Computer-aided Analysis of Structural Properties and Epitopes of Iranian HPV-16 E7 Oncoprotein

Fatemeh MOOSAVI, Hassan MOHABATKAR*, Sasan MOHSENZADEH
(Department of Biology, College of Sciences, Shiraz University, Shiraz 71454, Iran)

Received 12 May 2009 / Revised 16 January 2010 / Accepted 18 January 2010

Abstract: Infection by human papillomavirus type 16 (HPV-16) is the cause of 50% or more of cervical cancers in women. The E7 oncoprotein of HPV-16 has long been known as a potent immortalizing and transforming agent. We used different servers like PseAAC, MHC binding, MHC II binding and Expasy for the present computational prediction. The results for T cell epitopes showed that B1501, A0203, A0201, A0202, A0801 and DRB0405 alleles had lower IC50 than other alleles. We also predicted several peptides with the best binding affinities for alleles of the most frequent MHC class I and II alleles of the various ethnic groups living in the different region of Iran. Two peptides (26–35) and (44–52) were predicted as B-cell epitopes. According to this analysis 1 N-glycosylation site, 2 PKC sites, 4 CK2 sites and 3 disulfide sites were predicted. Our computational study predicted that B cell epitope 1 was Casein kinase II phosphorylated (site No. 31) and glycosylated (site No. 29). Putative MHC-I epitopes 3 and 5 and MHC-II epitopes 19, 21 and 26 were predicted to be casein kinase II phosphorylated. MHC-II epitopes 19 and 21 was predicted to be glycosylated. T cell epitopes 1, 13, 16 and 24 were demonstrated to be kinase C phosphorylated. The result of this analysis for Iranian HPV-16 E7 also indicated that 21.43%, 18.37% and 60.20% of the protein were in the α-helix, extended strand and random coil respectively.

Key words: HPV16, E7, epitope, posttranslation modification, bioinformatics, computational analysis, structure prediction.

1 Introduction

Human papillomaviruses (HPVs) are 8 kb, circular DNA viruses that specifically target the basal cells of the epithelial mucosa (Rachel et al., 2009), and they are comprised of more than 100 genotypes. The ability of HPV to transform epithelial cells is divided into high-risk and low-risk types. Low-risk types are associated with benign lesions such as warts, while infections by high-risk types progress to malignant lesions (Ragin et al., 2007). The HPV genome is comprised of several early (E) and late (L) genes, as well as a non-coding region, all of which play roles in viral replication, transcription, and carcinogenesis. The L genes encode L1 and L2 capsid proteins and are transcribed only in productively infected cells. The E genes encode E1, E2, E5, E6, and E7 proteins. E1 and E2 proteins regulate viral replication as well as the expression of other early viral genes. At least 3 proteins (E5, E6, and E7) coded by the high-risk HPVs are considered to be oncogenic, due to their transforming and growth-stimulating properties (Munger et al., 1989; Hubbert et al., 1992; Boyer et al., 1996).

HPV-16 is the most common type among cervical cancer cases worldwide, and about 54% of cervical carcinomas are attributed to the HPV-16 genotype (Munoz et al., 2003). Therefore, the goal of immunization against HPV-induced carcinomas is to activate cellular components of the immune system to recognize and to attack cells infected by HPV (Ohlschlager et al., 2006). Because the HPV-16 E7 oncoprotein is constitutively expressed in cervical cancers, it is considered as a good target for immune therapy of existing lesions and several E7-specific therapeutic vaccines which are currently under development in preclinical models (Brinkman et al., 2005). Computational methods have been used for the prediction of T cell and B cell epitopes (Yu et al., 2002) and in silico T cell and B cell epitope mapping using computational models is developing to be a novel approach to the study of peptide vaccines (de Groot et al., 2001).

Since E7 oncoprotein shows a lot of variation between different types of HPV, here we have studied the structural properties of the only available Iranian E7 protein of HPV 16 specifically. The aim of this study is to predict B cell epitopes, major histocompatibility complexes (MHCs) alleles, post translational modifications.
and secondary structure of the Iranian HPV-16 E7.

