SEEJPH Volume XXVI, S2, 2025, ISSN: 2197-5248; Posted:03-02-2025

NGS Analysis Human Papillomavirus Type 18 E2 DNA-Binding Domain Bound to its DNA Target with Biopython

Tanmay Bandbe [1], Juri Saikia [2], Uma Kumari [3*]

[1]*Project Trainee at Bioinformatics Project and Research Institute, Noida-201301, India*
[2]*Project Trainee at Bioinformatics Project and Research Institute, Noida-201301, India*
[3]*Senior Bioinformatics Scientist, Bioinformatics Project and Research Institute, Noida-201301, India. Email: uma27910@gmail.com*
**Corresponding Author: Uma Kumari*

KEYWORDS
HPV18, E2 Protein, DNA-binding domain, NGS analysis, Biopython, Protein-DNA interaction, Sequence alignment

ABSTRACT
Cervical cancer, a significant global health concern, is primarily caused by persistent infection with high-risk human papillomavirus (HPV) types, particularly HPV18. The E2 DNA-binding domain (DBD) of HPV-18 plays a crucial role in regulating viral replication and transcription. Disrupting the function of this domain can significantly impact the viral lifecycle, thereby preventing the progression of cervical cancer. By studying the molecular interactions between the E2 DNA-binding domain and its DNA target, the research aims to identify potential therapeutic targets and provide a framework for developing targeted treatments. This study leverages high-throughput sequencing technologies to decode proteomics data, illuminating critical genetic information and gene expression patterns. Through the multifaceted application of advanced bioinformatics tools, including sequence analysis using BLAST; multiple sequence alignment with CLUSTAL Omega and COBALT; protein domain and motif identification via InterProScan; structural analysis using CATH and PDBSum; and visualization with Jalview, we dissected the molecular mechanisms underpinning cancer pathogenesis. Furthermore, the study determined proteolytic cleavage site prediction by Expasy Peptide Cutter; and identified potential immunogenic regions within the E2 DBD using an in-silico tool, IEDB B and T-cell epitope. This integrated approach provides insights into the structural and functional properties of the HPV18 E2 DBD and its interaction with DNA, as well as potential implications for immunotherapeutic strategies targeting this critical viral protein. Additionally, Biopython was employed for bioinformatics analyses, enabling seamless data integration and processing to enhance the prediction accuracy of immunogenic regions and further characterize the E2 DBD. Through the application of NGS and bioinformatics tools, the study provides a thorough investigation of the molecular interactions and mechanisms by which the HPV-18 E2 DBD influences the pathogenesis of cervical cancer. By identifying key interactions and potential therapeutic targets, this research lays the groundwork for developing targeted treatments that can effectively disrupt the viral lifecycle and reduce the risk of cervical cancer.

1. Introduction

The most prevalent health issue affecting women globally Cervical cancer is considered the fourth cause of death among women worldwide with its establishment being linked with human papillomavirus (HPV) infection [1]. Cervical cancer is due to many environmental and genetic factors that lead to changes in epithelial cells and malignancy. The initial formation and growth of these tumors depend on two main oncogenes E6 and E7, whose expression leads to cancer.

It is believed that HPV E2 proteins regulate the transcription of viral genes by attaching to specific locations on the viral DNA, which are found in multiple copies within the viral upstream regulatory regions (URRs) [2]. The HPV E2 protein plays a crucial dual role in cancer development. It normally controls viral DNA replication and gene expression (including the E6/E7 oncogenes) by binding to specific DNA sites, thus preventing cancer. However, E2 can also prevent cancer by suppressing the expression of these same oncogenes. Therefore, loss or reduction of E2 activity is a key step in HPV-driven cancer. This loss often occurs due to disruption of the E2 gene during viral integration into the host's chromosomes, leading to uncontrolled production of the cancer-promoting E6 and E7 proteins, ultimately driving carcinogenesis [3]. Malignant tumor that develops in or on the lower part of the uterus and is divided into two categories depending on the type of cervical cells affected cervical squamous cell which affects the squamous cells of the cervix, and cervical adenocarcinoma which targets the mucus-producing cells of the cervix [4].

The main preventable risk factors for cervical cancer include human papillomavirus (HPV) infection, Chlamydia infection, multiple sex partners, smoking, AIDS, a weakened immune system, Exposure to miscarriage prevention medicine, and long-term use of oral contraceptives. In recent years, the integration of bioinformatics, molecular modeling, and machine learning into the drug development technique has extensively expanded the discovery of new drug candidates while reducing the costs associated [5].

