDLCIV	8.361	13	KLPDLCTEL	7.78	13	KLPDLCTEL	7.78
	18	KLPQLCTEL	7.607	47	FAFKDLFVV	7.667			
DRB1*0404	—	—	—	13	KLPDLCTEL	6.004	13	KLPDLCTEL	6.004
DRB1*0405	18	KLPQLCTEL	7.851	13	KLPDLCTEL	8.024	13	KLPDLCTEL	8.024
				47	FAFKDLFVV	7.663			
DRB1*0701	18	KLPQLCTEL	8.367	13	KLPDLCTEL	8.441	13	KLPDLCTEL	8.441
	52	FAFRDLCIV	7.814	47	FAFKDLFVV	7.956			
DRB1*0802	18	KLPQLCTEL	7.206	13	KLPDLCTEL	7.398	13	KLPDLCTEL	7.398
				47	FAFKDLFVV	7.153	47	FAFKDLCIV	7.242
DRB1*0901	18	KLPQLCTEL	7.302	13	KLPDLCTEL	7.468	13	KLPDLCTEL	7.468
							47	FAFKDLCIV	7.829
DRB1*1101	18	KLPQLCTEL	7.326	13	KLPDLCTEL	7.518	13	KLPDLCTEL	7.518
				47	FAFKDLFVV	7.275			
DRB1*1201	—	—	—	—	—	—	13	KLPDLCTEL	5.597
DRB1*1302	18	KLPQLCTEL	7.435	13	KLPDLCTEL	7.606	13	KLPDLCTEL	7.606
	52	FAFRDLCIV	6.805						
DRB1*1501	18	KLPQLCTEL	7.373	13	KLPDLCTEL	7.423	13	KLPDLCTEL	7.423
	52	FAFRDLCIV	6.174	47	FAFKDLFVV	6.294	44	VYQFAFKDL	6.607

KLPD(Q)LCTEL fragment, it holds ninth in HPV 16, third in HPV 18 and eighth position in HPV 45 E6 proteins (Table 5). Similarly, we obtained the BCpreds results for linear B-cell epitopes. It reveals each 4 epitopes for HPV 16 and HPV 45 and two epitopes for HPV 18 E6 which includes these two fragments (Table 6). From the B-cell epitope predictions, we understood that the fragments FAFK(R)DL and KLPD(Q)LCTEL could be consider as promiscuous epitopes against high risk HPVs.

To confirm whether the predicted promiscuous epitopes follows an acceptable secondary structure or not, we performed PSIPred analysis (Figure 3). It is known that the epitopic regions should either lie in helix or in coil regions. Intuitively, the predicted epitope FAFK(R)DL lies in coil region and KLPD(Q)LCTEL lies in helix region. These results confirm that of all the other fragments predicted, these two fragments namely, FAFK(R)DL and KLPD(Q)LCTEL can certainly be considered as promiscuous epitopes against E6 proteins of high risk HPVs.

