Linear epitope prediction in HPV type 16 E7 antigen and their docked interaction with human TMEM 50A structural model

Upasna Srivastava1*, Satendra Singh1, Budhyash Gautam1, Pramod Yadav1, Madhu Yadav1, George Thomas2, Gurmit Singh3

1Department of Computational Biology and Bioinformatics, Sam Higginbottom Institute of Agriculture, Technology and Sciences, Allahabad-211007, India; 2Jacob School of Biotechnology and Bioengineering, Sam Higginbottom Institute of Agriculture, Technology and Sciences, Allahabad-211007, India; 3Department of Computer Science and Information Technology, Sam Higginbottom Institute of Agriculture, Technology and Sciences, Allahabad-211007, India; Upasna Srivastava, E-mail: princy149@rediffmail.com; *Corresponding author

Received August 10, 2016; Accepted November 5, 2016; Published May 31, 2017

Abstract: Human Papilloma Virus (HPV) HPV type 16 E7 antigen is a known target in cervical cancer. We report the predicted potential epitopes in the E7 antigen. We further describe the subsequent interaction of these linear epitope peptides with the human TMEM 50 A structural model using molecular docking. This data finds application in the development of components towards HPV associated disease prevention.

Keywords: HPV, Cervical cancer, IEDB, Discovery Studio, Epitopes, TMEM 50

Background: Human papillomavirus (HPV) are a group of viruses associated with various proliferative diseases [1]. HPV types 6, 11, 16, and 18 are the most common and are associated with lesions in the genital tract [2]. Several clinical, molecular, and epidemiological investigations have identified human papillomavirus (HPV) as the major cause of cervical cancer and cervical dysplasia [3]. HPV infection is limited to the epithelium so when the HPV particles have entered host cells, infection is dealt with cell-mediated immunity. Cytotoxic T-lymphocytes (CTLs) recognize foreign peptide antigens presented on the infected target cell surface by molecules of major histocompatibility complex (MHC). In the presence of co-stimulatory molecules binding to a CTL may induce an immune response. Infiltrating T-cells are seen in regressing warts caused by low-risk HPVs. Cervical cancer show tumor infiltrating lymphocytes that are predominantly CTLs [4]. Cervical cancer is also defined as the cancer originating at the opening of the womb, which progressively migrates to the whole of cervix [5]. It is inferred to be hereditary [6] but in 99.7% of cases, is caused by persistent infection with HPV. Some available forms of treatment such as surgery, radiation therapy and chemotherapy are all cytoreductive treatment modalities. Healthy cells are also destroyed in addition to cancer cells in the process. Hence, there is a need to decrease the incidence of cervical cancer and develop better forms for its treatment [7]. Therefore, it is of interest to identify antigenic epitopes in HPV 16 E7 proteins and human TMEM 50A protein using advanced predictive models [8].

Methodology: Workflow:
The workflow for analysis of modeling and dynamics simulation of human TMEM 50A (Figure 1) and generation of putative linear epitopes (Figure 2) are given.

HPV 16 E7 Protein sequence retrieval:
The 98 residues HPV type 16 E7 protein was downloaded from the Protein sequence database of NCBI (NCBI gene ID: 1489079 and NCBI protein id: 9627105).

TMEM 50A sequence:
The 157 residues human Trans membrane (TMEM) protein sequence was downloaded from the protein sequence database of NCBI (Swiss Prot ID O95807). Searching the Protein Data Bank (PDB) using BLASTP [9] identified the structural homolog of the TMEM 50A protein.
Analysis of TMEM 50A
TMHMM 2.0 (http://www.cbs.dtu.dk/services/TMHMM-2.0/) and Target-P 1.1 server (http://www.cbs.dtu.dk/services/TargetP) were used to predict the transmembrane helices in TMEM 50A. Sub-cellular localization of the protein using amino acid composition was completed using PSORT II (http://psort.hgc.jp/cgi-bin/runpsort.pl). NetNGlyc 1.0 Server (http://www.cbs.dtu.dk/services/NetNGlyc/) predicted N-Glycosylation sites using an artificial neural networks (ANN) model. Physiochemical properties such as molecular weight, theoretical pl, total number of negatively (Asp+Glu) and positively (Arg+Lys) charged residues, extinction coefficients, instability index, aliphatic indexed grand average of hydropathicity (GRAVY) of the mature protein were computed using Expasy’s Prot param proteomics server.

Homology modeling for TMEM 50A:
A structural model of TMEM 50A was generated using the template (PDB ID: 3um7A) with Discovery Studio Modeler 3.5 (http://www.accelrys.com) (see Table 1 for details).