This approach uses molecular docking, computational modeling, and simulation techniques to develop drug candidates [6]. Next Generation Sequencing (NGS) has revolutionized virology and oncology through comprehensive, high-throughput analysis of disease genomes and their mutations. NGS enables genome-wide analysis to identify viral and homologous regions in the host DNA. Integration of NGS data with structural analysis of the E2 protein may provide a broader understanding of HPV pathogenesis and help improve treatment plans. In HPV research, NGS is used for a variety of purposes, including Genotyping and identifying specific HPV types present in clinical samples. Mutations that may alter DNA binding capacity or regulation of oncogenes, provide information on how mutations in the DNA targets affect oncogene regulation. NGS can also identify resistance mechanisms that may arise from mutations in the E2-DNA interaction site. Targeting the DNA binding domain of the E2 protein through drug design and combining it with insight from NGS may provide a new pathway for antiviral therapy. NGS has emerged as an effective tool in unraveling the complexities of cervical cancer and HPV interactions. Its potential to provide precise genomic and epigenomic insights contributes to higher diagnostic and prognostic techniques. NGS helps to improve cervical cancer prevention and treatment and decrease the global burden of HPV-related cancers.

2. Materials and Methods

2.1. Sequence Retrieval and Alignment Analysis:

We retrieved the nucleotide sequences for the HPV-18 E2 DBD and its target DNA, protein 1JJ4 from the Protein Data Bank (PDB) using the National Center for Biotechnology Information's (NCBI) Molecular Modeling Database (MMDB) Next, we utilized the NCBI's BLAST (Basic Local Alignment Search Tool) to perform a sequence alignment of this protein (input as FASTA format) and to understand the evolutionary relationships of the 1JJ4 protein with other proteins that can designate the E2's role in pathogenesis. Also, we identified the bit score for each BLAST hit to assess the similarity between the query sequence and the database sequences which is a log2 scaled and normalized raw score that indicates the quality of the alignment [7, 8].

2.2. Multiple Sequence Alignment Analysis:

We used the CLUSTAL Omega web server to perform a multiple sequence alignment (MSA) of the proteins 1JJ4 and 1F9F (provided in FASTA format as input) with other HPV oncogenic protein sequences. The server ran with default settings for both seeded guide trees and profile-profile techniques. By analyzing the resulting MSA, we aimed to identify conserved amino acid residues within the E2 DBD domain. In parallel, we uploaded the same FASTA sequence of proteins 1JJ4 and 1F9F as input and performed a multiple-sequence alignment using NCBI's COBALT (Constraint-based Multiple Alignment Tool) with customized parameters. We aimed to identify conserved and variable regions between the E2 DBD and its target in this multiple-sequence alignment, with a focus on identifying potential binding sites represented by peaks in the sequencing data [9, 10,17,18,19]

2.3. Domain and Motif Analysis (Signature Protein):

To understand how proteins 1JJ4 1F9F and their targets interact with DNA and how this contributes to the E2 DBD function, we used the InterProScan web tool. The analysis identified potential functional domains, families, and specific sites within the E2 protein, particularly focusing on the predicted E2 DBD [11].

2.4. Structural Classification and Analysis Using CATH:

We retrieved the nucleotide sequence for protein 1JJ4 (as 1F9F is homologous) and prepared the sequences in FASTA format as an input into the CATH (Class, Architecture, Topology, Homologous superfamily) web service. CATH uses the sequence to classify the protein's structure within a hierarchy, revealing its evolutionary connections to other proteins. By analyzing these results, we gained insights into the E2 DBD's structural features and functional domains. Furthermore, we compared this classification to known functional domains to pinpoint potential regions on the protein that could interact with other molecules and potentially be targeted for inhibition [12].

2.5. 3D Structure Analysis Using Protein DatabaseSum:

To comprehend the structure of the 1JJ4 (as 1F9F is homologous) protein-DNA complex in detail, we retrieved its crystallographic structure from the Protein Data Bank (PDB) in PDB format. We then used this structure as input into the PDBSum web service to analyze and visualize this complex structure. This analysis provided us with a comprehensive picture of the complex, including secondary structure elements (alpha-helices, beta-sheets, etc), 3D structure, specific interactions between the protein and DNA, including hydrogen bonds, salt bridges, and the involvement of any ligands or metal ions, functional domains, and structural features [13,21].

2.6. Sequence Alignment Visualizations Analysis:

We utilized protein 1JJ4 (as 1F9F is homologous) in FASTA format as input and employed Jalview version 2 to analyze and annotate the alignment of protein 1JJ4 sequences. The conservation Plot feature helped us generate a graphical representation of residue conservation across the alignment, highlighting important regions. Further, we enriched the alignment by adding annotations for functional domains, motifs, and secondary structure elements. Also, Jalview allowed us to import sequence features from external databases like UniProt and overlay them on the alignment, providing additional context. Likewise, leveraging Jalview's integrated tools, we predicted the presence of alpha-helices, beta-strands, and other secondary structures within the protein sequences [14].

2.7. Determination of Cleavage Sites Prediction in Protein Sequence:

We started by retrieving the amino acid sequence for protein 1JJ4 (UniProt accession P06790) in the UniProt database. This sequence likely plays a role in how the HPV18 E2 protein interacts with DNA. Next, we used a tool called ExPASy Cutter to predict where the protein might be cut by specific enzymes (proteases) or chemicals. We uploaded the amino acid sequence as input into the tool and chose the conditions we wanted to test. The tool then analyzed the sequence and provided a map that showed the sequence in easy-to-read blocks, with arrows or markers highlighting potential cleavage sites. Additionally, a table provides detailed information about each potential cleavage site [15].

