JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

MOLECULAR TARGET IDENTIFICATION AND NGS DATA ANALYSIS OF HUMAN PTPN14 IN A COMPLEX WITH CR3 DOMAIN OF HPV18 E7

Dia Saji[1] Uma Kumari[1]

[1]*Project Trainee at Bioinformatics Project and Research Insitute, Noida - 201301, India*

[2]*Senior Bioinformatics Scientist, Bioinformatics Project and Research Insitute, Noida - 201301, India*

[1*]Corresponding Author Uma kumari

Abstract: Cervical carcinoma is one of the major causes of cancer-related mortality among women in whole world, particularly in regions lacking robust screening and vaccination programs. Human papillomavirus (HPV), especially high-risk strains like HPV18, has been known as a primary etiological agent in the development of this malignancy. The E7 oncoprotein encoded by HPV18 is known to interfere with host tumor suppressor proteins, notably Protein Tyrosine Phosphatase Non-Receptor Type 14 (PTPN14). This study investigates the molecular interaction between PTPN14 and the CR3 domain of HPV18 E7 to explore therapeutic targeting opportunities. We utilized crystallographic data from the Protein Data Bank to assess the binding interface between these two proteins and conducted molecular docking studies with clinically relevant chemotherapeutic agents. Through computational docking using molecular docking, three potential ligands Mitomycin, Topotecan, and Paclitaxel C were evaluated for their binding affinity to the interaction interface. Topotecan exhibited the most favorable docking score, suggesting a strong binding affinity and potential disruption of the oncogenic protein complex. Structural validation was carried out using structure Vlaidation and Ramachandran plot analysis via PROCHECK, which confirmed the high stereochemical quality of the model. The binding pockets were visualized using PyMOL, and protein-ligand interactions were assessed for hydrogen bonding and electrostatic compatibility. Next-Generation Sequencing (NGS) data provided supporting evidence for the significance of PTPN14 in cervical carcinogenesis. Our findings suggest that targeting the PTPN14–HPV18 E7 interaction could serve as a novel strategy in the treatment of HPV-associated cervical cancer. The identified compounds, particularly Topotecan, demonstrate potential for further in-vitro and in-vivo validation as lead candidates. This study emphasizes the utility of integrative bioinformatics and structural biology in uncovering molecular mechanisms and therapeutic vulnerabilities in viral oncogenesis.

Keywords: Cervical Cancer; Insilico Drug designing; Molecular Docking; NGS Analysis; Ramachandran Plot; Sequence Alignment BioPython

INTRODUCTION

Cervical cancer poses a persistent threat to global women's health, ranking as the fourth most frequently diagnosed cancer and the fourth leading cause of cancer-related deaths in women. The burden of this disease disproportionately affects low- and middle-income countries (LMICs), where organized screening and preventive strategies are often inadequate. A majority of cervical cancer cases are directly linked to persistent infections by high-risk strains of the human papillomavirus (HPV), with HPV types 16 and 18 accounting for nearly 70% of the total global incidence. Among these, HPV18 is of particular concern due to its aggressive oncogenic potential and its association with adenocarcinoma, a subtype more difficult to detect and treat compared to squamous cell carcinoma [1, 2, 3].

HPV is a small, double-stranded DNA virus that infects epithelial tissues and promotes cellular proliferation by disrupting key regulatory checkpoints in the host cell cycle. Two major viral oncoproteins, E6 and E7, are central to HPV-mediated carcinogenesis. The E6 protein targets the tumor suppressor p53 for degradation, while E7 interacts with several cellular targets including the retinoblastoma protein (pRb), effectively disabling normal control mechanisms governing cell division and apoptosis. In recent years, increasing attention has been directed toward another important host target of the E7 protein: Protein Tyrosine Phosphatase Non-Receptor Type 14 (PTPN14). PTPN14 is a cytoplasmic tyrosine phosphatase with tumor suppressor functions. It plays a critical role in maintaining cellular architecture, regulating cytoskeletal dynamics, and controlling the Hippo signaling pathway, which governs cell proliferation, organ size, and tissue homeostasis. The Hippo pathway effector proteins YAP and TAZ are transcriptional coactivators that, when unchecked, can promote oncogenic transformation. PTPN14 functions to inhibit YAP/TAZ activity, thus serving as a barrier to uncontrolled cell growth. The degradation or functional inactivation of PTPN14 by HPV18 E7 disrupts this control and contributes significantly to the malignant phenotype observed in HPV-associated cancers [4, 5, 6, 7].

Recent structural biology studies, particularly those involving the protein complex with PDB ID: 6IWD, have revealed how the CR3 domain of HPV18 E7 binds directly to PTPN14, interfering with its tumor-suppressive activity. This direct protein-protein interaction opens a window for therapeutic intervention. Disrupting or weakening this interaction has the potential to restore PTPN14 functionality and hinder HPV-mediated oncogenesis. However, this approach requires a detailed molecular understanding of the binding interface and the identification of suitable molecular agents capable of targeting this interaction with high specificity and efficacy. The landscape of cervical cancer management has significantly evolved with the advent of prophylactic vaccines, particularly Gardasil and Cervarix, which protect against the most common oncogenic HPV types. Nonetheless, the vaccines are preventive, not therapeutic, and their efficacy is dependent on coverage before HPV exposure. Thus, for individuals already infected or presenting with high-grade lesions or invasive cancer, treatment options such as surgery, radiation therapy, and chemotherapy remain the mainstay of care. Despite their utility, these treatments come with considerable side effects and may not always guarantee a favorable prognosis, especially in advanced stages. This underscores the need for novel, targeted therapies that address the molecular underpinnings of HPV-driven malignancy. One promising approach lies in the integration of computational biology techniques such as molecular docking, structure-based drug design, and Next-Generation Sequencing (NGS). These tools enable a comprehensive understanding of the cancer genome, identification of key driver mutations, and exploration of protein-ligand interactions that may be exploited for therapeutic development [8, 9].

The understanding of HPV's role in oncogenesis has evolved significantly over the past several decades. Parallelly, advancements in structural biology and bioinformatics have enabled researchers to dissect protein interactions at atomic detail and explore therapeutic interventions using computational tools. The table below summarizes the historical trajectory of key discoveries in HPV biology and the evolution of structural bioinformatics technologies relevant to this study [15].

Table 1: Historical trajectory of key discoveries in HPV biology and the evolution of structural bioinformatics

Year	Milestone	Field	Description
1970s	Discovery of Papillomaviruses in warts	Virology	HPV initially linked to benign lesions; foundation laid for oncogenic research.
1983	HPV16 and HPV18 identified in cervical cancer tissues	Cancer Virology	Harald zur Hausen's groundbreaking discovery established HPV as a cause of cervical cancer.
1990s	Cloning and functional analysis of E6 and E7 oncoproteins	Molecular Oncology	E6 and E7 shown to inactivate tumor suppressors (p53, pRb); mechanism of viral carcinogenesis elucidated.
2000	Launch of Protein Data Bank (PDB) in its modern format	Structural Biology	Made high-resolution protein structures publicly accessible, including viral-host complexes.
2006	FDA approval of Gardasil (HPV vaccine)	Public Health	First vaccine targeting high-risk HPV types introduced, drastically reducing infection rates.
2010	AutoDock Vina introduced	Bioinformatics	Open-source docking software improving accuracy and speed in ligand-protein interaction predictions.
2015	CB-Dock and cavity detection algorithms developed	Computational Docking	Enhanced ligand docking precision by integrating pocket prediction with blind docking.
2020s	Integration of NGS and structural modeling	Genomics & Proteomics	Combined data-driven approaches used to identify mutational landscapes and druggable targets in viral oncology.