The molecular dynamic simulation of TMEM 50A
The Schrodinger software package (https://www.schrodinger.com) was used for the molecular dynamics simulation of the predicted TMEM 50A model.

Active sites prediction of TMEM 50A:
We used Q-SiteFinder to locate binding site. Q-SiteFinder (http://www.bioinformatics.leeds.ac.uk/qsitefinder) uses the interaction energy between the protein and a simple van der Waals probe to locate energetically favorable binding sites. It also generates predicted sites with the lowest average volumes of the cavity. The Active site finder (AADS) [15] was used to identify all cavities in a protein and scores them based on the physicochemical properties of functional groups lining the cavities in the protein. It is a freely available at http://www.scbio-iitd.res.in/dock/ActiveSite_new.jsp.

Computational epitope prediction for HPV type 16 E7 proteins:
The FASTA amino acid sequence of HPV type16 E7 protein was retrieved from Swiss-Prot (Accession No. P03129) at http://expasy.org/sprot/. The protein sequence of HPV 16 E7 was submitted to BCPREDs server (http://ailab.cs.iastate.edu/bcpreds/) and ABCPred online tool available at (http://www.imtech.res.in/raghava/abcpred) [16] for B-cell epitope prediction. IEDB analysis tool was used (http://tools.immuneepitope.org) was used for T-cell epitope prediction. This server covers a broad range of tools facilitating the prediction of new B-cell and T-cell epitopes in proteins of interest, and tools for the analysis of epitopes sets collected from the IEDB. Alleles considered for prediction in this study include HLA-A*01:01, HLA-A*02:01, HLA-DRB1*01:02 and HLA-DRB1*01:01.

Docking of Human TMEM 50A (Human Transmembrane Protein) with predicted T-cell epitopes of HPV type 16 E7 antigens:
The 3D structure prediction of all the HLA allele specific predicted epitopes was completed using the HHIPred server.
Prediction of protein-protein interaction of docked molecules:
InterProSurf was used for protein–protein interactions studies of the docked molecules, which investigated the role of hydrogen bond formation, hydrophobic residues and overall electrostatics with total accessible surface area of the complex protein molecules, which includes polar area/energy and apolar area/energy. InterProSurf analyzed each chain within the complex and provides interface residues, interface area of each residue and a change in the surface area of each residue upon complex formation. It is freely available at http://curie.utmb.edu/prosurf/html [19].

Results and Discussion:
HPV 16 E7 Protein sequence Analysis:
Analysis of HPV 16 E7 ProPram provides molecular weight (11.02 kD), theoretical pi (4.20), protein iso-electric point as pH 3.97 and total number of negatively charged residues (Asp + Glu) as 19, total number of positively charged residues (Arg + Lys) as five and ext. coefficient was 6335. The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro) >20 hours (yeast, in vivo), total number of atoms found as 1499 and instability index second was computed to be 63.00. This classified the protein HPV type 16 E7 as stable and it also measured aliphatic index as 78.57 and grand average of hydrophaticity (GRAVY) as 0.405.

Characterization of TMEM 50A (Human protein):
TMHMM server [20] and Target-P 1.1 server (http://www.cbs.dtu.dk/services/TMHMM/) and Target-P [20] predicted the presence of four transmembrane domains within the signal peptide, which may be required to direct the protein to secretory pathway. Physiological properties found as molecular weight (17400Da), basal is electronic point (5.57), 17388.70 (monoisotopic mass) and the result of K-NN prediction using PSORT- II gives percentage of plasma-membrane (43.5%), endoplasmic reticulum (26.1%), vacuolar (17.4%), Golgi bodies (4.3%), mitochondrial (4.3%). Post-translation modification using PhosphoSitePlus predicted two serine phosphorylation sites found at amino acid 82 and 84 residue, one possible N-linked glycosylation site located at amino acid 74, and one possible tyrosine phosphorylation site is found. Expasy’s ProPram proteomics server predicted the instability index (II) was computed to be 28.66, aliphatic index score as 95.10 and grand average of hydrophaticity (GRAVY) as 0.580 which classified the protein as stable.

Model refinement and evaluation of Human TMEM 50A
PROCHECK server provided Ramachandran plot analysis of the predicted model. The best model in terms of stereo chemical quality showed Overall G-factor value of -0.38 which indicates that geometry of the model corresponds to high probability confirmation with 85.8% Residues in most favored regions [A, B, L], 9.0% residues in additional allowed regions [a, b, l, p], 2.2% residues in generously allowed regions [~a, ~b, ~l, ~p] and 3.0% of the residues was present in the disallowed region of the plot. Verify 3D analysis revealed that 80% of the residues had an average 3D-1D score of <0.2, predicting that the model is compatible with its sequence. The amino acid environment was evaluated using ERRAT plots, which assess the distribution of different types of atoms with respect to one another in the predicted protein and is used for making decision about its reliability. ERRAT showed an overall quality factor of 31.544, a result expected for crystallographic models with resolution>2.5A. B-factor analysis is done with WHAT IF server reflected the mobility or flexibility of various parts of the molecule. Averaged B-factor deviation for protein backbone was 0.082 (Z score mean) and averaged standard deviation was 2.011. Since average deviation value was less than standard deviation, so it reflected a good quality model Figure 3.