2.8. Identification of Epitopes Analysis:

We used T cell epitopes (focusing on peptide binding to MHC class I molecules) and another for B cell epitopes (Bepipred Linear Epitope Prediction 2.0) to predict potential epitopes. Specifically, we uploaded the sequence for the E2 DBD region of protein 1JJ4 to each tool and submitted it for analysis. The resulting predicted T and B-cell epitopes were then examined in three ways: their location within the E2 DBD, their potential to trigger an immune response (immunogenicity), and the possibility of cross-reactivity with other HPV types or even human proteins [16].

2.9. Computational Analysis Using Biopython:

With Biopython (v1.26.4), we obtained the HPV18 E2 DBD nucleotide sequence (1JJ4) and its corresponding DNA target from NCBI GenBank, parsing the data for sequence ID, description, and sequence. We also retrieved the 1JJ4 3D structure from the RCSB PDB using Biopython, generating a Ramachandran plot to assess its stereochemical quality. Later, we extracted the crystallographic and

structural data from the CIF and PDBXML files, respectively. Finally, we used Biopython libraries to analyze the 1JJ4 structure, identifying key features and interactions within the E2 DBD.

3. Results

3.1. Sequence Retrieval and Alignment Analysis:

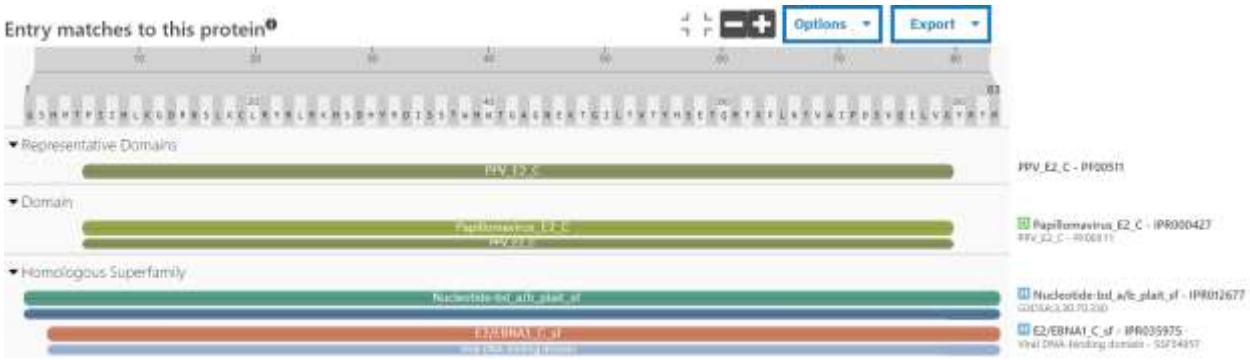
The BLAST analysis of protein 1JJ4 revealed a strong match with protein 1F9F. The high Max Score and Total Score indicate a robust alignment, while the 100% Query Cover signifies that the entire query sequence is involved in the match. The extremely low E-value (2e-36) provides strong statistical evidence that the match is not due to chance, suggesting a high degree of similarity between the two proteins. With a bit-score of 174, these two protein sequences are highly similar, indicating the possibility of homologous links. Overall, the robustness of these alignments is highlighted by the high Bit Score and low E-value, which imply that the discovered homologs have important structural and functional similarities with 1JJ4.

3.2. Multiple Sequence Alignment Analysis:

CLUSTAL Omega Analysis:

The constraint-based multiple alignments (COBALT) of the protein structures 1JJ4 and 1F9F have provided significant insights into their evolutionary relationships and functional domains. The alignment revealed several conserved regions across both proteins, indicating their shared evolutionary origins and potential functional similarities. These conserved motifs are critical for maintaining the structural integrity and biological functions of the proteins. For 1JJ4, the alignment highlighted regions that are essential for its role in specific biological processes, suggesting potential sites for targeted mutagenesis studies. Similarly, for 1F9F, the conserved regions identified through the alignment suggest its involvement in related biological functions, providing a basis for further experimental investigations. Overall, the COBALT has enhanced our understanding of the structural and functional characteristics of protein and paves the way for future research, including experimental validations and potential new insights into therapeutic interventions.

3.3. Domain and Motif Analysis (Signature Protein):



Papillomavirus type 18 E2 DNA-binding Domain

Figure 2: Molecular function, Ancestor chart of Human Papillomavirus type 18 E2 DNA- binding domain bound to its DNA Target

Through InterProScan analysis, we gain a detailed understanding of significant similarities in the domain structures and functional annotations of IJJ4 and IF9F proteins, suggesting they may share functional homology and play similar roles in DNA repair and maintenance mechanisms. The identification of specific domains and functional sites within these proteins provides valuable insights into their involvement in various biological processes.