Molecular docking is a computational technique that predicts the preferred orientation of one molecule (typically a drug candidate) when bound to a second molecule (usually a protein target). It allows researchers to evaluate the strength and nature of these interactions by calculating binding affinities and visualizing atomic-level contact points. When applied to the PTPN14–HPV18 E7 complex, molecular docking can identify small molecules capable of interfering with or modulating the interaction, potentially reactivating the tumor suppressive functions of PTPN14. In the present study, we have employed structure-based molecular docking to evaluate the binding potential of three known anticancer agents Mitomycin, Paclitaxel C (Taxol C), and Topotecan against the PTPN14-HPV18 E7 interface. These drugs were chosen based on their established therapeutic profiles and diverse mechanisms of action, offering the possibility of repurposing existing compounds for HPV-targeted therapy. Among them, Topotecan, a topoisomerase I inhibitor used in cervical and ovarian cancers, demonstrated the highest docking affinity, suggesting a favorable binding profile with the PTPN14 interface involved in E7 interaction. Complementary to molecular docking, we utilized NGS data analysis to contextualize the relevance of PTPN14 in cervical cancer. NGS offers deep insights into genomic alterations, expression profiles, and molecular pathways dysregulated in cancer. By integrating NGS data with protein structural analysis, we aim to bridge the gap between molecular biology and therapeutic application, thereby enabling a precision medicine approach. To ensure the accuracy and reliability of the protein-ligand interaction predictions, we employed several validation techniques. Structure Validation, a tool for assessing the quality of protein crystal structures, was used to detect potential regions of error based on nonbonded atomic interactions. Additionally, Ramachandran plot feature were used to evaluate the stereochemical integrity of the modeled protein. A model is considered high-quality if the majority of its residues fall within favored or allowed regions, and our structure satisfied these criteria with 92% of residues in the most favorable regions. Visualization of the docking interactions and identification of active binding pockets were achieved using PyMOL, a powerful molecular graphics system. Electrostatic surface representations and hydrogen bond mapping provided further clarity into the nature of drug-protein interactions. Moreover, we employed RasMol to explore secondary structural features such as alpha-helices and beta-sheets within the protein, which helped in confirming the integrity and folding of the active site regions [10, 11, 12, 13, 14, 15,32,38,39].

The novelty of this research lies in its integrative approach bridging structural biology, computational docking, and genomics—to identify and validate molecular targets in HPV-associated cervical cancer. By focusing on the interaction between PTPN14 and HPV18 E7, the study not only enhances our understanding of viral oncogenesis but also proposes a new direction for therapeutic intervention. Our results underscore the therapeutic promise of targeting viral-host protein interactions, particularly those involved in disrupting tumor suppressor

pathways. While traditional anticancer drugs often exert broad cytotoxic effects, rational drug design guided by molecular docking enables the development of highly specific agents with minimal off-target consequences. This approach is particularly important in treating virus-induced cancers, where the oncogenic mechanism is often precise and well-defined. Despite promising computational findings, the journey from in-silico prediction to clinical application involves rigorous experimental validation. The identified compounds must undergo in-vitro cell-based assays to confirm binding and functional inhibition, followed by in-vivo studies to assess pharmacodynamics, bioavailability, and toxicity. However, in-silico studies such as ours serve as an efficient and cost-effective starting point, narrowing down the pool of candidate compounds and informing future research. Above all, cervical cancer, though preventable and treatable when caught early, continues to exact a heavy toll due to gaps in screening, vaccination, and access to care. Novel therapeutic strategies are needed to complement existing approaches, particularly in advanced or vaccine-resistant cases. The interaction between PTPN14 and HPV18 E7 represents a compelling target for drug development, and our study offers a framework for how computational tools can accelerate the discovery of potential inhibitors. As we progress toward precision oncology, the integration of bioinformatics, structural biology, and pharmacology will become increasingly vital in combating complex diseases such as HPV-driven cervical cancer [14, 15].

MATERIALS AND METHODS

1. Retrieval of Protein Structure and ligands

The target protein complex investigated in this study is the human Protein Tyrosine Phosphatase Non-Receptor Type 14 (PTPN14) in interaction with the CR3 domain of Human Papillomavirus 18 (HPV18) E7 oncoprotein. The crystallographic structure of this protein-protein complex was obtained from the Protein Data Bank (PDB ID: 6IWD). This structure provides a detailed atomic representation of the interaction between host and viral proteins, essential for identifying potential druggable sites [16].

Three ligands with established roles in cancer therapy were selected: Mitomycin, Topotecan, and Paclitaxel C (Taxol C). These compounds were chosen based on their differing mechanisms of action and known clinical relevance in treating cervical and related cancers. The 3D structures of these ligands were retrieved from the PubChem database in SDF format [17].

2. Structural Visualization and Binding Site Analysis

Active binding sites were highlighted using surface and stick models to provide a clear spatial representation of ligand positioning. Electrostatic surface maps were generated to analyze polarity and charge complementarity between the protein and ligands. Secondary structural motifs including α-helices and β-sheets were examined using RasMol, which allowed for detailed observation of secondary structure conservation in the binding regions [18, 19, 20,27,28,29].

3. Structure Validation and Quality Assessment

To confirm the reliability of the crystal structure and the accuracy of docking simulations, we applied multiple validation tools:

- ERRAT was used to evaluate the quality of nonbonded atomic interactions in the protein model. A high overall quality factor indicated structural integrity [21].
- SAVES server, generated a Ramachandran plot to assess the stereochemical quality of the protein. These metrics confirm the high stereochemical reliability of the model used in docking analysis [22,31,35,34,36,38].

4. Next-Generation Sequencing (NGS) Analysis

NGS data was obtained from publicly available cancer genomics repositories, focusing on gene expression profiles and mutation data for PTPN14 in cervical cancer. Sequence analysis was conducted using BioPython scripts, and conserved domains were identified via InterProScan. COBALT has been performed to investigate the phylogenetic history of the protein. The integration of NGS data with docking results supports the biological significance of the PTPN14–HPV18 E7 interaction and validates the potential of ligand intervention in disrupting this oncogenic mechanism [23, 24, 25,31,32,35].

5. Molecular Docking via Cavity-Detection Based Docking

Molecular docking simulations were carried out using the Docking server, a cavity-detection guided blind docking platform that uses AutoDock Vina as its core engine. Server Docking automatically identified five potential binding pockets and performed docking simulations against each site. Binding affinities were calculated in kcal/mol, and docking poses were ranked based on their Vina scores [26,35,36,38,39].

