Journal of Computational Intelligence in Bioinformatics
ISSN 0973-385X Volume 6, Number 1 (2013) pp. 27-33

https://dx.doi.org/10.37622/JCIB/6.1.2013.27-33

Comparative Homology Modelling for HPV Type 16 E 7 Proteins by using MODELLER and its Validations with SAVS and ProSA Web Server

Upasna Srivastava*, Gurmit Singh

Department of Computational Biology & Bioinformatics,*
Department of Computer Science & Information Technology,
SHIATS, Allahabad - 211007, India.
**Corresponding Author: Upasna Srivastava*
Department of Computational Biology & Bioinformatics,
Sam Higginbottom Institute of Agriculture, Technology & Sciences
Deemed to be University, Allahabad - 211007, India.

Abstract

Human papillomaviruses (HPV) are a group of viruses which are associated with various proliferative diseases in the infected epithelium. HPV types 6, 11, 16, and 18 are the most common, are associated with lesions in the anogenital tract. The "benign" types 6 and 11 are mainly associated with condylomata acuminate of the oncogenic HPV types, HPV-16 is most frequently found in cervical carcinomas. The 3-dimentional structure of the protein was not yet available in Protein Data Bank; hence the present paper of predicting the 3D model of the HPV Type 16 E7 proteins. Template based homology modelling predicts the three-dimensional structure of hpv type 16 E 7 protein sequence (target) based primarily on its alignment to one or more proteins of known structure as template generated by MODELLER. The prediction process consists of target-template alignment, model building, and model evaluation. The model was checked for stereo chemical quality by PROCHECK, VERIFY 3D, WHAT IF, ERRAT AND ProSA servers were also used for the display of Z-scores and energy plots. Finally the protein was visualized with Swiss-PDB viewer.

Keywords: Homology modelling, SAVS, ProSA servers, Human papilloma virus.

Introduction

Three-dimensional (3D) protein structures are of great interest for the rational design of many different types of biological experiments, such as site-directed mutagenesis or structure-based discovery of specific inhibitors. The 3D structure of proteins is better conserved than their sequences, it is often possible to identify a homologous protein with a known structure (template) for a given protein sequence (target). In these cases, homology modelling has proven to be the method of choice to generate a reliable 3D model of a protein from its amino acid sequence as impressively shown in several meetings of the bi-annual CASP experiment [1]. Human papillomaviruses (HPV) are a group of viruses which are associated with various proliferative diseases in the infected epithelium [2].To date, more than 70 HPV types have been reported. Some of these, among which HPV types 6, 11, 16, and 18 are the most common, are associated with lesions in the anogenital tract [3]. The "benign" types 6 and 11 are mainly associated with condylomata acuminate of the oncogenic HPV types, HPV-16 is most frequently found in cervical carcinomas and the viral genome is often found integrated into the cellular genome [4]. The nuclear protein E7 of HPV-16 is considered to be one of the two major proteins involved in malignant transformation and the maintenance of the transformed phenotype of cells. The viral oncoprotein E7 is consistently transcribed in both HPV positive cervical cancer cell lines and cervical cancers [5, 6].In this MODELLER is used for homology or comparative modelling of protein three-dimensional structures (7,8). It implements comparative protein structure modelling by satisfaction of spatial restraints (9,10), and can perform many additional tasks, including de novo modelling of loops in protein structures, optimization of various models of protein structure with respect to a flexibly defined objective function, multiple alignment of protein sequences and/or structures, clustering, searching of sequence databases, comparison of protein structures. After generation of a suitable model, it is validated by using Structural Analysis and Verification Server (SAVS). SAV Server include various software's as PROCHECK (11) , WHAT IF (12,13,14), ERRAT(15,16), VERIFY-3D (17) and PROVE(18) and beside this , ProSA was also used for the display of Z-scores and energy plots. Finally the protein was visualized with Swiss-PDB viewer.