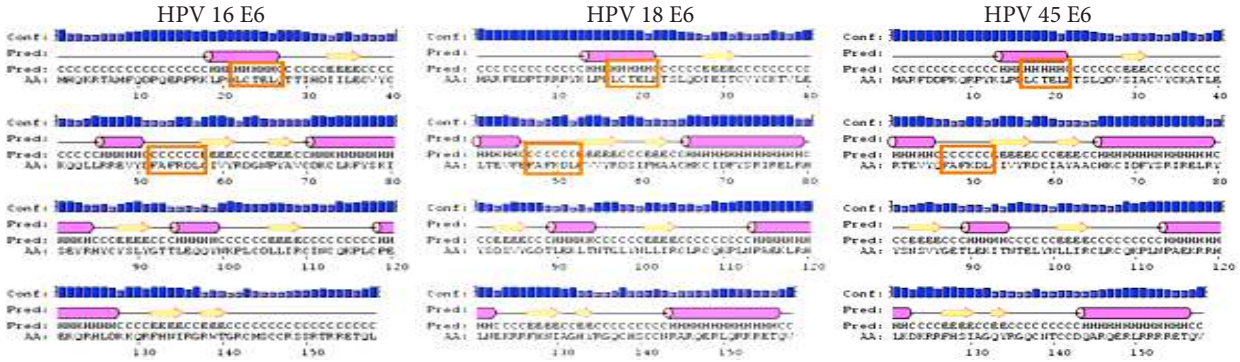

Figure 3. Prediction of Secondary Structure by PSIPred

Table 4. Prediction of Linear B-cell Epitopes by ABCpred

Rank	Sequence	Start position	Score
HPV 16 E6			
1	VYDFAFRDLCIVYRDS	49	0.87
2	RDLCIVYRDGNPYAVC	55	0.86
2	FHNIRGRWTGRCMSCC	132	0.86
3	TAMFQDPQERPRKLPQ	6	0.85
4	YRDGNPYAVCDKCLKF	61	0.84
5	LKFYSKISEYRHYCYS	74	0.83
5	RWTGRCMSCCRSSRTR	138	0.83
6	YAVCDKCLKFYSKISE	67	0.82
7	QTTIHDIILECVYCKQ	27	0.79
8	ISEYRHYCYSLYGTTL	80	0.78
8	KPLCDLLIRCINCQKP	101	0.78
9	ERPRKLPQLCTELQTT	14	0.73
10	QRHLDKKQRFHNIRGR	123	0.72
11	CVYCKQQLLRREVYDF	37	0.69
12	PLCPEEKQRHLDKKQR	116	0.68
HPV 18 E6			
1	FEFAFKDLCIVYRDSI	45	0.89
2	LQDIEITCVYCKTVLE	25	0.86
3	TRRPYKLPDLCTELNT	8	0.82
3	HKCIDFYSRIRELRHY	66	0.82
4	YNLLIRCLRCQKPLNP	99	0.78
4	YGDTLEKLTNTGLYNL	86	0.78
4	IRELRHYSDSVYGDTL	75	0.78
5	CIVYRDSIPHAACHKC	53	0.75
6	YRGQCHSCCNRARQER	134	0.74
7	AEKLRHLNEKRRFHNI	115	0.72
8	NRARQERLQRRRETQV	143	0.65
9	CQKPLNPAEKLRHLNE	108	0.64
10	PDLCTELNTSLQDIEI	15	0.59
11	TCVYCKTVLELTEVFE	31	0.58
12	HNIAGHYRGQCHSCCN	128	0.54
HPV 45 E6			
1	YQFAFKDLCIVYRDCI	45	0.87
2	YGETLEKITNTELYNL	86	0.85
2	SRIRELRYYSNSVYGE	73	0.85
3	HKCIDFYSRIRELRYY	66	0.83
4	MARFDDPKQRPYKLPD	1	0.82
5	CVYCKATLERTEVYQF	32	0.81
5	PDLCTELNTSLQDVSI	15	0.81
6	RRHLKDKRRFHSIAGQ	118	0.8
7	YNLLIRCLRCQKPLNP	99	0.78
8	QRPYKLPDLCTELNTS	9	0.77
9	SIAGQYRGQCNTCCDQ	129	0.74
10	DCIAYAACHKCIDFYS	58	0.72
11	DQARQERLRRRRETQV	143	0.69
12	TLERTEVYQFAFKDLC	38	0.67
12	QKPLNPAEKRRHLKDK	109	0.67

Table 5. Prediction of Linear B cell Epitopes by BCPreds

Amino acid position	Epitope sequence	BCPreds score
HPV 16 E6		
38	VYCKQQLLRREVYDFAFRDL	0.967
59	IVYRDGNPYAVCDKCLKFYS	0.889
7	AMFQDPQERPRKLPQLCTEL	0.833
117	PEEKQRHLDKKQRFHNIRGR	0.791
HPV 18 E6		
41	LTEVFEFAFKDLFVVYRDS	0.998
11	PYKLPDLCTELNTSLQDIEI	0.895
HPV 45 E6		
43	EVYQFAFKDLCIVYRDCIA	1
79	RYYSNSVYGETLEKITNTEL	0.889
107	RCQKPLNPAEKRRHLKDKRR	0.882
6	DPKQRPYKLPDLCTELNTSL	0.838