RESULT AND DISCUSSION

1. Structural and Binding Site Analysis

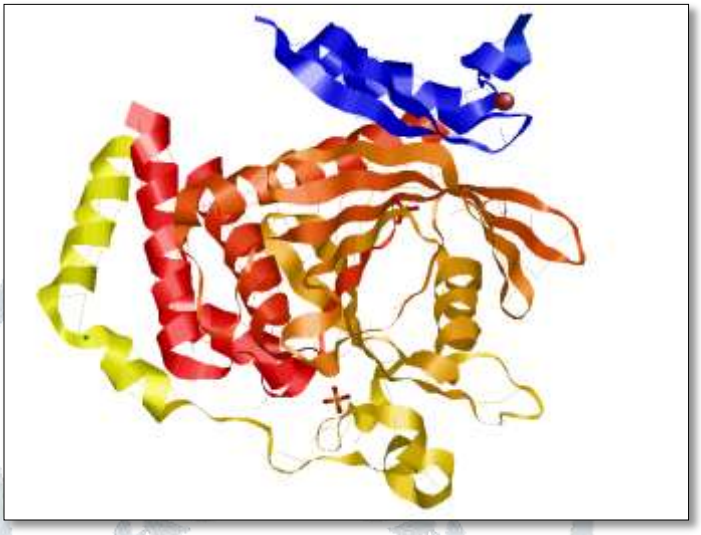

Figure 1: Hydrogen bond analysis

It illustrates the three-dimensional conformation of the protein complex comprising human PTPN14 (rendered in a rainbow gradient from N-terminus in blue to C-terminus in red) and the CR3 domain of the HPV18 E7 oncoprotein (highlighted in deep blue). The structural visualization provides critical insight into the stabilization of the protein-ligand complex through hydrogen bonding. Key hydrogen bonds were identified within the binding interface, with visible donor-acceptor interactions contributing to molecular stability and specificity. Notably, polar residues within the core of the active site region form hydrogen bridges with adjacent secondary structure elements, such as α-helices and β-sheets. These interactions are essential for maintaining the proper folding and alignment of the catalytic and binding domains.

The blue region (CR3 domain) is in close proximity to red and orange helices of PTPN14, indicating an interaction zone likely involved in hydrogen bonding and possibly van der Waals contacts. The spatial arrangement confirms that the ligand or interacting peptide is tightly bound within a hydrophilic pocket supported by electrostatic complementarity. Additionally, a water-mediated hydrogen bond network is suggested by the orientation of loop regions toward the core interface. The stability conferred by these interactions may be critical for the functional inhibition or modulation of PTPN14 activity by viral components or ligands. Overall, the structural evidence highlights the role of hydrogen bonding in supporting high-affinity binding at the PTPN14–HPV18 E7 interface. These interactions form the foundation for structure-guided drug design targeting this crucial oncogenic complex.

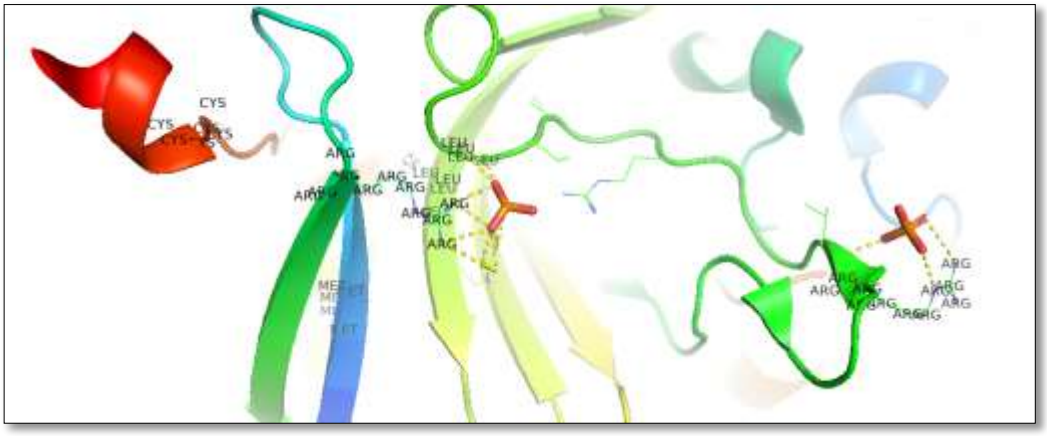

Figure 2: Phosphate Group (PO_4^{3-}) Interaction with Arginine Residues via Hydrogen Bonding

It presents a high-resolution molecular interaction map highlighting the binding behavior of a phosphate group (PO_4^{3-}) within the active domain of the protein complex. The phosphate is depicted in red and orange stick representation, positioned at the interface of multiple β-strands and loop regions. A network of hydrogen bonds (represented by yellow dashed lines) is observed between the negatively charged oxygen atoms of the phosphate group and the side chains of positively charged arginine (ARG) residues. These interactions play a crucial role in stabilizing the phosphate within the protein's binding cleft. The spatial geometry and the electrostatic compatibility between the phosphate group and nearby residues such as ARG, LEU, and MET indicate a strong polar attraction that enhances binding affinity.

This interaction site is surrounded by both α-helical and β-sheet secondary structures, suggesting a structurally conserved and functionally important binding region. The arginine-rich environment, particularly in the lower and right regions of the figure, demonstrates converging side chains engaging in bidentate hydrogen bonding with the phosphate. This reinforces the structural anchoring and possibly contributes to

catalytic or regulatory functions. Cysteine residues and a methionine cluster are observed in adjacent structural motifs, potentially contributing to local stabilization through hydrophobic or redox-sensitive interactions. The presence of such dense hydrogen bonding suggests that the phosphate group is tightly coordinated and may act as a cofactor mimic, a phosphorylation site, or a part of a ligand binding motif. This localized interaction landscape offers a promising target for the design of inhibitors that mimic phosphate-containing substrates or transition states in enzyme activity.

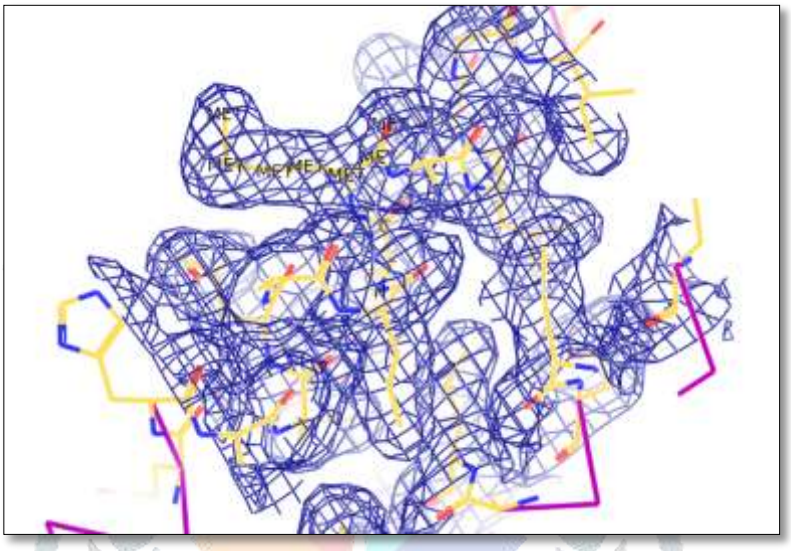

Figure 3: Electron Density Map Highlighting the Active Site of the Protein (PyMol)

It illustrates the electron density distribution surrounding the active site of the protein structure using the roving density feature in PyMOL. The mesh-like blue contours represent the electron density map, which corresponds to the experimental data derived from X-ray crystallography. This visualization is critical for validating the atomic positioning and confirming the accuracy of the modeled structure.