3.4. Structural Classification and Analysis Using CATH:

Figure 3: 1JJ4 Classification Representation of Alpha Domain in Proteomics (CATH Server)

Figure 4: 1JJ4 Highly Populated Superfamilies Representation of Alpha Domain in Proteomics (CATH Server)

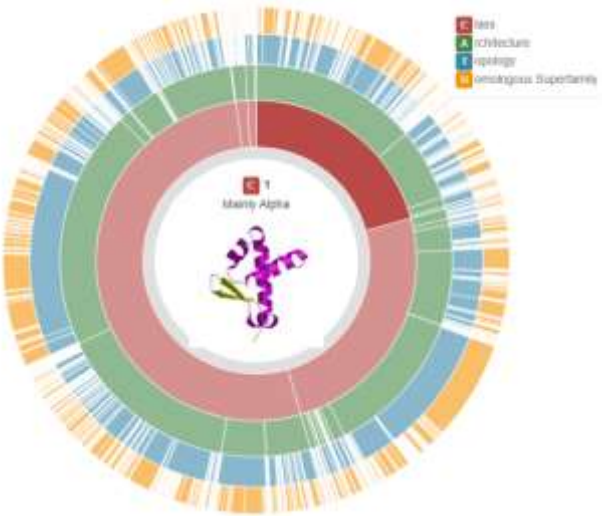


Figure 5: 1JJ4 Classification Representation of Beta Domain in Proteomics (CATH Server)

SEEJPH Volume XXVI, S2, 2025, ISSN: 2197-5248; Posted:03-02-2025

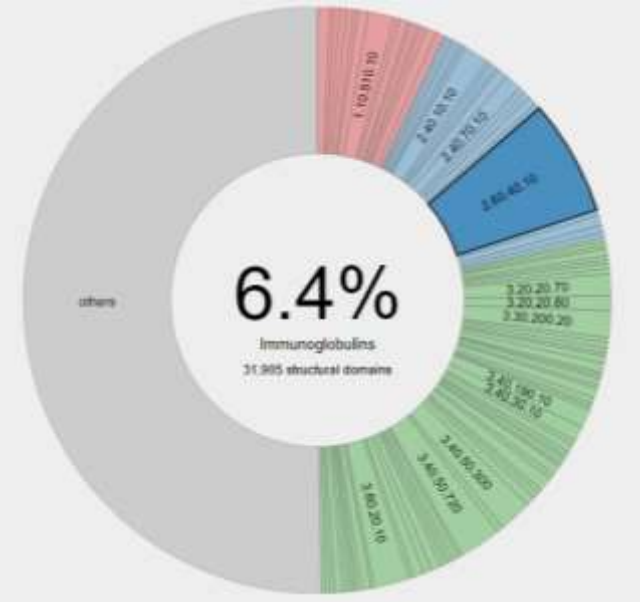


Figure 6: 1JJ4 Highly Populated Superfamilies Representation of Beta Domain in Proteomics (CATH Server)

The domains A and B of protein 1JJ4 are analyzed using the CATH Protein Structure Classification database, which provides useful insights into their structural and functional properties. CATH divides these domains into hierarchical levels, including Class, Architecture, Topology, and Homologous superfamily, to aid in understanding their evolutionary linkages and structural similarities. By analyzing domains A and B, we can identify similar structural motifs such as mixed alpha-helix, beta-sheet structure, and evolutionary patterns that may indicate functional parallels or differences. This analysis underscores the importance of structural databases in predicting protein behavior and interactions of various domains.

3.5. 3D Structure Analysis Using Protein DatabaseSum:

The PDBSum analysis of 1JJ4 shows the structural features, which include an alpha helix and a small beta-sheet, contribute to the binding interface's overall structure. It also reveals key residues and interactions, such as hydrogen bonds and hydrophobic interactions, which stabilize the protein-DNA complex. The protein-protein interface statistics showed 17 residues involved in the chain A interaction, while 18 residues were involved in the chain B interaction. Chain A covers an interface area of 812 (A^2), whereas chain B covers 809 (A^2). There are no salt bridges and disulfide bonds observed in this interface. But 14 hydrogen bonds stabilize the complex, and a total of 128 non-bonded contacts contribute to the interaction. This comprehensive knowledge of the B-chain interactions and structure sheds light on the molecular mechanisms of viral replication and pathogenicity, potentially guiding the development of targeted antiviral treatments. Overall, based on the PDBSum analysis of the protein 1JJ4, we can conclude that the B-chain is integral to the specific binding and interaction with its DNA target.

