RESEARCH COMMUNICATION

Prediction of Epitopes and Structural Properties of Iranian HPV-16 E6 by Bioinformatics Methods

Hassan Mohabatkar

Abstract

HPV-16 is the HPV most often linked to cervical carcinoma. E6 of the HPV-16 which expressed early in cancer cells is a target for immune therapeutic methods. In the present study, after fetching the sequence of HPV-16 E6 (accession No: ABC48950) from NCBI databank, by using hydrophilicity, flexibility, accessibility, turns, exposed surface, polarity and antigenic propensity scales, B cell epitopes of the protein were predicted. In addition, MHCPred version 2.0 program was used to predict MHC Class I and Class II alleles. The sequences of the epitopes were also found out. According to this computer-based prediction the results from A0203 and DRB0101 reveal lower IC50 than other alleles. For A0203 allele, peptide with the best binding affinity was 25ELQTTIHDI33. For DRB0101 allele, the peptide was 39YCKQQLLRR47. Different structural features of the protein were also predicted. These features were including glycosylation, kinase C phosphorylation, Casein kinase II phosphorylation and N-myristylation sites, and disulfide bonding states. By using these computational scales and programs, 0 glycosylation, 3 kinase C phosphorylation, 2 casein kinase II phosphorylation and 1 N-myristylation sites and 2 disulfide bonds were predicted. Development and approval of new vaccines are keys for control of cancer. Epitopes and structural features of proteins can be predicted and this information can help us in molecular and medical studies of viruses.

Key Words: Bioinformatics - computational analysis - structure prediction - T cell epitopes - B cell epitopes

Asian Pacific J Cancer Prev, **8**, 602-606

Introduction

Cervical cancer, the second most common malignancy in women worldwide, is almost invariably associated with infection by human papillomavirus (HPV) (Kaufmann et al., 2002; Kanjanavirojkul et al., 2006). In developing countries, cervical cancer is often the most common cancer in women and may constitute up to 25% of all female cancers (Harro 2001). Human papillomavirus type 16 (HPV-16) has been known as a major causative factor for the development of cervical carcinomas (Kim et al., 2004). The early proteins E6 and E7 of the cancer-related HPV-16 are constitutively expressed in cancer cells thus are targets for immune therapeutic approaches. HPV E6 is consistently expressed and is responsible for the malignant transformation of HPV-associated lesions. Thus, this protein represents an ideal target for therapeutic HPV vaccine development (Peng et al., 2005). E6 is a very important early protein of HPV and expression of this oncoprotein is widely considered to be responsible for the transforming ability of the virus (Liu et al., 2002). Whereas previous studies have mainly focused on the immunogenicity of E7 protein, little is presently known about E6 (Samorski et al., 2006).

A key step in the design of subunit vaccines is identification of epitopes by the overlapping synthetic peptides. This method decreases the possibility of missed epitopes, but lots of peptides need to be synthesized, at a high cost. Immunoinformatics, a newly emergent branch of bioinformatics, has already become a familiar and useful tool for selecting epitopes from immunologically relevent proteins, as well as for the further development of information about different epitopes. Epitopes are selected by prediction with software, which saves the expense of synthetic peptides and working time (Bian et al., 2003; Li et al., 2005).

Basically, the recognition of antigenic epitopes by the immune system, either small discrete T-cell epitopes or large conformational epitopes recognized by B cells and soluble antibodies is the key molecular event at the heart of the immune response to pathogens (Doytchinova and Flower, 2002). The objective of this study was computer-aided prediction of i) B cell epitopes, ii) major histocompatibility complexes (MHCs) alleles and iii) post translational modifications of Iranian HPV-16 E6. Although, perfect predictions with accuracy of 100% are not achievable (Hansen et al., 1996), the prediction of protein secondary structure is an important step in assessment of tertiary structure (Kim and Park, 2003), and because antigenicity of a protein depends on its secondary structure, in this study the secondary structure of HPV-16 E6 also was predicted.