Within the map, key residues such as methionine (MET) and cysteine (CYS) are observed to align precisely within well-defined electron clouds, indicating high confidence in their atomic positions. The side chains of these residues are neatly enclosed by the electron density, affirming the quality of crystallographic refinement in this region. The backbone of the polypeptide chain, represented in yellow, closely follows the contour of the blue density, further supporting the structural integrity of the model. Side chains extending from this backbone also exhibit strong electron density coverage, particularly in functionally relevant residues that are likely to participate in catalytic activity or ligand binding. The clear definition of the active site architecture as shown here lays the foundation for reliable computational docking and future mutagenesis studies.

It displays the electrostatic surface potential of the protein, visualized using PyMOL. This representation helps in identifying the charge distribution across the molecular surface, which is crucial for understanding potential ligand binding and protein-protein interactions. The color scale ranges from red (−3.600) indicating regions of strong negative electrostatic potential, to blue (0.400) indicating positively charged zones. Neutral or weakly charged areas appear in white. Notably, the red regions dominate the molecular surface, particularly at the concave center of the protein's topology, suggesting that these areas are enriched with acidic residues such as glutamate or aspartate. These negatively charged zones are potential hotspots for interaction with positively charged ligands, such as those containing arginine or lysine moieties. In contrast, small blue patches seen at the periphery may represent basic residues that contribute to electrostatic balance or serve in stabilizing charged ligands via complementary interactions. The distribution of surface charges also implies a high level of electrostatic anisotropy, possibly contributing to directional ligand docking.

2. Structural Validation

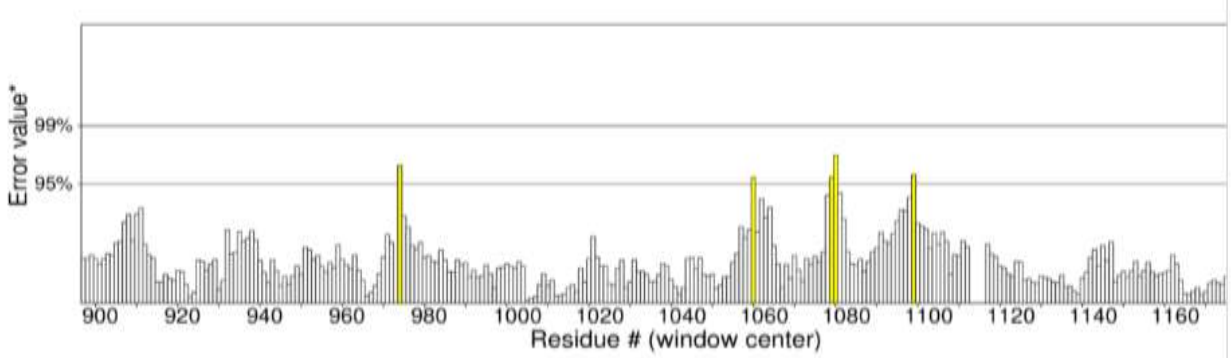

Figure 4: Structure Validation Analysis

The structural validation results obtained through the ERRAT2 program for Chain A of the protein structure . The plot shows the error value profile across the entire chain length, with each bar representing the error associated with a window of residues centered around the labeled position on the x-axis. The overall quality factor of the structure is reported as 96.865%, which exceeds the standard threshold of 95% for high-resolution models. This score signifies that the majority of the protein structure demonstrates excellent stereochemical reliability and that very few regions fall outside of acceptable error limits.

The horizontal lines in the plot represent the 95% and 99% confidence thresholds. Most of the error bars lie well below the 95% rejection limit, indicating that these regions of the model are highly accurate and structurally valid. Only a small number of segments—centered around residues ~960, ~1060, ~1080, and ~1100—show localized deviations where error values exceed the 95% threshold, marked in yellow. These regions may represent flexible loops or surface-exposed residues, which are more prone to variability or less defined electron density during crystallographic modeling. Such deviations are minimal and do not significantly compromise the model's global quality. They may be considered for refinement or closer inspection if these regions are functionally important or involved in ligand interaction. Overall, the structure is of high confidence and is deemed suitable for downstream applications, including docking and drug design. This ERRAT analysis reinforces the structural integrity of the protein used in this study

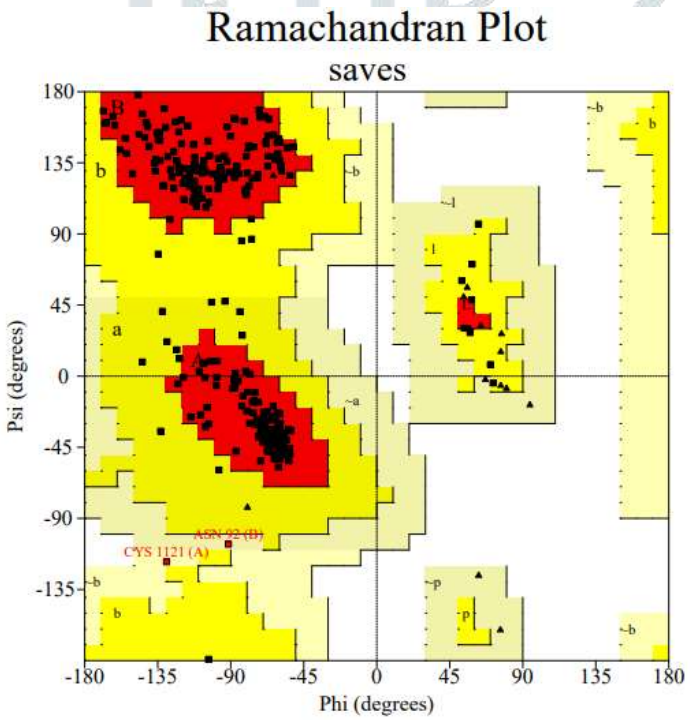

Figure 5: Representation of Ramachandran Plot

The plot evaluates the distribution of backbone dihedral angles (ϕ and ψ) for each amino acid residue, serving as a key metric for assessing stereochemical quality.

The statistical summary shown below the plot indicates that:

- 92.0% (276 residues) fall within the most favored regions (A, B, L), reflecting high stereochemical accuracy.
- 7.3% (22 residues) are in additionally allowed regions (a, b, l, p).
- 0.7% (2 residues) are in generously allowed regions (~a, ~b, ~l, ~p).
- Importantly, 0.0% of residues appear in the disallowed regions, indicating that the protein model lacks major backbone conformational outliers.