3.6. Sequence Alignment Visualizations Analysis:

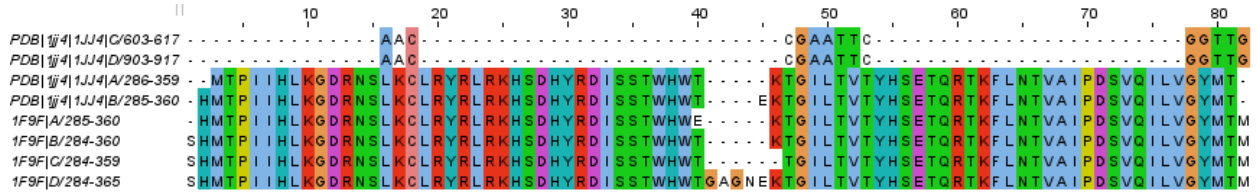


Figure 7: MuscleWS Alignment of 1JJ4 and 1F9F

The above CLUSTAL colour scheme helps in visualizing conserved and variable regions effectively and it represents as follows; Blue: Hydrophobic residues (L, V, M), Red: Positively charged residues, Green: Polar residues, Purple: Aromatic residues, Magenta: Proline (P), Yellow: Cysteine (C), Cyan: Glycine (G)

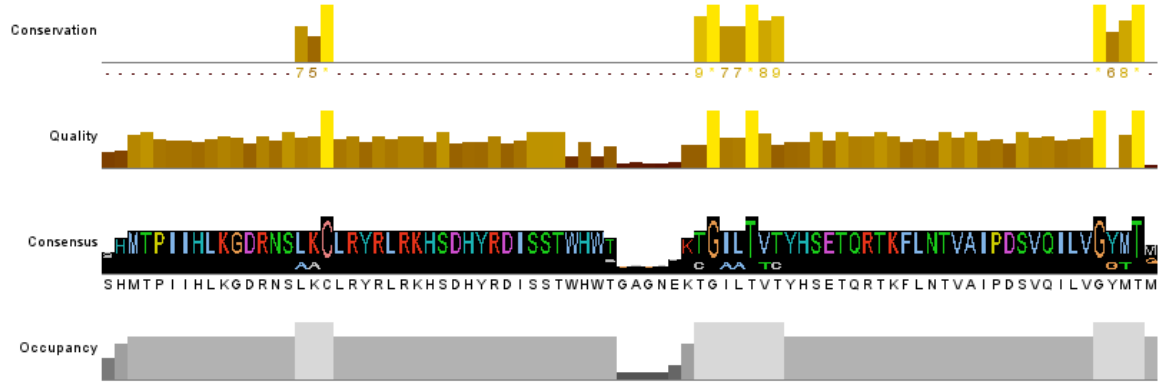


Figure 8: MuscleWS Jalview Analysis of 1JJ4 and 1F9F

The sequences 1JJ4|C and 1JJ4|D are identical, indicating a high level of conservation. The same is true for the sequences within the 1JJ4|A and 1JJ4|B pairs, which also show high conservation except for minor variations. This high level of conservation suggests that these sequences likely play critical roles in maintaining the structural integrity and function of the proteins. Across the sequences 1JJ4|A, 1JJ4|B, 1F9F|A, 1F9F|B, 1F9F|C, and 1F9F|D, several regions are highly conserved, such as the segments: MTPIIHLKGDRNSLKCLRYRLRKHSDHYRDISSTWH, which suggests these regions are critical for the proteins' structural stability or function. The sequences from 1F9F exhibit similar patterns with conserved regions, suggesting structural homology across these proteins. There are slight variations in specific residues, such as the substitution in the sequence 1F9F|A compared to 1JJ4|A (from W to E) at the 47th position and the presence of additional residues/insertions in 1F9F|D GAGNE at the 39th- 43rd positions, which might affect protein structure or function. Some minor differences were observed, particularly in the N-terminal and C-terminal regions of the sequences. For instance, the N-terminal regions of the sequences, particularly those from 1F9F, show variations in length and sequence (end in M). Likewise, 1JJ4|D has an additional G residue at the C-terminus compared to other chains. This substitution or insertion observed in these sequences could alter the interaction surfaces or binding affinities, influencing the protein's function or interactions with other molecules, or can be used to study mutagenicity. Further, the sequences show a high degree of occupancy with minimal gaps, indicating a stable and reliable alignment. Overall, the Jalview MuscleWS alignment of the PDB sequences 1JJ4 and 1F9F provides valuable insights into these proteins' structural and functional homology.

3.7. Determination of Cleavage Sites Prediction in Protein Sequence:

Your sequence:

```
        10         20         30         40         50         60
GSHMTPIIHL KGDRNSLKCL RYRLRKHSDH YRDISSTWHW TGAGNEKTGI LTVTYHSETQ

        70         80
RTKFLNTVAI PDSVQILVGY MTM
```

The sequence is 83 amino acids long.

Figure 9: Expasy Peptide Cutter Analysis of 1JJ4

The Expasy Peptide Cutter tool evaluated the 1JJ4 protein sequence and identified probable cleavage sites using a variety of proteolytic enzymes resulting in peptides of sequence. Each enzyme has particular cleavage preferences, and the tool has identified the sites where these enzymes may cleave the protein. The numbers following the enzyme name represent probable cleavage sites in the protein sequence. The cleavage positions are denoted by amino acid numbers. For example, the first line indicates that Arg-C proteinase may cleave the protein at locations 14, 21, 23, 25, 32, and 61.