Materials and Methods

Template search and sequence alignment for the amino acid sequence of HPV Type 16 E7 protein was performed after the sequence retrieval from the sequence database of Uniprot (http://www.expasy.org).The 3-dimentional structure of the protein was not yet available in Protein Data Bank; hence the present paper of predicting the 3D model of the HPV Type 16 E7 protein was undertaken. BLASTP search was performed against Brookhaven Protein Data Bank (PDB) with the default parameters to find suitable templates for homology modelling. Based on the maximum identity with high score and lower e-value crystal structure having PDB id 2EWLA was selected as the template. The sequence identity and similarity between the target and template were 47% and 64%, respectively. The sequence alignment of HPV Type 16

E7 protein and the template 2EWLA was carried out using the CLUSTALW (http://www.ebi.ac.uk/clustalw) program[20] .3D-Structure generation the academic version of MODELLER9V6(http://www.salilab.org/modeler) , was used for 3D structure generation based on the information obtained from sequence alignment [7] The MODELLER software employs probability density functions (PDFs) as the spatial restraints rather than energy. The 3D model of a protein is obtained by optimization of the molecular pdfs such that the model violates the input restraints as possible. The molecular pdfs was carried as a combination of pdfs restraining individual spatial features of the whole molecule [Table 1].Out of 10 models generated by MODELLER the one with the best G-score of PROCHECK [11], and with the best VERIFY3D [17] profile as subjected to energy minimization. VERIFY3D (structure evaluation server) was used to check the residue profile of the obtained three-dimensional models. In order to assess the stereo-chemical qualities of the three dimensional models, PROCHEK analysis was performed. Quality evaluation of the model for the environment profile was also predicted using ERRAT (structure evaluation server). The final refined model was evaluated for its atomic contacts using What-if program to identify bad packing of side chain atoms or unusual residue contacts. The predicted model was also analyzed for the recognition of errors in the three dimensional structure by using the ProSA web server [19]. This model was further subjected for the identification of active site and docking study (Figure 1: A).

Table 1: Top 10 models generated by Modeller for HPV Type 16 E7 proteins. In this Ident1 is the percentage sequence identity of the templates in the threading aligned region with the query sequence and Ident2 is the percentage sequence identity of the whole template chains with query sequence.

Rank	PDB Hit	Iden 1	Iden 2	Cov.	Norm Z-score
1.	2ewlA	0.41	0.24	O.57	2.25
2.	2ewlA	0.38	0.24	0.56	4.80
3.	2f8bA	0.38	0.24	0.57	2.47
4.	2b9dA	0.39	0.20	0.52	1.15
5.	2ewlA	0.38	0.24	0.57	4.71
6.	2ewlA	0.38	0.24	0.57	4.62
7.	2f8bA	0.41	0.24	0.57	2.23
8.	2ewlA	0.38	0.24	0.53	3.03
9.	1v87A	0.18	0.24	0.87	0.81
10.	2b9dA	0.38	0.20	0.51	4.67

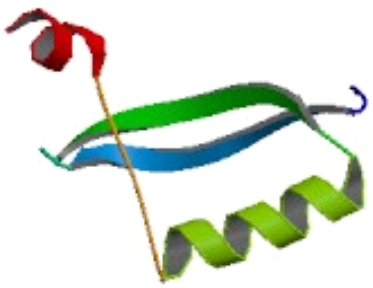

Figure 1 (A): Predicted model of HPV type 16 E 7 protein.

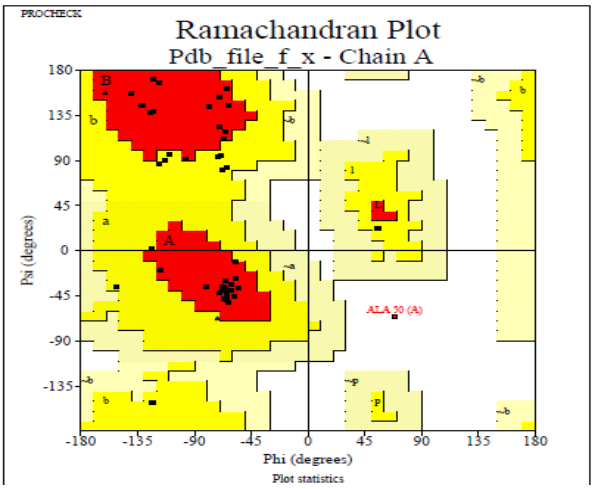

Figure 1(B): Ramachandran plot analysis for model Hpv type 16 E7 proteins. The plot statistics are: Total number of residues-49 with 72.7% Residues in most favoured regions [A,B,L](red); 25.0% Residues in additional allowed regions [a,b,l,p] (dark yellow); no residue in generously allowed regions [~a,~b,~l,~p] (light yellow); 2.3% Residues in disallowed regions. Number of glycine residues (shown as triangles) is 2 and Number of proline residues is 1.