Discussion

Prediction of immunogenic epitopes for HPV remains vital and challenging task using bioinformatics tools. Though there are various types of high risk HPVs, HPV 16, HPV 18 and HPV 45 are responsible for 94% of cervical adenocarcinomas. In this work we have made an attempt to predict the promiscuous epitopes among the E6 proteins of high risk HPVs. To achieve this task, we used different types of highly accurate bioinformatics tools and retrieved a huge amount of data and arrive at an interesting result. Our work focused on identifying the conserved residues, T-cells (HLA class I and HLA class II) and B-cells epitope and their corresponding secondary structure information. From the multiple sequence alignment results, the two sequences namely KLPD(Q)LCTEL and FAFK(R)DL appeared as most conserved residues with only one residue differing in their fourth positions. The FAFK(R)DL sequence varies with Arginine in HPV 16 E6 protein and lysine in HPV 18 and HPV 45 E6 proteins. The KLPD(Q)LCTEL fragment varies with glutamine in HPV 16 E6 protein and aspartic acid in HPV 18 and HPV 45 E6 proteins. The varying residue in HPV 16 E6 protein is specified in parenthesis. Most surprisingly, the same fragments FAFK(R)DL and KLPD(Q)LCTEL were predominantly restricted by different alleles in all the high risk HPVs. Also, the presence of hydrophobic residues, alanine and phenylalanine in FAFK(R)DL fragment, leucine and proline in KLPD(Q)LCTEL fragment at second and third positions supports that these fragments can be considered as promiscuous epitopes since TAP

transportation always favors a hydrophobic residues at second and third positions. As an experimental support to our prediction, Nakagawa et al. 2004; Nakagawa et al. 2007 reported a region from 52 to 61 (FAFKDLCIVY) of HPV 16 E6 protein and Rudolf et al. (2001) reported KLPDLCTEL of HPV 18 E6 to be an endogenously processed T-cell epitope. As an extension, we were interested to find whether same epitopes are present other high risk HPV types. The same epitopes were also found in HPV 45 E6 protein. The secondary structure prediction also reveals that the fragment FAFK(R)DL preferred to be in coil region and KLPD(Q)LCTEL to be in helix region which are most important qualities for the consideration to become promiscuous epitopes.

There are major drawbacks in recent years which prevent the scientific community for an effective treatment against HPV (Brusic et al., 2004; Flower et al., 2010). Therapeutic vaccines play a significant role in preventing the metastatic spread of the disease with immediate impact rather than prophylactic vaccines which normally takes many years to reduce deaths from this disease (Frazer, 2004). Also, the prophylactic vaccine development requires very long duration. Thus, identifying promiscuous epitopes for therapeutic vaccine development could only be the possible way for new findings (Aidan et al., 2009; Stanley, 2009). Hence, we conclude that the fragments FAFK(R)DL followed by KLPD(Q)LCTEL should be considered as most promiscuous epitopes among E6 proteins of high risk HPVs - HPV 16, HPV 18 and HPV 45.

Acknowledgements

The authors thank the management of VIT University for providing the facilities and encouragement to carry out this work. The authors would like to express their heartfelt gratitude to the anonymous reviewers for their intuitive suggestions.

References

Ackerman AL, Cresswell P (2004). Cellular mechanisms governing cross-presentation of exogenous antigens. *Nature Immunology*, **5**, 678-84.

Aidan M, Ulrich K, Charles, et al (2009). T-cell epitope prediction: rescaling can mask biological variation between MHC molecules. *PLoS Comput Biol*, **5**, 1000327.

Agarwal SM, Raghav D, Singh H, et al (2011). CCDB: a curated database of genes involved in cervix cancer. *Nucleic Acids Res*, **39**, 975-9.

Agnieszka K, Grabowska, Riemer AB (2012). The invisible enemy-how human papillomaviruses avoid recognition and clearance by the host immune system. *Open Virol J*, **6**, 249-56.

Alessia M, Marco T, Nello C (2009). Support vector machines. wiley interdisciplinary reviews: computation stat. *Computational Statistics*, **1**, 283-9.