Department of Biology, College of Sciences, Shiraz University, Shiraz, Iran E-mail: mohabatkar@susc.ac.ir Fax: +98711-2280926

```
.....10.....20.....30.....40.....50.....60.....70.....
MHQKRTAMFQDPQERPGKLPQLCTELQTTIHDIILECVYCKQQLLRREVYDFAFRDLCIVYRDGNPYAVCDKCLKFYSK

80.....90.....100.....110.....120.....130.....140.....150.....
ISEYRHYCYSVYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLDKKQRFHNIRGRWTGRCMSCCRSSRTRRETQL
```

Figure 1. Sequence of E6 Protein of Iranian Human Papillomavirus Type 16

Patients and Methods

Amino acid Sequence

The sequence of Iranian HPV-16 E6 was obtained from NCBI databank. The accession number was ABC48950. The isolation source of the virus was cervical cancer tissue.

B cell epitope prediction

All prediction calculations were based on propensity scales for each of the 20 amino acids. The amino acid sequence of each protein was read as a moving window. In order to compare the profiles obtained by different methods, various scales were normalized where the original values of each scale were set between +3 and -3. Hydrophilicity (Parker et al., 1986), flexibility (Karplus and Schulz, 1985), accessibility (Emini et al., 1985), turns (Pellequer et al., 1993), exposed surface (Kolaskar and Tongaonkar, 1990), polarity (Janin and Wodak, 1978) and antigenic propensity (Ponnuswamy et al., 1980) scales were applied to predict B cell epitopes.

T cell epitope prediction

MHCPred version 2.0 (Guan et al., 2003 and Guan et al., 2006) a server for quantitative prediction of MHC binding peptides has been used. This server is available from the URL: http://www.jenner.ac.uk/MHCPred. MHCPred runs as a CGI sever, written in Perl, operating under Microsoft Windows NT. The sequence of a protein is entered, MHC allele and affinity threshold are selected and the program run. MHCPred covers a range of different human MHC allele peptides specificity models. These include Class I alleles (HLA-A*0101, HLA-A*0201, HLA-A*0202, HLA-A*0203, HLA-A*0206, HLA-A*0301, HLA-A*1101, HLA-A*3301, HLA-A*6801, HLA-A*6802 and HLA-B*3501) and Class II alleles (HLA-DRB*0401, HLA-DRB*0401 and HLA-DRB*0701).

Prediction of post-translational modifications

Different scales were used to predict glycosylation, N-myristoylation, protein kinase C phosphorylation, and casein kinase II phosphorylation sites and disulfide bonds (Vullo and Frasconi, 1994; Bairoch et al, 1997; Hubbard and Ivatt, 1981; Bause, 1983). N-glycosylation sites are searched as Asn-X-Thr or Asn-X-Ser sequences, where X is any residue.

Secondary structure prediction

A scale of secondary structure, which was based on the prediction of turns and loops obtained from statistical analysis of proteins of known structure, was considered for secondary structure prediction (Garnier et al., 1978).

Results

The amino acid sequence of the protein of the study is shown in figure 1. The length of the protein was 158 residues.

The percentage of different amino acids in the protein was calculated (table 1). The most prevalent amino acid was Arginine (17 residues), followed by Leucine and Cysteine respectively. The least amino acid was tryptophan, followed by Adenine and Methionine.

Hydrophilicity, flexibility, accessibility, turns, exposed surface, polarity and antigenic propensity scales were applied to predict B cell epitopes. Results of B cell epitope prediction has been shown in table 2.

Above-mentioned parameters have been correlated with the location of continuous epitopes. As a result, 5 regions predicted to be B-cell epitopes. The shortest epitope was epitope number 1 (7 residues), and the longest one was epitope number 4 (12 residues).

A0101, A0201, A0202, A0203, A0206, A0301, A1101, A3101, A6801, A6802, B3501, DRB0101, DRB0401 and DRB0701 were the alleles chosen for this computation analysis. Peptides with the lowest predicted IC50, corresponding to the best predicted binding affinities are shown in table 3. According to this computer-based prediction the results from A0203 and DRB0101 reveal lower IC50 than other alleles. For A0203 allele, the three peptides with the best binding affinities are 25ELQTTIHDI33 (IC50 = 2.88), 21QLCTELQTT29 (IC50 = 3.58) and 94TLEQQYNKP102 (IC50 = 6.28), respectively. For DRB0101 allele, the three peptides with the best binding affinities are 39YCKQQLLRR47 (IC50 = 1.03), 99YNKPLCDLL107 (IC50 = 1.07) and 91YGTTLEQQY99 (IC50 = 1.69), respectively.