The black squares and triangles represent non-glycine and glycine residues, respectively. The clustering of most data points within red-shaded core areas suggests a properly folded and energetically favorable structure. The presence of no residues in disallowed regions further validates the high-resolution quality of the modeled protein backbone. Additionally, the data report that the model contains:

- 300 non-glycine/non-proline residues
- 18 glycine residues
- 14 proline residues
- 6 terminal residues (excluding Gly and Pro)
- Total of 338 residues analyzed

This high percentage of residues in favored and allowed regions exceeds the accepted quality benchmark for protein structures, confirming that Chain A is conformationally well-resolved and suitable for docking, simulation, and functional interpretation studies.

3. Next-Generation Sequencing (NGS) Analysis

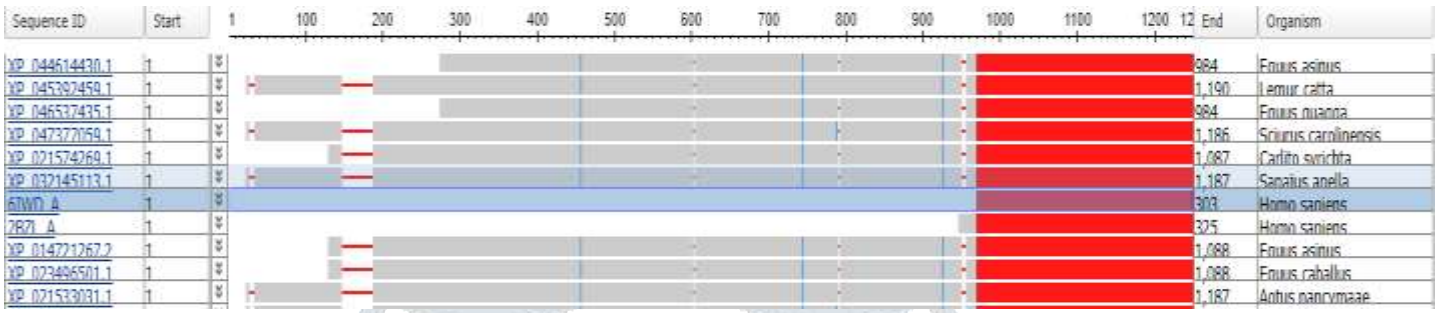

Figure 6: Panorama view result in MSA

Multiple sequence alignment (MSA) results obtained from COBALT, highlighting the sequence conservation and domain architecture of PTPN14 homologs across various vertebrate species. The graphical panorama representation demonstrates the relative sequence alignment and coverage against the query structure 6IWD_A (Homo sapiens). Each row in the figure represents a homologous sequence from a different organism, with start and end positions labeled and aligned to a common scale. Notably, XP_044144430.1 (Equus asinus), XP_045392459.1 (Lemur catta), and XP_046537445.1 (Equus quagga) show nearly full-length alignment with the query, extending beyond 984 residues, while Carlito syrichta, Saimiri sciureus, and Sapajus apella extend up to ~1187 residues, indicating larger domain coverage and suggesting evolutionary conservation of the PTPN14 protein among primates and mammals. The aligned blocks, shaded in grey and blue, indicate regions of strong conservation, particularly within the central and C-terminal portions of the protein. These conserved regions likely correspond to functionally important domains, such as the phosphatase catalytic site and protein interaction motifs. The red blocks at the C-terminal end represent highly conserved sequences, possibly indicating structurally or catalytically crucial regions.

The reference entry 6IWD_A aligns up to residue 303, representing a truncated or domain-specific structure used in crystallographic analysis. In contrast, sequences like 2B7J_A, another human isoform, span 325 residues, suggesting structural overlap. This alignment demonstrates high interspecies conservation of the PTPN14 protein, reinforcing its essential biological function. The presence of consistent alignment across distant species highlights the evolutionary pressure to preserve key domains, making PTPN14 a promising and conserved target for cross-species drug development strategies.

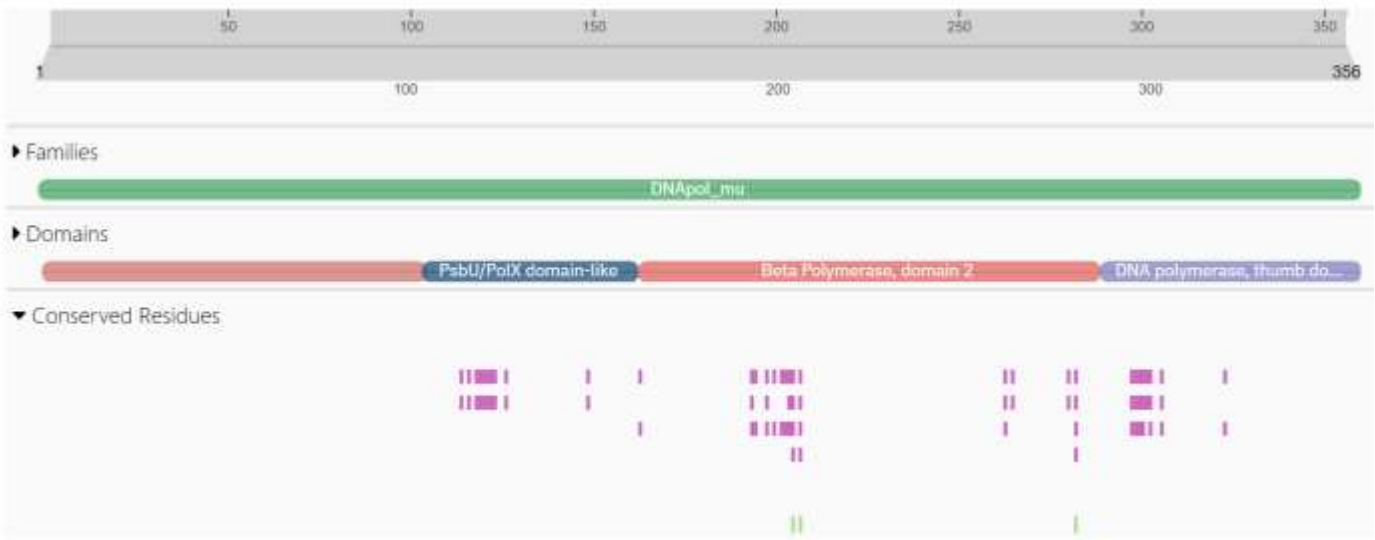

Figure 7: Protein Domain and Motif Analysis of Proteomic Sample

The visualization includes predictions for protein families, functional domains, and conserved residues, providing a comprehensive insight into the structural and functional characteristics of the sequence. The family classification indicates a match with the DNApol_mu (DNA polymerase mu) family, as shown by the continuous green bar spanning the entire length of the sequence from residue 1 to 356. This suggests that the protein likely performs a DNA polymerization function, with roles in DNA repair or replication. The domain architecture reveals the presence of three distinct functional domains:

- PsbU/PolX domain-like (blue) located in the N-terminal region, associated with polymerase accessory activity.
- Beta Polymerase, domain 2 (red) spans the central portion of the protein, likely corresponding to the catalytic core involved in nucleotide incorporation.
- DNA Polymerase, thumb domain (lavender) at the C-terminal end, which typically plays a role in template binding and stability during polymerization.