3.8. Identification of Epitopes Analysis:

3.8.1. T Cell Epitope Prediction Tools

Class I Immunogenicity

Masking: **default**

Masked variables: **[1, 2, 'cterm']**

Table 1: T-Cell Epitope Prediction of 1JJ4

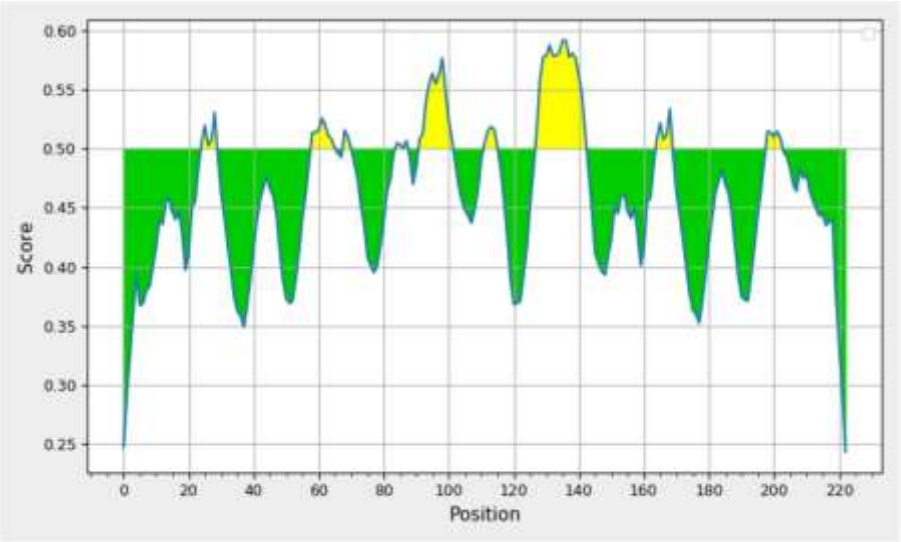
Peptide	Length	Score
RTKFLNTVAI	10	0.12494
GSHMTPIIHL	10	0.1173
YRDISSTWHW	10	0.0613
PDSVQILVGY	10	0.04884
LTVTYHSETQ	10	0.03181
MTM	3	0
TGAGNEKTGI	10	-0.01244
RYRLRKHSDH	10	-0.25017
KGDRNSLKCL	10	-0.33206

To identify potential T-cell epitopes within the protein 1JJ4, we employed the IEDB T-cell Epitope Prediction and Analysis Tool. This tool utilizes a combination of algorithms to predict peptide binding affinities to MHC class I molecules which present antigens (epitopes) to CD8+ cytotoxic T-cells. The analysis was performed with default masked variables including positions 1, 2, and the C-terminus. The strong binding affinity is indicated by a negative score, and the weak binding affinity is shown by a positive value. Peptides with negative scores, including TGAGNEKTGI, RYRLRKHSDH, and KGDRNSLKCL, are therefore more likely to attach to MHC Class I molecules and possibly trigger a more robust T cell immunological response. The peptides with positive scores (RTKFLNTVAI, GSHMTPIIHL, YRDISSTWHW, PDSVQILVGY, LTVTYHSETQ, and MTM) indicate that the epitope has a moderate to weak immunogenic potential when presented by MHC Class I molecules. They display some chance of being identified by CD8+ T-cells, even though they might not be the most effective epitope for eliciting a T-cell response. However, further experimental validation is necessary to confirm their immunogenicity.

3.8.2. B Cell Epitope Prediction Tools

Bepipred Linear Epitope Prediction 2.0

Center position: 4 **Threshold: 0.500**



Average: 0.456

Minimum: 0.243 **Maximum:** 0.592

Figure 10: Graphical Analysis of B-cell Epitope Prediction (Threshold)

To identify potential B cell epitopes within protein 1JJ4, we utilized the IEDB B Cell Epitope Prediction and Analysis Tool, specifically the Bepipred Linear Epitope Prediction 2.0. The analysis aimed to discover linear epitopes based on the sequence characteristics of the protein. This tool generated a graphical representation of the predicted prominent B cell epitopes, along with numerical data summarizing the epitope prediction scores. The green regions indicate non-epitope areas with lower threshold scores, while the yellow regions highlight the potential epitope areas with high threshold scores. The average score across the protein sequence is 0.456, which is below the threshold of 0.500, suggesting that the overall sequence has a moderate potential for epitope formation. The lowest score observed is 0.243, indicating regions with the weakest epitope potential. The highest score observed is 0.592, indicating regions with the strongest epitope potential. The figure highlights the regions with high epitope scores, indicating potential antibody binding sites.

