Journal of Virological Methods 177 (2011) 26–30

Contents lists available at ScienceDirect

Journal of Virological Methods

journal homepage: www.elsevier.com/locate/jviromet

Identification of epitopes in Indian human papilloma virus 16 E6: A bioinformatics approach

Ajay Kumar Singh [a], Srikanta Kumar Rath [b], Krishna Misra [a,*]

[a] Centre for Biomedical Magnetic Research, SGPGI Campus, Lucknow, India
[b] Genotoxicity Laboratory, Toxicology Division, Central Drug Research Institute, Lucknow 226001, India

Article history:
Received 26 January 2011
Received in revised form 31 May 2011
Accepted 7 June 2011
Available online 14 June 2011

Keywords:
Human papilloma virus 16
Protein E6
Cervical cancer
Computational analysis
Structure prediction
B cell epitopes

A B S T R A C T

HPV-16 is reported as the cause of cervical and other related carcinomas. The early expressed protein E6 in cancer cells is found to be the target for immune therapeutic methods. The sequence of HPV-16 E6 (Accession No: ABK32509) from NCBI databank has been taken for this study. Hydrophilicity, flexibility, accessibility, turns, exposed surface, polarity and antigenic propensity scales were used for the B cell epitope prediction. MHC Class I and Class II alleles for the accession were predicted by the MHCPred 2.0 Program. The epitope sequences were also found out. Computer-based prediction program results show, A0203 and DRB0101 lower IC50 than other alleles. The best peptide binding affinity was 21HLCTELQTT30 of A0203 allele. In DRB0101 allele the peptide found was 39YCKQQLLRR48. Different structural features of the protein have also been predicted including glycosylation, kinase C phosphorylation, casein kinase II phosphorylation and N-myristylation sites. These computational prediction programs show four glycosylation, five kinase C phosphorylation, two casein kinase II phosphorylation, zero N-myristylation sites and seven disulphide sites. Development and approval of new vaccines are the keys for control of cancer. Epitopes and other structural features of protein prediction could be the best source of information and can help in molecular and medical studies of viral infection and development of HPV associated cancer drugs.

1. Introduction

Cervical cancer, observed and reported worldwide as the second most common malignancy in women, is almost invariably associated with human papilloma virus (HPV) infection (Kaufmann et al., 2002; Kanjanavirojkul et al., 2006). In most of the developing countries 25% of the total female cancer patients have cervical cancer (Harro et al., 2001).

Human papillomavirus type 16 (HPV-16) has been found to be major causative factor for the development of cervical carcinomas (Kim et al., 2004). The two major oncogenic proteins E6 and E7 of HPV-16 are consistently expressed in cancer cells and are major targets for immune therapeutic approaches. It has been found that HPV E6 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). The early oncoprotein E6 expression is responsible for the transforming ability of the virus (Liu et al., 2002). The previous study was focused mainly on the immunogenicity of E7 protein, little is known presently about E6 (Samorski et al., 2006). The epitopes of Iranian HPV 16 E6 protein have been reported (Mohabatkar, 2007). Epitope identification by overlapping synthetic peptides is key step for vaccine development. The present method decreases the possibilities of missed epitopes. However, peptides need to be synthesized even if at high cost. Cell mediated immune responses have also been considered important in the control of HPV infections (Farhat et al., 2009). Immunoinformatics, a new emerging branch of bioinformatics, has already become a familiar and useful tool for selecting epitopes from immunologically relevant proteins, as well as for further development of information about different epitopes. Epitope prediction with software is cost and labour effective and saves the expense of synthetic peptides and working time (Bian et al., 2003; Li et al., 2005). The development of an adenoviral vaccine against E6 and E7 oncoproteins to prevent growth of human papillomavirus has also been reported (Lee et al., 2008). The antigenic recognition of 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 of the immune response to pathogenicity (Doytchinova and Flower, 2002).

Although the host effective vaccine development is still to be targeted, but recently for this purpose the combination of the fowlpox virus and HPV 16 were used for cervical carcinoma in rabbits, which can lead to human cancer as well (Radaelli et al., 2010). This study

* Corresponding author. Tel.: +91 9415247579; fax: +91 5222668215.
E-mail addresses: ajjay2@gmail.com (A.K. Singh), skrath@cdri.res.in (S.K. Rath), krishnamisra@hotmail.com (K. Misra).

0166-0934/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jviromet.2011.06.006

was focused on Indian HPV 16 E6 by computer aided prediction of (A) B cell epitopes, (B) major histocompatibility complexes (MHCs) alleles and (C) post translational modifications of Indian HPV-16 E6. Earlier, perfect predictions with accuracy of 100% were 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 was also predicted.