Molecular dynamic simulation of TMEM 50A
The Molecular dynamic stability of TMEM 50A protein has been established by performing the molecular dynamic simulation study using Schrodinger software packages showed the energy of the molecule, which was found constant throughout the time period of simulation. In Figure 4 the radius of gyration increases in between 200 to 400 ps time but after 400 ps it decreases up to 1000 ps and finally omens almost constant for rest of the duration of the simulation. This suggests that the Human TMEM 50A protein model has a compact structure, which provides the required stability to the molecule. In Figure 5 the root mean square deviation (RMSD) during the simulation was increasing in the beginning and at 600 ps it showed highest peak value but after that it becomes almost constant for rest of the duration of the simulation. This suggests that the Human TMEM 50A model has lesser RMSD for Its backbone and it also has less flexibility, indicating that this protein has a stable dynamic behavior structure. In Figure 6 the root mean square (RMS) fluctuations were very less because most of the atoms were free from the RMS fluctuations but some atoms showed this at C and N terminal due to the presence of loop regions of modeled TMEM 50A.

Active site prediction result:
Q-SiteFinder predicted site volume and protein volume in cubic Angstrom with the minimum and maximum co-ordinates [21] and Active site finder (AADS) generated top 15 cavity points for TMEM 50A each pocket was represented by a single cavity point. Then it was sorted into cavities in the descending order of their
volumes. The cavities generated in protein along with the Cartesian coordinates of the cavities are shown in Table 2.

MHC Class I binding epitope prediction:

Table 3 and Table 4 shows the predicted T-cell MHC class-I epitopes for alleles HLA-A*01:01 and HLA-A*02:01 respectively. These small peptides (epitopes) were predicted by the IEDB analysis tool and ranked according to their consensus percentile rank. These percentile ranks refer to the percentage of rank out of one million random peptides from swiss-prot proteins [22]. In this the tap scores of the predicted epitopes were measured as local sequence to structure fitness based on torsion angle propensities normalized against the global minimum and maximum. In table 3 top four epitopes MHGDTPTLHEL, HGDTPTLHEY, GDTPTLHEYM and DTPTLHEYM were selected for docking with receptor protein. While in the table 4 the score rank of ANN and SMM were shown in the units of IC50 (nM), which demonstrated that better binders have a low value of ANN and SMM rank. In this data include measured binding affinities for a total of 15 ranked epitopes for allele HLA-A*02:01. The top ranked epitopes YMLDLQPETT, DLLMGTLGIV, LLMGTLGIVC and TLEDDLLMGTGTL were selected for epitope modeling and docking process. In particular a value of 500 (IC50) is often used as the threshold between binders and non-binders. So, peptides sequence YMLDLQPETT, DLLMGTLGIV, LLMGTLGIVC and LLMGTLGIVC represents highest affinity for binding with receptor proteins.

MHC class II binding epitope prediction for allele HLA DRB1*01:01:

Table 5 and Table 6 shows peptides for MHC Class-II alleles HLA-DRB1*01:01 and HLA-DRB1*01:02 predicted by IEDB analysis tool which were ranked according to their consensus percentile rank. In these predicted epitopes HVDIRTLEDLLMTGTL, IRTLEDLLMLGTVG, VDIRTLEDLLMGTLG and RTLEDLLMGTGTVG were selected for docking with TMEM 50A. In Table 6, for each peptide, a percentile rank for Sturniolo method was generated by comparing the peptides score against the score of five million random isomers selected from Swiss-Prot database. It should be noted that small numbered percentile rank indicates high affinity. So the peptide sequences DLLMTGLGIVCPICSQ, LLMGTLGIVCPICSQK and MGTLGIVCPICSQKP represents high affinity towards receptor proteins because these contained same percentile score and Sturniolo rank of 78.0.