Table 1. Residue Compositions for Iranian HPV-16 E6 Proteins

%A: 1.9	%C: 8.9	%D: 5.1	%E: 5.7	%F: 3.2
%G: 3.2	%H: 3.2	%I: 5.1	%K: 7.0	%L: 9.5
%M: 1.9	%N: 2.5	%P: 4.4	%Q: 8.2	%R: 10.8
% S: 3.8	%T: 5.7	%V: 3.2	%W: 0.6	%Y: 6.3

Table 2. Amino Acid Sequences of Iranian HPV-16 E6 Predicted to be B Cell Epitope

No.	Domains of Iranian HPV-16 E6 predicted to be B cell epitope
1	11DPQERPG17
2	80ISEYRHYC87
3	95LEQQYNKP102
4	121EKQRHLDKKQRF132
5	145SCCRSSRTRRE155

Table 3. Peptides of Iranian HPV-16 E6 with the Best Predicted Binding Affinities for Each Allele

No.	Alleles	Peptides	IC50 values
1	A0101	80ISEYRHYCY88	35.89
2	A0201	99YNKPLCDLL107	38.90
3	A0202	106LLIRCINCQ114	32.14
4	A0203	25ELQTTIHDI33	2.88
5	A0206	41KQQLLRREV49	8.22
6	A0301	107LIRCINCQK115	25.41
7	A1101	93TTLEQQYNK101	8.32
8	A3101	116PLCPEEKQR124	83.75
9	A6801	52FAFRDLCIV60	6.89
10	A6802	28TTIHDIILE36	19.68
11	B3501	52FAFRDLCIV60	177.83
12	DRB0101	39YCKQQLLRR47	1.03
13	DRB0401	38VYCKQQLLR46	74.99
14	DRB0701	41KQQLLRREV49	15.38

Results of computer-assisted prediction of the number of glycosylasion, Phosphorylation, myristoylation, and disulfide sites are shown in table 4. According to this analysis no asparagines was predicted to be glycosylated. Results also showed that 3 residues were predicted to be kinase C phosphorylated, 2 residues were predicted to be casein kinase II phosphorylated and 1 glycine was predicted to be myristoylated. This computational study predicted that B cell epitope No. 5 was kinase C phosphorylated (sites no. 149 and 152). None of the putative epitopes was predicted to be myristoylated or glycosylated. Putative MHC-I epitopes number 4 and 7 were predicted to be Casein kinase II phosphorylated. This computational study predicted that neither MHC-I nor MHC-II epitopes were myristoylated.

The result of this analysis for Iranian HPV-16 E6 indicates that 3 regions predicted to be a-helix and 6 regions predicted to be b-sheet (Table 5). According to this analysis 16.5 %, 20.3 % and 46.8 % of the protein was in the a-helix, b-sheet and turn forms respectively. Most parts of the protein (63.92%) were exposed to the solvent.

Discussion

The aim of this investigation was to apply bioinformatics methods to study the B and T cell epitopic sites and some other structural properties of Iranian HPV-16 E6. HPV-16 is the HPV most frequently associated with cervical carcinoma in humans. For the prevention or treatment of cervical carcinoma, the E6 and E7 oncoproteins appear to be good targets for vaccine-induced

Table 4. Molecular Characterization of Iranian HPV-16 E6. PKC sites are the number of Protein kinase C phosphorylation sites, CK2 sites are the number of casein kinase II phosphorylation site, Myr sites are the number of N-myristoylation sites and Dis sites are the number of disulfide bonds.