Results and Dissection

Model building, refinement and evaluation for the modelled structure of the Hpv type 16 E7 proteins The predicted model was subjected to model refinement followed by validation using SAVS and PROSA server. Ramachandran plot was drawn and the structure was analyzed by PROCHEK, then it was found that the 72.7% of the residues fell in the most favoured region, 25.0% of the residue fell in additional allowed region and 2.3% of the residue fell in disallowed region and there is no any residue fell in generously allowed regions. All these finding suggest a stereo chemical good, model. The overall PROCHECK G factor for the homology-modelled structure was (Dihedrals: -0.26 Covalent: 0.40) was zero. The final structure was validated by an ERRAT graph. The quality factor of 97.561 indicated good quality, as score of > 50 are acceptable for a reasonable model (Figure 3).High quality of model is also

confirmed from VERFY 3D server of modelled protein showed a score higher than 0.2 thus the model showed satisfactory 3D-1D scores for all the residues in the sequence. 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 conformation was (-0.493), averaged standard deviation of Omega values was 5.877.

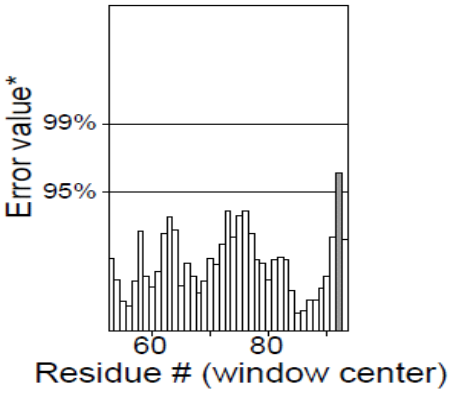

Figure 3: Overall quality factor of predicted model by ERRAT2 program. In this, on the error axis, two lines are drawn to indicate the confidence with which it is possible to reject regions that exceed that error value.

ZScore Analysis

RMS Z-scores for anomalous bond lengths and bond angles as determined by WHAT IF 0.628 and 0.886 respectively, which is very close to 1.0 suggesting high model quality.

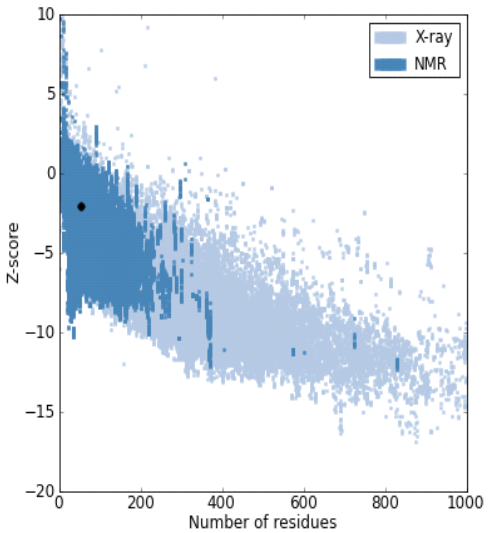

Figure 4: ProSA-web z-scores of all protein chains in PDB determined by X-ray crystallography (light blue) or NMR spectroscopy (dark blue) with respect to their

length. The z-scores of calculated model of Hpv type 16 E7 proteins are highlighted as large dots.

ProSA was used to check three dimensional models of Hpv type 16 E7 proteins for potential errors. The program displays two characteristics of the input structure: its Z-score and a plot of its residue energies. The Z-score indicates overall model quality and measures the deviation of the total energy of the structure with respect to an energy distribution derived from random conformations. The energetic architecture as predicted by PROSA score was negative (-2.06) for the modelled protein. The value is quite nearer to that of template (-2.11), which indicates the reliability of the model (Figure 4 and Figure 5).