Docking of MHC Class-I epitopes and TMEM 50A

Top ranked predicted epitopes for different alleles of MHC Class I were docked with human TMEM 50A receptor using Patch Dock server. The results have been shown in the Table 7, docked structure of modeled epitope MHGDTPTLHEL of MHC class-I (HLA-A*01:01) and TMEM 50A showed global energy as -16.45 while other modeled epitopes HGDTPTLHEY and DTPTLHEYM of this class showed global energy as -60.15 and -40.25 with TMEM 50A receptor protein while in Table 8 docked structure of modeled epitope YMLDLQPETT of MHC class I (HLA-A*02:01) and TMEM 50A showed global energy as -60.59.

Docking of MHC class-II epitopes and TMEM 50A

Top ranked predicted epitopes for different alleles of MHC class-II were docked with human TMEM 50A receptor using Patch Dock server. The results are shown in the Table 9 and Table 10. The top ranked predicted MHC class-I and MHC class-II modeled epitopes were docked with their respective receptor Human TMEM 50 A. HVDIRTLEDLLMGTGTL docked with human TMEM 50 A showed global energy as -38.45 while another docked molecule IRTLEDLLMGTVG, RTLEDLLMGTGVIC and VDIRTLEDLLMGTGTL contained global energy as -66.57, -9.25 and -9.25. In Table 10 docked molecule DLLMTGLGIVCPICS and human TMEM 50A have global energy as -47.59 while another docked molecules LLMGTLGIVCPICSQ, LMGTLGIVCPICSQK and MGTLGIVCPICSQKP have global energy as -40.36, -54.74 and -39.75 respectively. The modeled epitope DLLMTGLGIVCPICS.pdb was chosen for docking with human TMEM 50 A protein because it has -11.27 as atomic contact energy (ACE) and global energy as -47.59 was given suitable binding affinity to with human protein. In Figure 7 docked molecule showed binding of modeled epitope YMLDLQPETT of MHC Class-I (HLA-A*02:01) with human TMEM 50A receptor protein with global energy -60.59, attractive Vdw -24.24 and repulsive Vdw 26.54. In this, the contribution of the atomic contact energy (ACE) and hydrogen bonds (HB) energy to the global binding energy was found as -9.33 and -1.39. After this InterProSurf server predicts the total accessible surface area of this docked structures as 1568.43 in which polar area/energy was 610.73 and apolar area/energy was 957.70 with 1.4 prob radiiues.

In Figure 8 docked molecule showed binding of modeled epitope MHGDTPTLHEL of MHC class-I (HLA-A*01:01) to with human TMEM 50A with global energy -16.45, attractive Vdw -26.45, repulsive Vdw 10.97, ACE -0.99, HBs -0.52 which gave total accessible surface area of this docked structure as 11286.18 in which polar energy was 2870.52 and apolar energy was 8415.65 with 1.4 prob radiiues. In Figure 9 docked molecules showed binding of modeled epitope DLLMTGLGIVCPICS of MHC- II (HLA-DRB1*01:02) to with human TMEM 50A receptor protein with global energy -47.59, attractive vdw -29.83 and repulsive vdw 7.75, which gave more accurate vaccine candidate structure in comparison to rest of the molecules. In this the contribution of the atomic contact energy (ACE) and hydrogen bonds (HBs) energy to the global binding energy was found as -11.27 and -0.95.
Figure 3: Predicted TMEM 50A model visualized using Discovery Studio 3.5.

Figure 4: Radius of gyration (Rg) for human TMEM 50A protein.

Figure 5: RMSD of the human TMEM 50A protein model for 10 ns.

Figure 6: RMS fluctuation at 10 ns for the TMEM 50A model. X-axis represents the time in picoseconds and Y-axis represents RMSD in nm unit.
Table 1: Top 10 templates selected for homology modeling of TMEM 50A.

| Rank | Pdb hit | Iden 1 | Iden 2 | Cov. | Norm Z-Score |
|------|---------|--------|--------|------|--------------|
| 1    | 3umA    | 0.11   | 0.13   | 0.64 | 1.20         |
| 2    | 2ksA    | 0.20   | 0.10   | 0.41 | 1.17         |
| 3    | 4qbyA   | 0.17   | 0.20   | 0.96 | 0.73         |
| 4    | 3kbPA   | 0.20   | 0.17   | 0.44 | 0.52         |
| 5    | 1iyA    | 0.16   | 0.20   | 0.83 | 0.79         |
| 6    | 4akRA   | 0.09   | 0.17   | 0.94 | 0.73         |
| 7    | 2wCA    | 0.09   | 0.19   | 0.68 | 1.06         |
| 8    | 3kbPA   | 0.21   | 0.17   | 0.45 | 0.96         |
| 9    | 3qCA    | 0.10   | 0.15   | 0.98 | 0.47         |
| 10   | 1lyC    | 0.14   | 0.12   | 0.54 | 0.60         |

Note: In this Iden 1 shows the percentage sequence identity of the templates in the threading aligned region and Iden 2 is the percentage sequence identity of the whole template chains with the TMEM 50A sequence and Norm Z-score is the normalized Z-score of the threading alignments. Alignment with a Normalized Z-score > 1 means a good alignment.