Character	Gly. Sites	PKC Sites	CK2 Sites	Myr Sites	Dis Sites
Number	0	3	2	1	2

Table 5. Prediction of Secondary Structure and Accessibility Type of Iranian HPV-16 E6

No. helices	No. sheets	Helices*	Sheets*	Turns*	Accessibility type* Burred	Exposed
3	6	16.5	20.3	46.8	36.1	63.9

* Percentage Value

cytotoxic T lymphocytes (Bourgault Villada et al., 2000). An important factor in the development of cervical neoplasia is the role of HPV variants (Giannoudis and Herrington, 2001). HPV variants differ in biological and chemical properties and pathogenicity (Conrad-Sto¨ppler, et al., 1996, Veress et al., 1999). The oncogenicity of specific HPV variants appears to vary geographically and also with the ethnic origin of the population. Intratypic sequence variation has been found in the E2, E4, E5, E6, and E7 genes of HPV-16, and this variation can be of functional significance. Recently, screening strategies, effective therapeutic and preventive vaccines are developing that have the potential to contribute significantly to the control and prevention of cervical cancer (Burd, 2003). To date, there is no experimentally determined structural information, specifically, regarding Iranian HPV-16 E6, so this investigation has been conducted.

In a protein, antigenic determinants lie in regions which are hydrophilic, exposed and polar, and accessibility and flexibility of these segments are high. This has led to the rules that would allow the position of B-cell epitopes to be predicted from these features of the protein sequence (Pellequer et al., 1991, Mohabatkar and Kar, 2004). Further more, there is a need to understand the structural and immunological basis for the Cell-Mediated Immune response to help the formulation of therapeutic vaccines (Sarkar et al., 2005). The perfect bioinformatics prediction of T-cell epitopes can to a great extent reduce the experimental cost in candidate epitope identification. In the present study MHCPred program has been used to predict MHC Class I and Class II alleles of HPV 16 E6, and it gives information to design new experiments (Chen et al., 2006).

There are a number of studies on the effects of modifications of amino acid residues on the functions of proteins. For example, in Alzheimer disease, through study of effects of phosphorylation of epitopes investigators found out that epitopes require the phosphorylation of some residues to be effective (Goedert et al., 1994). As this computational study predicted that B-cell epitope number 5 and MHC-I epitope number 4 and 7 were phosphorylated, they can be good epitopes. Although HPV-16 E6 was not predicted to be glycosylated and none of the potential epitopes of Iranian HPV-16 E6 was myristoylated, in a research on Nef protein of HIV-1, it was shown that factors which inactivate the myristoylation sites significantly enhance the specific T cell response (Liang et al., 2002). The powerful effect of carbohydrates on the antigenicity of proteins has been proved for glycosylated sites too (Huang et al., 1997). Warnock et al (1993) has explained that post-translational phosphorylation and acetylation of the tumor suppressor

protein p53, elicit important effects on the function and the stability of the resultant protein. However, as phosphorylation and acetylation are dynamic events subject to complex controls, elucidating the relationships between phosphorylation and acetylation is difficult. Additionally, in a research on breast cancer, Schuman et al (2003) has discussed about the possible roles that peptide epitope secondary structure and glycosylation state may play in mucin tumor immunogenicity.

Since the residue composition of any protein is important, in the present investigation the residue composition for Iranian HPV-16 E6 protein has been calculated. In a previous study on HPV-6 E7, it was shown that single amino acid substitutions in low-risk HPV enhanced features of the high-risk HPV E7 oncoproteins (Sang and Barbosa, 1992).

In protein, turns are located on the surface; these parts are accessible and hydrophilic. In contrast, the core mostly devoid of water molecules (Pellequer et al., 1991). In a research on herpes simplex virus type I, it has been shown that the secondary structure is very important for antibody binding and even a minor modification of the secondary structure can affect the immune identification of antigens (Schlosser et al., 2003). Like any other protein, prediction of secondary structure of Iranian HPV-16 E6 can provide us important information about the interactions and functions of this protein.

In conclusion, identification of epitopes is crucial in understanding the rules of B and T cell activation and designing of synthetic vaccines. Identification of these epitopes has paved a way towards cancer immunotherapy and identification of many other infectious diseases. This kind of studies can help the developing of the experimental methodologies, by omitting non functional sequences.

Acknowledgment

Support of this study by Shiraz University is acknowledged.