Altschul S, Madden T, Schaffer A, et al (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl Acids Res*, **25**, 3389-402.

Androphyl EJ, Hubbert NL, Schiller JT, et al (1987). Identification of the HPV-16 E6 protein from transformed mouse cells and human cervical carcinoma cell lines. *EMBO J*, **6**, 989-92.

Bernard HU, Burk RD, Chen Z, et al (2010). Classification of Papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. *Virology*, **401**, 70-9.

Bian H, Olson JFR, Hammer J (2003). The use of bioinformatics for identifying class II-restricted T-cell epitopes. *Methods*, **29**, 299-309.

Boccardo E, Lepique AP, Luisa L (2010). The role of inflammation in HPV carcinogenesis. *Carcin*, **31**, 1905-12.

Brusic V, Bajic VB, Petrovsky N (2004). Computational methods for prediction of T-cell epitopes-a framework for modeling, testing, and applications. *Method*, **34**, 436-43.

Chandra S, Singh TR (2012). Linear B cell epitope prediction for epitope vaccine design against meningococcal disease and their computational validations through physicochemical properties. *Netw Model Anal Health Inform Bioinforma*, **1**, 153-9.

Cole ST, Danos O (1987). Nucleotide sequence and comparative analysis of the human papillomavirus type 18 genome. Phylogeny of papillomavirus and repeated structure of the E6 and E7 gene products. *J Mol Bio*, **193**, 599-608.

De Groot A S (2004). Immunome derived vaccine. *Expert Opin Biol Ther*, **4**, 767-72.

Delius H, Hofmann B (1994). Primer-directed sequencing of human papillomavirus types. *Curr Top Microbial Immunol*, **186**, 13-31.

De Sanjose S, Quint WGV, Alemany L, et al (2010). Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. *Lancet Oncol*, **11**, 1048-56.

Dimitrov I, Garnev P, Flower DR (2010). EpiTop-a proteochemometric tool for MHC class II binding prediction. *Bioinformatics*, **26**, 2066-8.

Doytchinova IA, Guan P, Flower DR (2006). EpiJen: a server for multistep T cell epitope prediction. *BMC Bioinformatics*, **7**, 131.

Engelhard VH (1994). How cells process antigens. *Sci Am*, **271**, 54-61.

Fakhry C, Gillison ML (2006). Clinical implications of human papillomavirus in head and neck cancers. *J Clin Oncol*, **24**, 2606-11.

Flower DR, Phadwal K, Macdonald IK, et al (2010). T-cell epitope prediction and immune complex simulation using molecular dynamics: state of the art and persisting challenges. *Immunome Res*, **6**, 2-4.

Frazer IH (2004). Prevention of cervical cancer through papillomavirus vaccination. *Nature Rev Immun*, **4**, 46-54.

Gallagher KME, Man S (2007). Identification of HLA-DR1- and HLA-DR15-restricted human papillomavirus type 16 (HPV16) and HPV18 E6 epitopes recognized by CD4+ T cells from healthy young women. *J General Virology*, **88**, 1470-8.

Gnanamony M, Peedicayil A, Abraham P (2007). An overview of Human Papillomavirus and current vaccine strategies. *Indian J Med Microbiol*, **25**, 10-17.

Govan VA (2008). A novel vaccine for cervical cancer: quadrivalent human papillomavirus (types 6, 11, 16 and 18) recombinant vaccine (Gardasil). *Ther Clin Risk Manage*, **4**, 65-70.

Hakenberg J, Nussbaum A, Schild H, et al (2003). MAPPP -MHC-I antigenic peptide processing prediction. *Appl Bioinformatics*, **2**, 155-8.

Hellberg S, Sjostrom M, Skagerberg B, et al (1987). Peptide quantitative structure-activity relationships, a multivariate approach. *J Med Chem*, **30**, 1126-35.