Table 2: Coordinates of cavity points generated by active site finder in TMEM 50A along with an approximate volume in the respective cavity.

| S. No. | Predicted cavity on surface of protein | Cavity Point X-axis | Cavity Point Y-axis | Cavity Point Z-axis | Volume of the Cavity |
|--------|---------------------------------------|---------------------|---------------------|---------------------|----------------------|
| 1      | Cavity_1_INQFMLGKVRSACTPDEYT           | -9.44               | -0.44               | -0.44               | 1.08                |
| 2      | Cavity_2_FVWLAICSMGNEYDHQ              | -12.05              | -0.47               | -0.47               | 1.29                |
| 3      | Cavity_3_FILAMWC2CSTVHEYDQQR           | -12.05              | -0.47               | -0.47               | 1.29                |
| 4      | Cavity_4_OYQVIANFKLMWSCPHD             | -12.05              | -0.47               | -0.47               | 1.29                |
| 5      | Cavity_5_FILMLACWQTVGHHYH              | -12.05              | -0.47               | -0.47               | 1.29                |
| 6      | Cavity_6_NIFMKLCISKACTVWY              | -12.05              | -0.47               | -0.47               | 1.29                |
| 7      | Cavity_7_FILSAICNTIGHVYKDW             | -12.05              | -0.47               | -0.47               | 1.29                |
| 8      | Cavity_8_AGQIVYQFKMLMWSPTD             | -12.05              | -0.47               | -0.47               | 1.29                |
| 9      | Cavity_9_OYQVIANFKLMWSC                | -12.05              | -0.47               | -0.47               | 1.29                |
| 10     | Cavity_10_HFGLKARTVWSES               | -12.05              | -0.47               | -0.47               | 1.29                |

Table 3: Prediction of linear epitopes in HPV 16 E7 with HLA-A*01:01 using NetMHCpan-IEDB

| S. No. | Position Start-End | Pep | Sequence | Proteasome Score | TAP Score | MHC Score | Processing Score | Total Score | MHC IC50 [nM] |
|--------|-------------------|-----|----------|------------------|-----------|-----------|------------------|-------------|--------------|
| 1      | 1:1-10            | 10  | MLMELTLGIV 2.10 | 398.58 | 2.8       | 125.23    | 1.4              | 38150.11    |
| 2      | 1:2-11            | 10  | MLMELTLGIV 2.10 | 398.58 | 2.8       | 125.23    | 1.4              | 38150.11    |
| 3      | 1:3-12            | 10  | MLMELTLGIV 2.10 | 398.58 | 2.8       | 125.23    | 1.4              | 38150.11    |
| 4      | 1:4-13            | 10  | MLMELTLGIV 2.10 | 398.58 | 2.8       | 125.23    | 1.4              | 38150.11    |
| 5      | 1:5-14            | 10  | MLMELTLGIV 2.10 | 398.58 | 2.8       | 125.23    | 1.4              | 38150.11    |
| 6      | 1:6-15            | 10  | MLMELTLGIV 2.10 | 398.58 | 2.8       | 125.23    | 1.4              | 38150.11    |
| 7      | 1:7-16            | 10  | MLMELTLGIV 2.10 | 398.58 | 2.8       | 125.23    | 1.4              | 38150.11    |
| 8      | 1:8-17            | 10  | MLMELTLGIV 2.10 | 398.58 | 2.8       | 125.23    | 1.4              | 38150.11    |
| 9      | 1:9-18            | 10  | MLMELTLGIV 2.10 | 398.58 | 2.8       | 125.23    | 1.4              | 38150.11    |
| 10     | 1:10-19           | 10  | MLMELTLGIV 2.10 | 398.58 | 2.8       | 125.23    | 1.4              | 38150.11    |

Table 4: Prediction of linear epitopes in HPV 16 E7 with HLA-A*02:01 using ANN and SMM methods

| S. No. | Position Start-End | Peptide Length | Sequence | Percentile Rank | ANN IC50 [nM] | ANN Rank | SMM IC50 [nM] | SMM Rank |
|--------|-------------------|---------------|----------|-----------------|---------------|----------|---------------|----------|
| 1      | 1:1-20            | 10            | MLMELTLGIV 2.10 | 398.58 | 2.8       | 125.23    | 1.4              | 38150.11    |
| 2      | 1:2-21            | 10            | MLMELTLGIV 2.10 | 398.58 | 2.8       | 125.23    | 1.4              | 38150.11    |
| 3      | 1:3-22            | 10            | MLMELTLGIV 2.10 | 398.58 | 2.8       | 125.23    | 1.4              | 38150.11    |
| 4      | 1:4-23            | 10            | MLMELTLGIV 2.10 | 398.58 | 2.8       | 125.23    | 1.4              | 38150.11    |
| 5      | 1:5-24            | 10            | MLMELTLGIV 2.10 | 398.58 | 2.8       | 125.23    | 1.4              | 38150.11    |