2 Methods

2.1 Amino acid sequence and residue composition

The sequence of Iranian HPV-16 E7 was obtained from NCBI databank. The accession number was ABC54573 and the isolation source of the virus was cervical cancer tissue. Residue composition of the Iranian HPV-16 E7 protein was predicted by PseAAC server. This server is accessible at http://chou.med.harvard.edu/bioinf/PseAA/ (Shen et al., 2008).

2.2 Sequence alignment

The E7 sequences were obtained from the NCBI protein database. Accession numbers were as follows: CAA52562 (HPV31), ABR08478 (HPV33), AAA47051 (HPV39), ACR24226 (HPV52), AF478136 (HPV58), ABL96586 (HPV59), ABK32511 (Italy), ACJ66713 (China), AAB70739 (Italy), ABL96586 (Korea), AAO15692 (Thailand), AAD33253 (USA), and AAD83253 (USA). The sequences were aligned with ClustalW.

2.3 B cell epitope prediction

All prediction calculations were based on propensity scales for each of the 20 amino acids. The amino acid sequence of the protein was read as a moving window. Hydrophilicity (Parker et al., 1980), flexibility (Karplus and Schulz, 1985), accessibility (Emini et al., 1986), exposed surface (Janin and Wodak, 1978) and antigenic propensity (Ponnuswamy et al., 1980) scales were applied to predict B cell epitopes. 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.

2.4 T cell epitope prediction


The sequence of a protein was entered, MHC allele and affinity threshold (IC50=500 nM) were selected and the program began to run. A lower number of IC50 indicates higher affinity. Peptides with IC50 values lower than 50 nM were considered as high affinity, lower than 500 nM as intermediate affinity and lower than 5000 nM as low affinity.

2.5 Prediction of post-translational modifications

Different programs were used to predict glycosylation, N-myristoylation, protein kinase C phosphorylation, and casein kinase II phosphorylation sites and disulfide bonds (Baiocco et al., 1997; Vullo and Frasconi, 1994; Bause, 1983; Hubbard and Ivatt, 1981). Asn-X-Thr or Asn-X-Ser sequences were considered as N-glycosylation sites where X was any residue. Expasy available at www.expasy.ch/tools, was used for prediction of post-translational modifications. Predicting the disulfide bond topology in a protein is of crucial importance for the understanding of protein function and can be of great help for tertiary structure prediction methods. The web server http://clavius.bc.edu/~clotelab/DiANNA, a disulfide connectivity prediction server was applied for this purpose (Ferre et al., 2005).

2.6 Secondary structure prediction

A program based on the prediction of turns and loops obtained from statistical analysis of proteins of known structure, was applied for secondary structure prediction (Garnier et al., 1996).

3 Results and Discussion

3.1 Amino acid sequence and residue composition

The percentage of different amino acids in the protein was calculated (Table 1). The most prevalent amino acid was leucine, followed by aspartic acid and glutamic acid respectively. The least amino acid was tryptophan, followed by phenylalanine.

| Amino acid | M | M | L | V | W | A | C | D | E | F | G | H | I | K | L |
|------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Percentage (%) | 4.1 | 7.1 | 10.2 | 9.2 | 5.1 | 4.1 | 1.0 | 5.1 | 1.0 | 4.1 | 5.1 | 1.0 | 11.2 |

Total number of negatively charged residues (Asp + Glu) was 19 and total number of positively charged residues (Arg + Lys) was 5. The molecular formula was C_{468}H_{712}N_{124}O_{161}S_{10}. There was no Trp in the sequence. We also found that there was a great number of Leu, Asp and Glu, which are all polar amino acids. Polar amino acids were mainly localized in the N-terminal half of the protein. These polar amino acids are important because they have the tendency of being on the surface of the protein.

By ClustalW alignment, Iranian E7 oncoprotein of HPV 16 and other high risk HPVs (18, 31, 33, 39, 52 and 58) were compared. Although the overall sizes of the E7 proteins of high risk HPV are rather similar, there is a lot of variation between different types of HPV, and only 66% of the amino acids on average are conserved in these proteins. The lowest and the highest similarities occurred between Iranian E7 and E7 of type 18, and of type 31 respectively.