Holzhutter HG, Frommel C, Kloetzel PM (1999). A theoretical approach towards the identification of cleavage-determining amino acid motifs of the 20 S proteasome. *J Mol Bio*, **286**,

1251-65.
Iurescia S, Fioretti D, Fazio VM, et al (2012). Epitope-driven DNA vaccine design employing immunoinformatics against B-cell lymphoma: a biotech's challenge. *Biotechnol Adv*, **30**, 372-83.
Jacob AG, Hansen NC, Vinther J, et al. (2003). A promoter within the E6 ORF of human papillomavirus type 16 contributes to the expression of the E7 oncoprotein from a monocistronic mRNA. *J Gen Virol*, **84**, 3429-41.
Jones DT (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol*, **292**, 195-202.
Kavitha KV, Saritha R, Vinod Chandra SS (2013). Computational methods in linear b-cell epitope prediction. *Int J Comput Appl*, **63**, 28-32.
Khan HA, Arif IA, Bahkali AH, et al (2008). Bayesian, maximum parsimony and UPGMA models for inferring the phylogenies of antelopes using mitochondrial markers. *Evol Bioinform Online*, **4**, 263-70.
Koichiro T, Joel D, Masatoshi N, et al (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol*, **24**, 1596-9.
Larkin MA, Blackshields G, Brown NP, et al (2007). Clustal w and clustal X version 2.0. *Bioinformatics*, **23**, 2947-48.
Lapinsh M, Prusis P, Gutcaits A, et al (2001). Development of proteochemometrics: A novel technology of use for analysis of drug-receptor interactions. *Biochim Biophys Acta*, **1525**, 180-90.
Lin SY, Cheng CW, Su ECY (2013). Prediction of B-cell epitopes using evolutionary information and propensity scales. *BMC Bioinformatics*, **14**, 2-10.
Longworth MS, Laimins LA (2004). Pathogenesis of human papillomaviruses in differentiating epithelia. *Microbiol Mol Biol R*, **68**, 362-72.
Lundegaard C, Lamberth K, Harndah M, et al (2008). NetMHC-3.0: Accurate web accessible predictions of human, mouse, and monkey MHC class i affinities for peptides of length 8-11. *Nucleic Acids Res*, **36**, 509-12.
Ma B, Xu Y, Hung CF, et al (2010). HPV and Therapeutic Vaccines: Where are we in 2010? *Curr Cancer Ther Rev*, **6**, 81-103.
Ma B, Roden R, Wu TC (2010). Current status of HPV vaccines. *J Formos Med Assoc*, **109**, 481-3.
Mark HE, Suzanne L, Lydia GC (2009). Genetic variants in TAP are associated with high-grade cervical neoplasia. *Clin Cancer Res*, **15**, 1019-23.
Morrow MP, Yan J, Sardesai NY (2013). Human papillomavirus therapeutic vaccines: targeting viral antigens as immunotherapy for precancerous disease and cancer. *Expert Rev Vaccines*, **12**, 271-83.
Munger K, Howley PM (2002). Human papillomavirus immortalization and transformation functions. *Virus Res*, **89**, 213-28.
Nakagawa M, Kim KH, Moscicki AB (2004). Different methods of identifying new antigenic epitopes of human papillomavirus type 16 E6 and E7 Proteins. *Clin Diagn Lab Immun*, **11**, 889-96.
Nakagawa M, Kim KH, Gillam TM, et al (2007). HLA class I binding promiscuity of the CD8 T-Cell epitopes of human papillomavirus type 16 E6 protein. *J Virol*, **81**, 1412-23.
Nielsen M, Lundegaard C, Woming P, et al (2003). Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci*, **12,** 1007-17.
Nielsen M, Lundegaard C, Woming P, et al (2004). Improved prediction of MHC class I and II epitopes using a novel Gibbs sampling approach. *Bioinformatics*, **20**, 1388-97.
Oyarzun P, Ellis JJ, Boden M, et al (2013). PREDIVAC: CD4+ T-cell epitope prediction for vaccine design that covers 95% of HLA class II DR protein diversity. *BMC Bioinformatics*, **14**, 52.
Patronov A, Doytchinova I (2013). T-cell epitope vaccine design by immunoinformatics. *Open Biol*, **3**, 120139.