The conserved residue track below the domain map (marked in magenta) highlights multiple conserved amino acid positions dispersed across the sequence. These conserved sites likely contribute to enzymatic activity, substrate binding, or structural integrity of the polymerase domains. Their presence supports the functional importance of the detected domains and validates the protein's predicted role.

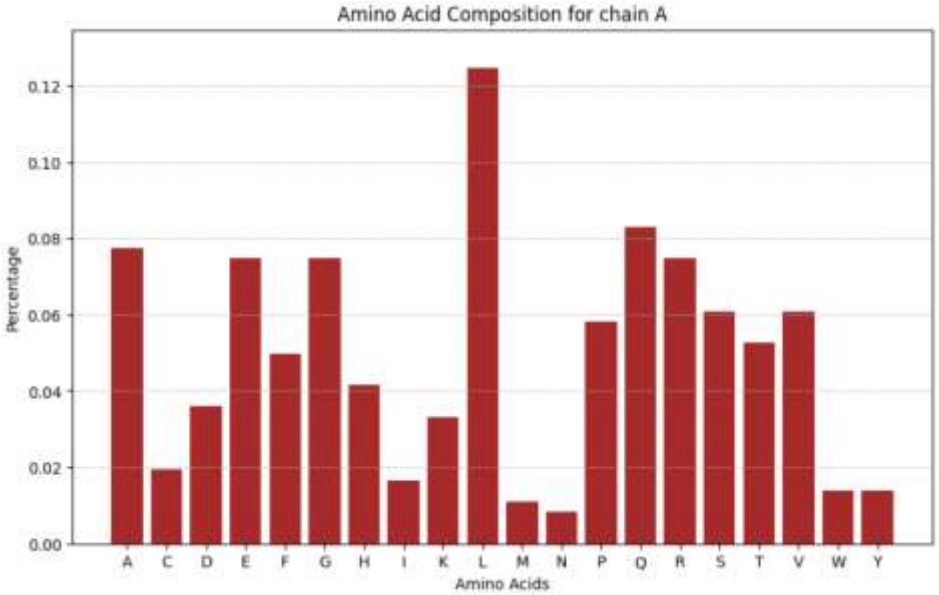

Figure 8: Amino Acid composition graph obtained through BioPython

The bar graph illustrates the amino acid composition of chain A, expressed as a percentage of each residue type. Among all the amino acids, leucine (L) displays the highest abundance, making up over 12% of the chain's composition. In contrast, methionine (M), asparagine (N), and cysteine (C) are among the least represented, each constituting less than 3% of the total.

Moderately abundant residues include glutamine (Q), glutamic acid (E), glycine (G), and alanine (A), all of which contribute roughly 7–8% to the overall sequence. Amino acids such as lysine (K), serine (S), threonine (T), valine (V), and phenylalanine (F) exhibit average representation, generally ranging from 5% to 6%. The distribution reflects a noticeable preference for hydrophobic and polar uncharged residues, which may be related to the structural or functional roles of this chain. Overall, the composition suggests a diverse but uneven distribution of amino acids in chain A, with certain residues being significantly more prominent than others.

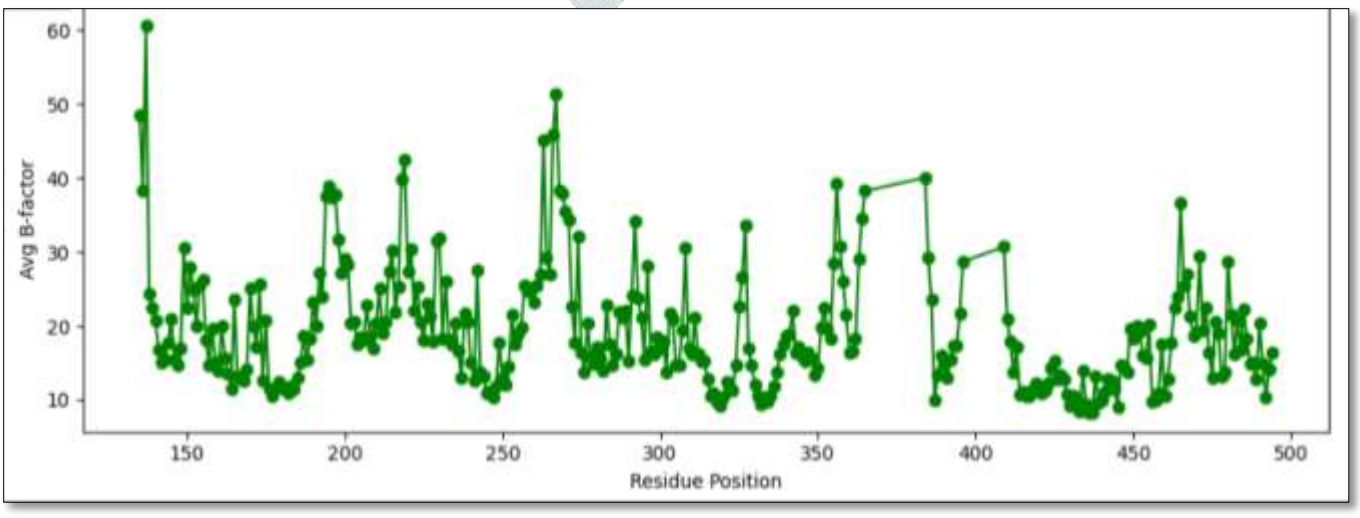

Figure 9: Average B-factor per residue obtained through BioPython

The plot presents the average B-factor values for each residue along the protein chain, highlighting variations in atomic flexibility. B-factors, also known as temperature factors, reflect the mobility or disorder of atoms within the structure. In this graph, several peaks are observed, indicating regions with higher flexibility. The most prominent peak appears around residue position 130, where the B-factor exceeds 60, suggesting significant atomic movement or structural disorder in this region. Additional fluctuations are visible near residues 250 and 370, also representing segments of increased mobility. Conversely, multiple regions particularly between residues 170–190, 320–340, and 440–460 exhibit consistently lower B-factor values, indicating more stable or rigid structural elements. The graph reveals a heterogeneous distribution of atomic flexibility across the protein, which may correspond to functionally dynamic domains or surface-exposed loops, while the lower B-factor regions likely represent the protein’s stable core or well-ordered secondary structures.