Table 5: Prediction of linear epitopes in HPV 16 E7 protein with HLA-DRB1*01:01 using iEDB Consensus tool

| S. No. | Position Start-End | Peptide Length | Peptide Sequence | Percentile Rank | Comb. Lib IC50 [nM] | SMM IC50 [nM] | SMM Align Rank |
|--------|-------------------|---------------|-----------------|-----------------|---------------------|---------------|---------------|
| 1      | 73-87             | 10            | HVDIRTLLEDLLMGTL 18.87 | 4983.51 | 75.0       | 14.01         | 127           |

ISSN 0973-2063 (online) 0973-8894 (print)

©2017 Bioinformation 13(5): 122-130 (2017)
Table 6: Prediction of epitopes in e7 with HLA-DRB1*01:02 using IEDB-Sturniolo method

| S. No. | Position Start-End | Peptide Sequence | Percentile Rank | Sturniolo Core | Sturniolo Score | Sturniolo Rank |
|-------|--------------------|------------------|-----------------|----------------|----------------|----------------|
| 1.    | 181-95             | DLMGTLGIVCPICS   | 0.78            | LGIVCPICS      | 2.05           | 0.78           |
| 2.    | 182-96             | LLMGTLGIVCPICS   | 0.78            | LGIVCPICS      | 2.05           | 0.78           |
| 3.    | 183-97             | MLGTLGIVCPICSQK  | 0.78            | LGIVCPICS      | 2.05           | 0.78           |
| 4.    | 184-98             | MGLGTLGIVCPISQKP | 0.78            | LGIVCPICS      | 2.05           | 0.78           |
| 5.    | 170-84             | QSTHVDIRLDEVLM   | 3.28            | IRTLEDLLM      | 0.78           | 3.28           |
| 6.    | 171-85             | STHVDIRLDEVLM    | 3.28            | IRTLEDLLM      | 0.78           | 3.28           |
| 7.    | 172-86             | THVDIRLDEVLM     | 3.28            | IRTLEDLLM      | 0.78           | 3.28           |
| 8.    | 173-87             | HVDIRLDEVLMGL    | 3.28            | IRTLEDLLMGL    | 0.78           | 3.28           |
| 9.    | 174-88             | VDIRLDEVLMGL     | 3.28            | IRTLEDLLMGL    | 0.78           | 3.28           |
| 10.   | 175-89             | DILEQDIRLDEVLM   | 3.28            | IRTLEDLLM     | 0.78           | 3.28           |
| 11.   | 176-90             | IRTLEDLLGLTVG    | 3.28            | IRTLEDLLGTV    | 0.78           | 3.28           |
| 12.   | 177-91             | RTELDLTLGTVG     | 3.65            | LMGLTGIVC      | 0.6           | 3.65           |
| 13.   | 178-92             | TLEDLTLGTVGCP    | 3.65            | LMGLTGIVC      | 0.6           | 3.65           |
| 14.   | 179-93             | LEDLMLGTVGCPI    | 3.65            | LMGLTGIVC      | 0.6           | 3.65           |
| 15.   | 180-94             | EDDLMGLTVGCPI    | 3.65            | LMGLTGIVC      | 0.6           | 3.65           |

Table 7: Docked energy score using FireDock for HLA-A*01:01 specific epitopes with Human TMEM 50 A protein.

| Epitope       | Global energy (KJ/mol) | Attractive vDW | Repulsive vDW | ACE (atomic contact energy) | HB (hydrogen and disulphide bonds) | Transformation (ligand transformation after refinement) |
|---------------|------------------------|----------------|---------------|----------------------------|-----------------------------------|-----------------------------------------------------|
| MHGDTPITLHE   | -16.45                 | -26.45         | 10.97         | -0.99                      | -0.52                             | 0.461751 0.5365/0.479643 0.148394 0.325716 0.359675 0.58814320 |
| HCDTPITLHEY   | -60.15                 | -37.44         | 50.00         | -17.29                     | -1.06                             | 1.650077 0.501054 1.488182 0.437999 0.33281659 0.1930298 |
| DTPITLHEYML   | -40.25                 | -35.06         | 11.92         | -5.85                      | -5.27                             | 0.125927 -0.101452 -0.280155 -0.851133 0.27405537 0.21195085 |