Peters B, Sidney J, Bourne P, et al (2005). The design and implementation of the immune epitope database and analysis resource. *Immunogenetics*, **57**, 326-36.
Pingping S, Wenhan C, Yanxin H, et al (2011). Epitope prediction based on random peptide library screening: benchmark dataset and prediction tools evaluation. *Molecules*, **16**, 4971-94.
Rudolf MP, Man S, Melief CJM, et al (2001). Human T-Cell responses to HLA-A-restricted high binding affinity peptides of human papillomavirus type 18 proteins E6 and E7. *Clin Cancer Res*, **7**, 788-95.
Saha S, Raghava GPS (2006). Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins*, **65**, 40-8.
Seedorf K, Krammer G, Durst M, et al (1985). Human papillomavirus type 16 DNA sequence. *Virology*, **145**, 181-5.
Sette A, Sidney J, Oseroff C, et al (1993). HLA DR4w4-binding motifs illustrate the biochemical basis of degeneracy and specificity in peptide-DR interactions. *J Immunol*, **151**, 3163-70.
Shehzadi A, Rehman SU, Idrees M, (2011). Promiscuous prediction and conservancy analysis of CTL binding epitopes of HCV 3a viral proteome from Punjab Pakistan: an In-silico approach. *Virol J*, **8**, 55.
Smith KL, Tristram A, Gallagher KM, et al (2004). Epitope specificity and longevity of a vaccine-induced human T cell response against HPV18. *Int Immunol*, **17**, 167-76.
Soliman PT, Slomovitz BM, Wolf JK (2004). Mechanisms of cervical cancer. *Drug Discovery today: Disease mechanism*, **1**, 253-8.
Sirskyj D, Mitoma FD, Golshani A, et al (2011). Innovative bioinformatic approaches for developing peptide-based vaccines against hypervariable viruses. *Immunol Cell Biol*, **89**, 81-9.
Sudandiradoss C, George Priya Doss C, Rajasekaran R, et al (2008). Analysis of binding residues between scorpion neurotoxins and D2 dopamine receptor: a computational docking study. *Comput Biol Med*, **38**, 1056-67.
Stanley MA (2009). Immune responses to human papilloma viruses. *Indian J Med Res*, **130**, 266-76.
The Universal protein resource (Uniprot) in 2010. *Nucleic Acids Res*, **38**, 142-8.
Tomar N, De RK (2010). Immunoinformatics: an integrated scenario. *Immunol*, **131**, 153-68.
Tong JC, Tan TW, Shoba R (2006). Methods and protocols for prediction of immunogenic epitopes. *Briefings in Bioinformatics*, **8**, 96-108.
Tumban E, Peabody J, Peabody DS, et al (2011). A pan-HPV vaccine based on bacteriophage PP7 VLPs displaying broadly cross-neutralizing epitopes from the HPV minor capsid protein, L2. *PLoS One,* **6**, 23310.
Waldmann TA (2003). Immune responses to human Papilloma viruses. *Nat Med*, **9**, 269-77.
Walshe VA, Hattotuwagama CK, Doytchinova IA, et al (2009). Integrating *in silico* and *in vitro* analysis of peptide binding affinity to HLA-Cw*0102: a bioinformatic approach to the prediction of new epitopes. *PLoS One*, **4**, 8095.
Whelan S, Lio P, Goldman N (2001). Molecular phylogenetics: state-of-art methods for looking into the past. *Trends Genet*, **17**, 262-72.
Wu CY, Monie A, Pang X, Hung CF, Wu TC (2010). Improving therapeutic HPV peptide-based vaccine potency by enhancing CD4+ T help and dendritic cell activation. *J*

Biomed Sci, **17**, 88-97.
Yao B, Zheng D, Liang S, et al (2013). Conformational B-Cell epitope prediction on antigen protein structures: a review of current algorithms and comparison with common binding site prediction methods. *PLoS One*, **8**, 62249
Yasser ELM, Dobbs D, Honavar V (2008). Predicting linear B-cell epitopes using string kernels. *J Mol Recognit*, **21**, 243-55
Zhang W, Niu Y (2010). Presented at 3rd International conference on biomedical engineering and informatics. *BMEI*, **10**, 5640578.