References

Bairoch A, Bucher P, Hofmann K. (1997). The PROSITE database, its status in 1997. *Nuc Aci Res*, **25**, 217-23.

Bause E (1983). Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes. *Biochem J*, **209**, 331-6.

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

Bourgault Villada I, Bénéton N, Bony C, et al (2000). Identification in humans of HPV-16 E6 and E7 protein epitopes recognized by cytolytic T lymphocytes in association with HLA-B18 and determination of the HLA-B18-specific binding motif. *Eur J Immunol*, **30**, 2281-9.

Burd EM, (2003). Human Papillomavirus and Cervical Cancer. *Clinical Microbiology Reviews*, **16**, 1–17.

Conrad-Stöppler MCK, Ching H, Stöppler K, et al (1996). Natural variants of the human papillomavirus type 16 E6 protein differ in their abilities to alter keratinocyte differentiation and to induce p53 degradation. *J Virol*, **70**, 6987–93.

Giannoudis A, Herrington CS (2001). Human papillomavirus variants and squamous neoplasia of the cervix. *J Pathol*, **193**, 295–302.

Harro CD, Pang Y-Y S, Roden RBS (2001). Safety and immunogenicity trial in adult volunteers of a human papillomavirus 16 L1 virus-like particle vaccine. *J Natl Cancer Inst*, **93**, 284–92.

Chen CC, Hwang JK, Yang JM (2006). (PS)2: protein structure prediction server. Nuc Aci Res, 34 (Web Server issue), 152-7.

Doytchinova IA, Flower DR (2002). Quantitative approaches to computational vaccinology. *Immunol Cell Biol*, **80**, 270-9.

Emini EA, Hughes JV, Perlow DS, et al (1985). Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol*, **55**, 836-9.

Garnier J, Osguthorpe DJ, Robson B (1978). Analysis and implications of simple methods for predicting the secondary structure of globular proteins. *J Mol Biol*, **120**, 97-120.

Goedert M, Jakes TR, Crowther A, et al (1994). Epitope mapping of monoclonal antibodies to the paired helical filaments of Alzheimer's disease: identification of phosphorylation sites in tau protein. *Biochem J*, **301**, 871-7.

Guan P, Doytchinova IA, Zygouri C, et al (2003). MHCPred: bringing a quantitative dimension to the online prediction of MHC binding. *Appl Bioinformatics*, **2**, 63-6.

Guan P, Hattotuwagama CK, Doytchinova IA, et al (2006). MHCPred 2.0: an updated quantitative T-cell epitope prediction server. *Appl Bioinformatics*, **5**, 55–61.

Hansen JE, Lund O, Neilsen JO, et al (1996). Prediction of the secondary structure of HIV-1 gp120. *Proteins*, **25**, 1-11.

Huang X, Barchi JJ Jr, Lung FD, et al (1997). Glycosylation affects both the three dimensional structure and antibody binding properties of the HIV-1IIIB GP120 peptide RP135. *Biochemistry*, **36**, 10846-56.

Hubbard SC, Ivatt RJ (1981). Synthesis and processing of asparagine-linked oligosaccharides. *Annu Rev Biochem*, **50**, 555-83.

Janin J, Wodak S (1978). Conformation of amino acid side-chains in proteins.*J Mol Biol*, **125**, 357-86.

Kanjanavirojkul N, Pairojkul C, Yuenyao P, et al (2006). Risk Factors and Histological Outcome of Abnormal Cervix with Human Papilloma Infection in Northeastern Thai-women. *Asian Pacific J Cancer Prev*, **7**, 567-70.

Karplus PA, Schulz GE (1985). Prediction of chain flexibility in proteins: A tool for the selection of peptide antigens. *Naturwissenschaften*, **72**, 212-13.

Kaufmann AM, Stern PL, Rankin EM, et al (2002). Safety and immunogenicity of TA-HPV, a recombinant vaccinia virus expressing modified human papillomavirus (HPV)-16 and HPV-18 E6 and E7 genes, in women with progressive cervical cancer. *Clin Cancer Res*, **8**, 3676-85.

Kim H, Park H (2003). Protein secondary structure prediction based on an improved support vector machines approach. *Protein Eng*, **16**, 553-60.