Table 8: Docked energy score using FireDock for HLA-A*02:01 specific epitopes with Human TMEM 50 A protein.

| Epitope       | Global energy (KJ/mol) | Attractive vDW | Repulsive vDW | ACE (atomic contact energy) | HB (hydrogen and disulphide bonds) | Transformation (ligand transformation after refinement) |
|---------------|------------------------|----------------|---------------|----------------------------|-----------------------------------|-----------------------------------------------------|
| YMLDLPQETT    | -60.59                 | -24.24         | 6.54          | -15.33                     | -1.68                             | -1.799130 0.078032 -0.747186 -0.563374 -0.1689145 0.2483170 |
| DLMLGTGLV     | -55.57                 | -28.15         | 11.13         | -14.56                     | -3.41                             | 1.133225 -0.037606 -0.816487 0.1096378 -0.143226 0.577003 |
| LLMGTLGIVC    | -55.24                 | -24.39         | 3.40          | -11.40                     | -2.64                             | 0.720709 -0.173591 -0.310262 0.17513165 -0.295737 0.188661 |
| RTLEDLLGLT    | -55.16                 | -22.14         | 2.12          | -9.65                      | -1.20                             | -0.184480 0.320455 0.783296 0.989859 0.16813620 -0.5468778 |

Table 9: Docking using FireDock for HLA-DRB1*01:01 specific epitopes with human MEM 50A protein.

| Epitope       | Global energy (KJ/mol) | Attractive vDW | Repulsive vDW | ACE (atomic contact energy) | HB (hydrogen and disulphide bonds) | Transformation (ligand transformation after refinement) |
|---------------|------------------------|----------------|---------------|----------------------------|-----------------------------------|-----------------------------------------------------|
| HVDIRLDEVLMGT | -38.45                 | -26.15         | 27.01         | -9.33                      | -1.39                             | 0.699969 -0.397265 -1.029715 -1.292001 0.1410763 2.199753 |
| IRTLEDLLGLTVG | -66.57                 | -36.11         | 52.99         | -33.14                     | -1.22                             | 1.075338 -0.722708 -2.897819 -0.189286 -2.212364 -0.7311763 |
| RTILEDLLGLTVC | -49.25                 | -19.25         | 7.71          | 4.96                       | -2.61                             | -2.138732 0.461615 2.041796 0.1667150 0.2018180 0.587845 |
| VDIRLDEVLMGL  | -49.25                 | -19.25         | 7.71          | 4.96                       | -2.61                             | -2.138732 0.461615 2.041796 0.1667150 0.2018180 0.587845 |

Table 10: Comparison of parameters HLA allele specific peptides docked to TMEM 50A

| S. No | Epitopes and alleles | HLA alleles | Global energy | Attractive vDW | Repulsive vDW | Total accessible surface area | Polar energy/Apolar energy | No. of residue at interface | Prob Radius |
|-------|----------------------|-------------|---------------|----------------|---------------|-------------------------------|--------------------------|-------------------------------|------------|
| 1.    | YMLDLPQETT           | A*02:01     | -60.59        | -24.24         | 26.54         | 11286.18                      | 2870.52/4815.65          | LEU                          | 1.4        |
| 2.    | MHGDTTPITLHE         | A*01:01     | -16.45        | -26.45         | 10.97         | 11286.18                      | 2870.52/4815.65          | LEU                          | 1.4        |
| 3.    | DMLGTLGIVCPICS      | DRB1*01:02  | -47.59        | -29.83         | 7.75          | 11286.18                      | 2870.52/4815.65          | CY5                          | 1.4        |
| 4.    | HVDIRLDEVLMGT       | DRB1*01:01  | -38.45        | -26.15         | 27.01         | 11286.18                      | 2740.68/4828.70          | Glu, ARG and CY5             | 1.4        |

ISSN 0973-2063 (online) 0973-8894 (print) 128 ©2017 Bioinformation 13(5): 122-130 (2017)
Figure 7: Molecular docking of YMLDLQPETT (E7 antigen) peptide with predicted HLA-A*02:01 specificity to human TMEM50A.

Figure 8: Molecular docking of MHGDTPTLHE (E7 antigen) peptide with predicted HLA-A*01:01 specificity to human TMEM50A.

Figure 9: Molecular docking of HVDIRTLDDLGMGL (E7 antigen) peptide with predicted HLA-DRB1*01:01 specificity to human TMEM50A.