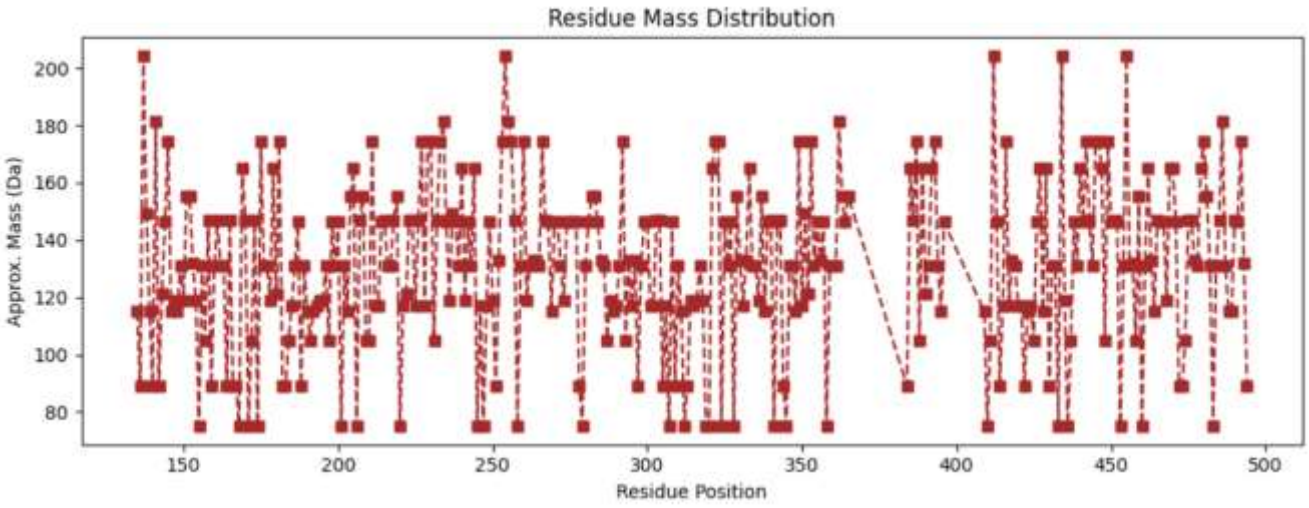

Figure 10: Residue Mass Distribution Across the Protein Chain

Each data point corresponds to a residue, plotted by its sequence position and estimated molecular mass.The mass values fluctuate considerably, generally ranging between 75 Da and just over 200 Da. This variation reflects the inherent differences in molecular weights of individual amino acids, such as glycine and alanine on the lower end, and tryptophan or tyrosine on the higher end. No uniform pattern is observed across the chain; rather, the distribution is highly irregular, with spikes and dips appearing at various intervals. Some regions around positions 140, 250, 370, and 450 exhibit frequent shifts between high and low mass residues, suggesting a diverse amino acid composition in those segments.

4. Molecular Docking

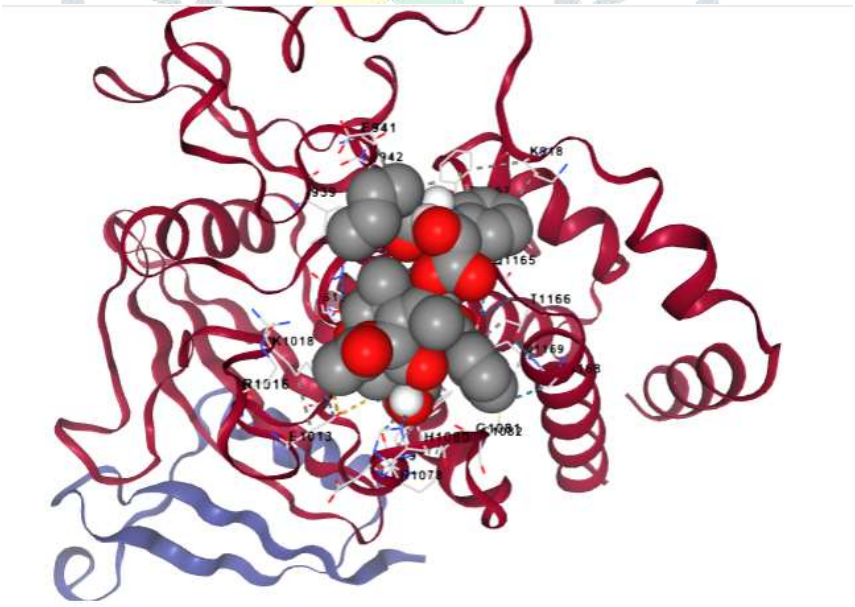

Figure 11: Docking with Taxol C

Table 2: Docking score with Taxol C

CurPocket ID	Vina Score	Cavity Volume (Å³)	Center (x, y, z)	Docking Size (x, y, z)
C4	-6.9	281	-19, 34, -23	26, 26, 26
C3	-6.4	330	-12, 9, -20	26, 26, 26
C2	-5.4	641	-31, 25, 0	26, 26, 26
C1	-5.1	817	-13, 17, -7	26, 26, 26
C5	-5.1	239	10, 31, -12	26, 26, 26

The table summarizes the docking characteristics of five predicted binding pockets. Among them, C4 exhibits the most favorable Vina score of -6.9, indicating the strongest binding affinity. It is followed by C3 with a score of -6.4. Although C1 has the largest cavity volume (817 Å³), its Vina score is lower, suggesting weaker binding potential despite the spacious pocket.

Interestingly, C2 also shows a relatively large cavity (641 Å³) but a moderate binding score (-5.4), whereas C5 has the smallest cavity (239 Å³) and the least favorable score among the group (-5.1, tied with C1). This data suggests that pocket C4, despite having a medium cavity size (281 Å³), offers the best balance between volume and binding affinity, making it a prime candidate for docking experiments.

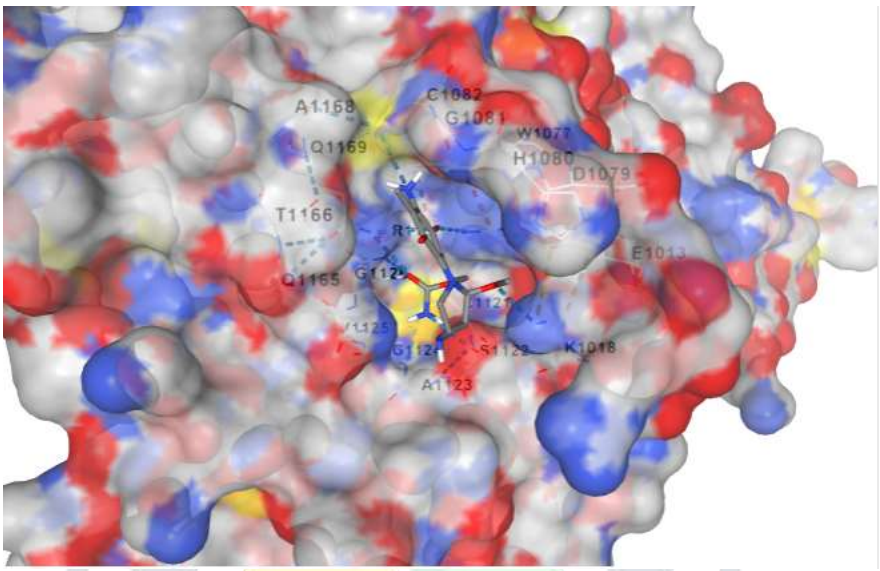

Figure 12: Docking with Mtomycin

Table 3: Docking score with Mtomycin

CurPocket ID	Vina Score	Cavity Volume (Å³)	Center (x, y, z)	Docking Size (x, y, z)
C4	-7.2	281	-19, 34, -23	19, 19, 19
C3	-6.7	330	-12, 9, -20	19, 19, 28
C2	-6.3	641	-31, 25, 0	19, 19, 19
C1	-6.2	817	-13, 17, -7	27, 19, 19
C5	-5.2	239	10, 31, -12	19, 19, 19

The data highlights comparative insights into the five docking pockets evaluated based on Vina binding scores, cavity volumes, and grid box sizes. C4 exhibits the most favorable Vina score of -7.2, indicating the strongest binding potential, despite its moderate cavity volume (281 Å³) and smaller docking grid size.