Summary and Conclusion

The present study analyses the three dimensional structure of Human papilloma virus type 16 E 7 proteins. It was predicted using HPV45 oncoprotein E7 as template (PDB Id: 2EWLA). Structure validation by PROCHECK, Verify 3D, WHAT IF, ERRAT and ProSA confirmed the reliability of the model. The predicted properties of hpv type 16 E7 protein helps in drug designing and vaccine designing against cervical cancer.

Acknowledgement

The authors are deeply grateful to Sam Higginbottom Institute of Agriculture, Technology and Sciences for providing the necessary guidance, facilities and support required for completing the research work.

References

[1] Tremonton, A. and Morea, V. (2003) Assessment of homology-based predictions in CASP5. Proteins, 53 (Suppl. 6), 352–368.

[2] Pfister H. Biology and biochemistry of papillomaviruses. Rev Physiol Biochem Pharmacol 1984; 99: 112-81.

[3] Pfister H. Human papillomaviruses and genital cancer. Adv Cancer Res 1987; 48:113-47

[4] Cullen AP, Reid R, Campion M, Lorincz AT. Analysis of The physical state of different human papillomavirus DNAs in intra-epithelial an invasive cervical neoplasm. J Viro1991; 65:606-12.

[5] Pfister H. Human papillomaviruses and genital cancer. Adv CancerRes 1987; 48:113-47. Cullen AP, Reid R, Campion M, Lorincz AT. Analysis of

[6] The physical state of different human papillomaviruses DNAs in intra-epithelial an invasive cervical neoplasm. J Viro1991; 65:606-12.

[7] Yee C, Krishnan-Hewlett I, Baker CC, Schlegel R, Howley PM. Presence and expression of human papillomaviruses sequences in human cervical carcinoma cell lines. Am Pathol 1985; 119:361-6.

[8] Van den Brule AJ, Cromme FV, Snijders PJ, Smit L, Oudejans CB, Baak JP, *et al.* Nonradioactive RNA in situ hybridization detection of human papillomaviruses 16-E7.

[9] N. Eswar, M. A. Marti-Renom, B. Webb, M. S. Madhusudhan, D. Eramian, M. Shen, U. Pieper, A. Sali. Comparative Protein Structure Modelling With MODELLER. Current Protocols in Bioinformatics, John Wiley & Sons, Inc., Supplement 15, 5.6.1-5.6.30, 2006.

[10] M.A. Marti-Renom, A. Stuart, A. Fiser, R. Sanchez, F. Melo, A. Sali. Comparative protein structure modelling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. 29, 291-325, 2000.

[11] A. Sali & T.L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779-815, 1993.

[12] A. Fiser, R.K. Do & A. Sali. Modelling of loops in protein structures, Protein Science 9. 1753-1773, 2000.

[13] Laskowski R A, MacArthur M W, Moss D S and Thornton J M (1993). PROCHECK: a program to check the stereo chemical quality of protein structures. J. Appl. Cryst., 26, 283-291.

[14] WHAT IF: a molecular modelling and drug design program, G.Vriend, J. Mol. Graph. 8, 52—56 (1990).

[15] WHAT_CHECK (verification routines from WHAT IF) R.W.W.Hooft, G.Vriend, C.Sander and E.E. Abola, Errors in protein structures .Nature 381, 272 (1996).

[16] Ramachandran plot: G.N.Ramachandran, C.Ramakrishnan and V.Sasisekharan, Stereochemistry of Polypeptide Chain Conformations. J. Mol. Biol. 7, 95--99 (1963).

[17] Sippl, M.J. (1993), Recognition of Errors in Three-Dimensional Structures of Proteins. Proteins 17, 355-362.

[18] Colovos C & Yeates TO, Protein Sci. 1993 2: 1511 [PMID: 8401235]

[19] VERIFY3D: assessment of protein models with three-dimensional profiles: Eisenberg D, Lüthy R, Bowie JU.Source Laboratory of Structural Biology and Molecular Medicine, University of California, Los Angeles 90095, USA. [PMID: 9379925].

[20] Pontius J, Richelle J, Wodak SJ. (1996). Deviations from standard atomic volumes as a quality measure for protein crystal structures. J. Mol. Biol. 264, 121-136.

[21] Wildenstein & Sippl (2007), ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Research 35, W407-W410.

[22] Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994 Nov 11; 22(22):4673-80.