3.9. Computation Analysis Using Biopython:

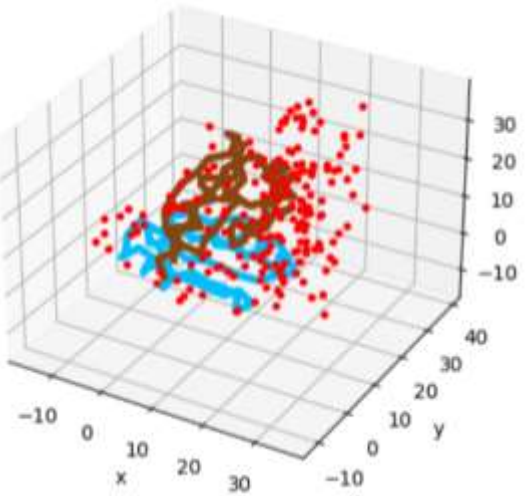


Figure 11: Parsed PDB Structure of 1JJ4 using Biopython

A detailed structural analysis of the HPV 18 E2 DBD (1JJ4) and a potential drug molecule was performed using Biopython. We successfully retrieved the nucleotide sequences of both the HPV18 E2 DBD (NCBI accession code: 1JJ4) and its DNA target from the NCBI GenBank database. The 3D structure of the 1JJ4 protein was fetched from the RCSB PDB database using Biopython's PDB module. Further, the generated Ramachandran plot, created with Biopython's Plot module,within the favored and allowed regions, indicating a high-quality stereochemical structure for the protein. Also, it facilitated the extraction of crystallographic data from the CIF file. The data revealed key crystallographic parameters essential for structural validation. Furthermore, the PDBXML format of protein 1JJ4 was parsed to extract detailed structural information, including atomic coordinates, residue names, and bonding interactions. Using Biopython libraries, we performed an extensive structural analysis to identify key features and interactions within the E2 DNA-binding domain. The analysis highlighted several critical interactions between the E2 protein and its DNA target, including hydrogen bonds, electrostatic interactions, and hydrophobic contacts that stabilize the protein-DNA complex.

4. Conclusion

By employing Biopython and a suite of computational tools, we conducted a comprehensive NGS analysis of the HPV-18 E2 DBD interacting with its DNA target. This study delves into the genetic, immunological, structural, and functional characteristics of this viral protein, offering valuable insights into its molecular mechanisms in HPV infection and disease progression as well as its potential as an immunogenic target.

The sequence analysis tools BLAST performed to identify homologous sequences and assess the evolutionary conservation of the E2 DBD. The resulting high Bit-score indicated a quantitative measure of sequence similarity, emphasizing the significance of conserved residues in facilitating the DNA-binding function of the E2 protein. Multiple sequence alignments were performed using the multiple alignment and COBALT tools, which identified conserved motifs and structural elements crucial to the E2 DNA-binding function. The alignments revealed crucial residues that are involved in interactions between proteins and DNA. The functional annotation tool InterProScan was used to identify functional domains and motifs in protein sequences, thereby confirming the presence of the E2 DBD and predicting other functional areas within the protein with its role in viral replication. Further, structural analysis tools like PDBSum provided precise summaries of protein structures, whereas CATH provided a hierarchical classification of protein domain structures, which aided in viewing the 3D structure of the E2 DBD and determining the spatial arrangement of key residues. The visualization and annotation using Jalview haves enriched our comprehension of the E2 protein's interaction dynamics.

The Expasy Peptide Cutter tool predicted proteolytic cleavage sites within the E2 DNA-binding domain, which aids in understanding how the protein is processed and presented in the immune system. Moreover, the immunological analysis tool such as the IEDB T Cell Epitope Prediction Tool revealed potential T cell epitopes in the E2 DBD, indicating regions that may trigger a cellular immunological response, and the Bepipred Linear Epitope Prediction 2.0 tool identified linear B cell epitopes, which are likely to be recognized by B cells and contribute to antibody-mediated humoral immunity. These identified epitopes could serve as valuable targets for the development of diagnostic tools, vaccines, and immunotherapies to combat HPV-associated diseases. Through the identification of conserved regions and potential functional motifs within the E2 protein, we gained valuable insights for targeted drug design. By leveraging the open-source Biopython library, we efficiently processed NGS data and integrated bioinformatics tools. This streamlined approach enabled the analysis of large-scale biological data, including protein sequences and structures. The use of Biopython facilitated efficient data handling and integration of various analysis pipelines.

Overall, the comprehensive NGS analysis of the HPV-18 E2 DNA-binding domain, supported by a robust suite of computational tools, advances our understanding of the molecular underpinnings of HPV-associated carcinogenesis. These findings underscore the potential of computational biology and therapeutic approaches and enhance our ability to design targeted interventions against HPV-related diseases. By leveraging advanced bioinformatics tools, this study contributes to the broader efforts in cancer research and personalized medicine, offering a future experimental validation and clinical applications.