Figure 10: Molecular docking of DLLMGTLGIVCPICS (E7 antigen) with predicted HLA-DRB1*01:02 specificity to human TMEM50A.
Beside this, InterProSurf server predicted the total accessible surface area of the docked structure as 11286.18 in which polar area/energy were 2870.52 and apolar area/energy was 8415.65 with 1.4 prob radiiuses. It also predicted the total number of surface atoms as 859 and total buried atoms were 371. In Figure 10, docked molecule showed binding of modeled epitope HVDIRTLEDLLMGTL of MHC Class-II (HLA-DRB1*01:01) to human TMEM 50A receptor protein. In which modeled epitope protein showed in blue color and human TMEM 50A shown in green color visualized with Discovery Studio 3.5 client tool with global energy -38.45, attractive Vdw -26.15 and repulsive Vdw 27.01 which gave more accurate vaccine candidate structure in comparison to rest of the molecules. In this the contribution of the atomic contact energy (ACE) and hydrogen bonds (HBs) energy to the global binding energy was found as -9.33 and -1.39. After this InterProSurf server predicted the total accessible surface area of this docked structure as 9569.38 in which polar area/energy was 2740.68 and apolar area/energy was 8415.65 surface atoms as 859 and total buried atoms were 371.

Conclusion:
We document the predicted binding of linear peptide epitopes YMLDLQPETT (HLA-A*02:01), MHGDTPTLHE.HLA-A*01:01, DLLMGTLLGCPISCHLA-DRB1*01:02 and HVDIRTLEDLLMGTL (HLA-DRB1*01:01) in HPV type 16 E7 antigen with the human TMEM 50A protein structural model. The data documented in this study finds application in development of components towards HPV associated disease prevention.

Acknowledgement:
The authors are grateful to Sam Higginbottom Institute of Agriculture, Technology and Sciences, Allahabad for providing facilities and support to complete the present research work.

References:
[1] Pfister H et al. Rev Physiol Biochem Pharmacol. 1984 99: 111-81 [PMID: 6320328].
[2] Pfister H et al. Adv Cancer Res. 1987; 48: 113-47 [PMID: 3039808].
[3] Walboomers JM & Jacobs MV, J Pathol. 1999 189: 12-9 [PMID: 10451482].
[4] Konya J & Dillner J, Adv Cancer Res. 2001 82: 205-38 [PMID: 11447764].
[5] Müller-Schiffmann A & Beckmann J, J Virol. 2006 80: 8718-28 [PMID: 16912319].
[6] Bosch FX & Lorincz A, J Clin Pathol. 2002 55: 244-65 [PMID: 11919288].
[7] Burd EM et al, Clin Microbiol Rev. 2003 16: 1-17 [PMID: 12525422].
[8] Petrovsky N & Silva D, Novartis Found Symp. 2003 254: 23-25 [PMID: 14712930].
[9] Altschul SF et al. J Mol Biol. 1990 215: 403-10 [PMID: 2231712].
[10] Vriend, G et al., J Mol Graph. 1990 8: 52-56 [PMID: 2268628].
[11] Laskowski RA & MacArthur MW, J. Appl. Cryst. 1993 26: 283-291.
[12] Sippl MJ et al, Proteins. 1993 17: 355-62 [PMID: 8108378].
[13] Eisenberg D & Lüthy R, Methods Enzymol. 1997 277: 396-404 [PMID: 9379925].
[14] Ramachandran G N & Ramakrishnan C, J Mol Biol. 1963 7: 95-9 [PMID: 13990617].
[15] Singh T & Biswas D, J. Chem. Inf. Model. 2011 51: 2515–2527 [DOI: 10.1021/ci200193z].
[16] Saha S & Raghava GP, Proteins. 2006 65: 40-8 [PMID: 16894596].
[17] Schneidman-Duhovny D & Inbar Y, Nucleic Acids Res. 2005 33: 363-7 [PMID: 15980490].
[18] Andrusier N & Nussinov R, Proteins. 2007 69: 139-59 [PMID: 17598144].
[19] Negi SS & Schein CH, Bioinformatics. 2007 23: 3397-9 [PMID: 17933856].
[20] Emanuelsson O & Nielsen H, J Mol Biol. 2000 300: 1005-16 [PMID: 10891285].
[21] Laurie AT & Jackson RM, Bioinformatics. 2005 21: 1908-16 [PMID: 15701681].
[22] Nielsen M & Lundegaard C, BMC Bioinformatics. 2007 8: 238 [PMID: 17608956].
[23] Zhang C & Cornette JL, Protein Sci. 1997 6: 1057-64 [PMID: 9144777].

Edited by P Kangueane
Citation: Srivastava et al, Bioinformation 13(5): 122-130 (2017)