Kim SH, Kim KS, Lee EJ, et al (2004). Human keratin 14 driven HPV 16 E6/E7 transgenic mice exhibit hyperkeratinosis. *Life Sci*, **75**, 3035-42.

Kolaskar AS, Tongaonkar PC (1990). A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett*, **276**, 172-4.

Li GF, Wang Y, Zhang ZS, et al (2005). Identification of immunodominant Th1-type T cell epitopes from Schistosoma japonicum 28 kDa glutathione-S-transferase, a vaccine candidate. *Acta Biochim Biophys Sin (Shanghai)*, **37**, 751-8.

Liang X, Fu T, Xie H, et al (2002). Development of HIV-1 Nef

vaccine components: immunogenicity study of Nef mutants lacking myristoylation and dileucine motif in mice. *Vaccine*, **20**, 3413-21.

Liu Y, Li JZ, Yuan XH, et al (2002). An AP-1 binding site mutation in HPV-16 LCR enhances E6/E7 promoter activity in human oral epithelial cells. *Virus Genes*, **24**, 29-37.

Mohabatkar H, Kar SK (2004). Prediction of exposed domains of envelope glycoprotein in Indian HIV-1 isolates and experimental confirmation of their immunogenicity in humans. *Braz J Med Res*, **36**, 675-81.

Parker JM, Guo D, Hodges RS (1986). New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry*, **23**, 5425-32.

Pellequer JL, Westhof E, Regenmortel MHV (1991). Predicting location of continuous epitopes in proteins from their primary structures. *Methods Enzymol*, **203**, 176-201.

Pellequer JL, Westhof E, Van Regenmortel MH (1993). Correlation between the location of antigenic sites and the prediction of turns in proteins. *Immunol Lett*, **36**, 83-99.

Peng S, Trimble C, Ji H, et al (2005). Characterization of HPV-16 E6 DNA vaccines employing intracellular targeting and intercellular spreading strategies. *J Biomed Sci*, **12**, 689-700.

Ponnuswamy PK, Prabhakaran M, Manavalan P (1980). Hydrophobic packing and spatial arrangement of amino acid residues in globular proteins. *Biochim Biophys Acta*, **623**, 301-16.

Samorski R, Gissmann L, Osen W (2006). Codon optimized expression of HPV 16 E6 renders target cells susceptible to E6-specific CTL recognition. *Immunol Lett*, **7**, 41-9.

Sang BC, Barbosa MS (1992). Single amino acid substitutions in "low-risk" human papillomavirus (HPV) type 6 E7 protein enhance features characteristic of the "high-risk" HPV E7 oncoproteins. *Proc Natl Acad Sci USA*, **89**, 8063-7.

Sarkar AK, Tortolero-Luna G, Follen M, et al (2005). Inverse correlation of cellular immune responses specific to synthetic peptides from the E6 and E7 oncoproteins of HPV-16 with recurrence of cervical intraepithelial neoplasia in a cross-sectional study. *Gynecol Oncol*, **99 (3 Suppl 1)**, S251-61.

Schlosser G, Mezo G, Kiss R, et al (2003). Synthesis, solution structure analysis and antibody binding of cyclic epitope peptides from glycoprotein D of Herpes simplex virus type I. *Biophys Chem*, **106**, 155-71.

Schuman J, Campbell AP, Koganty RR, et al (2003). Probing the conformational and dynamical effects of O-glycosylation within the immunodominant region of a MUC1 peptide tumor antigen. *J Pept Res*, **61**, 91-108.

Veress G, Szarka K, Dong XP, et al. (1999). Functional significance of sequence variation in the E2 gene and the long control region of human papillomavirus type 16. *J Gen Virol*, **80**, 1053–143.

Vullo A, Frasconi P (2004). Disulfide connectivity prediction using recursive neural networks and evolutionary information. *Bioinformatics*, **20**, 653-60.

Warnock LJ, Raines SA, Mee TR, et al (2005). Role of phosphorylation in p53 acetylation and PAb421 epitope recognition in baculoviral and mammalian expressed proteins. *FEBS J*, **272**, 1669-75.