References:

[1] Almeida, A. M., Queiroz, J. A., Sousa, F., & Sousa, Â. (2019). Cervical cancer and HPV infection: Ongoing therapeutic research to counteract the action of E6 and E7 oncoproteins. *Drug Discovery Today*, *24*(10), 2044–2057. https://doi.org/10.1016/j.drudis.2019.07.011

[2] McLaughlin-Drubin, M. E., & Münger, K. (2009). Oncogenic activities of human papillomaviruses. *Virus Research*, *143*(2), 195–208. https://doi.org/10.1016/j.virusres.2009.06.008

[3] Soeda, E., Ferran, M. C., Baker, C. C., & McBride, A. A. (2006). Repression of HPV16 early region transcription by the E2 protein. *Virology*, *351*(1), 29–41. https://doi.org/10.1016/j.virol.2006.03.016

[4] Šarenac, T., & Mikov, M. (2019). Cervical Cancer, Different Treatments and Importance of Bile Acids as Therapeutic Agents in This Disease. *Frontiers in Pharmacology*, *10*, 484. https://doi.org/10.3389/fphar.2019.00484

[5] Sun, D., Gao, W., Hu, H., & Zhou, S. (2022). Why 90 % of clinical drug development fails and how to improve it? *Acta Phramaceutica Sinica B*, *12*(7), 3049-3062. https://doi.org/10.1016/j.apsb.2022.02.002

[6] Kumari, U., Bandbe, T., & Brahma, D. (2024, September 1). Structure-Based Drug Design and Molecular Target Prediction of Transthyretin Variant A97S Complexed. *Internation Journal of Emerging Technologies and Innovative Research (JETIR)*, ISSN:2349-5162, 11(9), a316-a326. http://doi.one/10.1729/Journal.41352

[7] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

[8] Madej, T., Lanczycki, C. J., Zhang, D., Thiessen, P. A., Geer, R. C., Marchler-Bauer, A., & Bryant, S. H. (2014). MMDB and VAST+: Tracking structural similarities between macromolecular complexes. *Nucleic Acids Research*, *42*(D1), D297–D303. https://doi.org/10.1093/nar/gkt1208

[9] Sievers, F., & Higgins, D. G. (2018). Clustal Omega for making accurate alignments of many protein sequences. *Protein Science*, *27*(1), 135–145. https://doi.org/10.1002/pro.3290

[10] Papadopoulos, J. S., & Agarwala, R. (2007). COBALT: Constraint-based alignment tool for multiple protein sequences. *Bioinformatics*, *23*(9), 1073–1079. https://doi.org/10.1093/bioinformatics/btm076

[11] Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

[12] Orengo, C., Michie, A., Jones, S., Jones, D., Swindells, M., & Thornton, J. (1997). CATH – a hierarchic classification of protein domain structures. *Structure*, *5*(8), 1093–1109. https://doi.org/10.1016/S0969-2126(97)00260-8

[13] Laskowski, R. A., Jabłońska, J., Pravda, L., Vařeková, R. S., & Thornton, J. M. (2018). PDBsum: Structural summaries of PDB entries. *Protein Science*, *27*(1), 129–134. https://doi.org/10.1002/pro.3289

[14] Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—A multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189–1191. https://doi.org/10.1093/bioinformatics/btp033

[15] Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., De Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., … Stockinger, H. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research*, *40*(W1), W597–W603. https://doi.org/10.1093/nar/gks400

[16] Vita, R., Mahajan, S., Overton, J. A., Dhanda, S. K., Martini, S., Cantrell, J. R., Wheeler, D. K., Sette, A., & Peters, B. (2019). The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Research*, *47*(D1), D339–D343. https://doi.org/10.1093/nar/gky1006

[17] Uma kumari, Devanshi Gupta, Computational analysis of Glioma Transcriptome and proteome with Bio python ;2023Volume-13, Issue-1, Pp 1-14,IJBTR

[18] Kumari Uma & Tripathi, Kartik. (2023). Computational Analysis and Molecular Docking Approach for Liver Cancer. International Journal for Research in Applied Science and Engineering Technology. 11. 1828 to 1633. 10.22214/ijraset.2023.54927.

[19] Uma kumari,Meenakshi Pradhan,Saptarshi Mukherjee,Sreyashi Chakrabarti;,Ngs Analysis approach for neurodegenerative disease with Biopython",2023 ,volume 10,issue 9 , http://doi.one/10.1729/Journal.36043

[20] Vineeta Johri , Ekta Tyagi , Uma Kumari. (2024). Next Generation Sequencing Analysis of NSP5 NSP6 and NSP7 SARS CoV-2 Nonstructural Protein with Bio-Python. *South Eastern European Journal of Public Health*, 1653–1666. https://doi.org/10.70135/seejph.vi.2936 Dec 2024

[21] Uma Kumari,Gurpreet Kaur *et al*,"Biopython/Network Of Protein Identification And NGS Analysis Of Glioma Cancer ATP Competitive Type III C-MET Inhibitor : Volume 11, Issue 2 : 27-Jun-2024 :pp 41-51 : http://doi.one/10.1729/Journal.40229