2. Materials and methods

2.1. Amino acid sequence

The sequence of Indian HPV-16 E6 was obtained from NCBI databank. The accession number is ABK32509. The isolation source of the virus was cervical cancer tissue.

2.2. B cell epitope prediction

The criteria for all prediction calculations were based on propensity scales for each of the 20 amino acids. The moving window was used for amino acid sequence of each protein. The normalized comparative profiles were obtained by different methods, whereas the original values of each scale were set between +3 and −3. The B-cell epitope prediction was based on 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).

2.3. T cell epitope prediction

For the quantitative prediction of MHC binding peptides MHCPred version 2.0 (Guan et al., 2003, 2006) has been used. This server is available from the URL: http://www.jenner.ac.uk/MHCPred. The program runs as a CGI sever, written in Perl, operating under Microsoft Windows NT.

The sequence of a protein was entered and selections of MHC alleles and affinity threshold were chosen for the run of program. MHCPred covers a range of different human MHC allele peptides specificity models. These analyse Class I alleles (HLA-A*0101, HLA-A*0201, HLAA*0202, HLA-A*0203, HLA-A*0206, HLAA*0301, HLA-A*1101, HLA-A*3301, HLA5A*6801, HLA-A*6802 and HLA-B*3501) and Class II alleles (HLA-DRB*0401, HLADRB*0401 and HLADRB*0701).

2.4. Prediction of post-translational modifications

Different parameters were used to predict glycosylation, N-myristoylation, protein kinase C phosphorylation, casein kinase II phosphorylation sites and disulphide sites (Vullo and Frasconi, 2004; Bairoch et al., 1997; Hubbard and Ivatt, 1981; Bause, 1983). N-glycosylation sites are identified as Asn-Ø-Thr or Asn-Ø-Ser sequences, where Ø is any residue.

Table 1
Results through ProtParam for the analysis of residue ratio.

% A: 1.9	% C: 8.9	% D: 5.1	% E: 5.7	% F: 3.2
% G: 2.5	% H: 3.2	% I: 5.1	% K: 7.0	% L: 10.1
% M: 1.9	% N: 2.5	% P: 4.4	% Q: 7.0	% R: 11.4[a]
% S: 3.8	% T: 5.7	% V: 3.2	% W: 0.6[b]	% Y: 7.0

[a] Maximum percentage residue R (arginine).
[b] Minimum percentage residue W (tryptophan).

2.5. 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).

3. Results

The amino acid sequence of HPV 16 E6 protein with 158 residues is shown in Fig. 1. Percentage of different amino acids in this protein was calculated (Table 1). HPV 16 E6 protein sequence consists of mostly arginine (18 residues), followed by leucine (16 residues) and cysteine (14 residues) respectively. Minimum amino acids were histidine, phenylalanine, valine followed by alanine, methionine and tryptophan. Hydrophilicity, flexibility, accessibility, turns, exposed surface, polarity and antigenic propensity parameters were used to predict B cell epitopes. The results of B cell epitope prediction have been shown in Table 2.

Above-mentioned parameters have been correlated with the location of continuous epitopes. As a result, 6 regions were predicted to be B-cell epitopes. The shortest epitope was epitope number 6 (7 residues), and the longest one was epitope number 2 (35 residues) (Table 3). 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 **21HLCTELQTT29** (IC50 = 2.81), **25ELQTTIHDI33** (IC50 = 2.88) and **48SSRTRRETL157** (IC50 = 4.74), 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.

Results of computer-assisted prediction of the number of glycosylation, phosphorylation, myristoylation, and disulphide sites are shown in Table 4. According to this analysis no asparagines were predicted to be glycosylated. However, number of glycosylated residues was four. Results have also predicted five residues as kinase C phosphorylated, two residues as casein kinase II phos-

```
---------10---------20---------30---------40---------50---------60
MHQKRTAMFQDPQERPRKLPHLCTELQTTIHDIILECVYCKQQLLRREVYDFAFRDLCIV

---------70---------80---------90---------100---------110---------
120
YRDGNPYAVCDKCLKFYSKISEYRYYCYSVYGTTLEQQYNKPLCDLLIRCINCQKPLCPE

---------130---------140---------150--------158
EKQRHLDKKQRFHNIRGRWTGRCMSCCRSSRTRRETLL
```

Fig. 1. Sequence of E6 protein of Indian human papillomavirus type 16.