C3 and C2 also show strong binding scores (-6.7 and -6.3, respectively), but with slightly larger cavity volumes. Notably, C1 possesses the largest cavity (817 Å³) but ranks lower in binding affinity (-6.2), while C5, with the smallest cavity (239 Å³), shows the weakest interaction (-5.2). C4 remains the most promising candidate due to its high binding affinity, making it a suitable target for molecular docking and ligand optimization studies, even though it doesn't have the largest volume or docking grid.

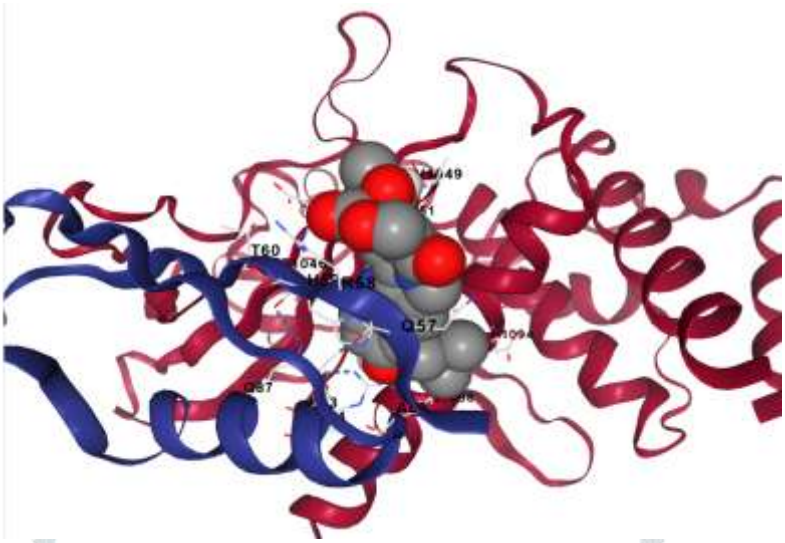

Figure 13: Docking with Topotecan

Table 4: Docking score with Topotecan

CurPocket ID	Vina Score	Cavity Volume (Å³)	Center (x, y, z)	Docking Size (x, y, z)
C3	-7.3	330	-12, 9, -20	24, 24, 24
C1	-6.5	817	-13, 17, -7	24, 24, 24
C4	-6.5	281	-19, 34, -23	24, 24, 24
C2	-6.4	641	-31, 25, 0	24, 24, 24
C5	-5.8	239	10, 31, -12	24, 24, 24

Among the evaluated docking pockets, C1 demonstrated the strongest binding affinity with a vina score of -7.8, paired with the largest cavity volume (817 Å³), suggesting its potential as a highly accessible and energetically favorable binding site.

Closely following is C3, which recorded a slightly lower Vina score of -7.7, despite having a smaller cavity size. C4, C2, and C5 showed moderate to low binding affinities, with C5 having both the weakest Vina score (-5.8) and the smallest cavity volume. All pockets were docked using a uniform grid box size of 23 × 23 × 23 Å³, and interactive contact residue information was available for detailed analysis. The results position C1 as the most promising candidate for further molecular interaction studies, given its optimal balance of volume and binding strength.

CONCLUSION

This study successfully employed computational techniques to analyze the structural and functional characteristics of a target protein, providing valuable insights into its amino acid composition, structural flexibility, residue mass distribution, and potential binding sites. The amino acid profile revealed a predominance of leucine, suggesting its potential role in maintaining structural stability. The B-factor analysis highlighted flexible regions across the protein, indicating dynamic segments that may be involved in functional conformational changes. The residue mass distribution confirmed a diverse composition of residues, with no extreme outliers, ensuring overall molecular balance.Through cavity detection and molecular docking, five prominent binding pockets were identified and evaluated. Among these, Pocket C1 emerged as the most favorable site, exhibiting the best binding affinity (Vina score: -7.8) and the largest cavity volume (817 Å³), indicating its high accessibility and likelihood to accommodate ligands effectively. These computational findings form a strong foundation for structure-based drug design, enabling the selection of optimal binding sites for future inhibitor development. Further validation through experimental studies and ligand screening is recommended to confirm the binding potential of the identified pockets, particularly C1, in therapeutic applications.

REFERENCES

1. Kusakabe M, Taguchi A, Sone K, Mori M, Osuga Y. Carcinogenesis and management of human papillomavirus-associated cervical cancer. *Int J Clin Oncol*. 2023;28(8):965-974. doi:10.1007/s10147-023-02337-7
2. Rahangdale L, Mungo C, O'Connor S, Chibwesha CJ, Brewer NT. Human papillomavirus vaccination and cervical cancer risk [published correction appears in BMJ. 2023 Dec 7;383:p2901. doi: 10.1136/bmj.p2901.]. *BMJ*. 2022;379:e070115. Published 2022 Dec 15. doi:10.1136/bmj-2022-070115
3. Oyouni AAA. Human papillomavirus in cancer: Infection, disease transmission, and progress in vaccines. *J Infect Public Health*. 2023;16(4):626-631. doi:10.1016/j.jiph.2023.02.014
4. Williamson AL. Recent Developments in Human Papillomavirus (HPV) Vaccinology. *Viruses*. 2023;15(7):1440. Published 2023 Jun 26. doi:10.3390/v15071440
5. Nelson CW, Mirabello L. Human papillomavirus genomics: Understanding carcinogenicity. *Tumour Virus Res*. 2023;15:200258. doi:10.1016/j.tvr.2023.200258
6. Crosbie EJ, Einstein MH, Franceschi S, Kitchener HC. Human papillomavirus and cervical cancer. *Lancet*. 2013;382(9895):889-899. doi:10.1016/S0140-6736(13)60022-7

7. Wei F, Georges D, Man I, Baussano I, Clifford GM. Causal attribution of human papillomavirus genotypes to invasive cervical cancer worldwide: a systematic analysis of the global literature. *Lancet*. 2024;404(10451):435-444. doi:10.1016/S0140-6736(24)01097-3
8. Sharifian K, Shoja Z, Jalilvand S. The interplay between human papillomavirus and vaginal microbiota in cervical cancer development. *Virol J*. 2023;20(1):73. Published 2023 Apr 19. doi:10.1186/s12985-023-02037-8
9. Zhang Z, Ma Q, Zhang L, et al. Human papillomavirus and cervical cancer in the microbial world: exploring the vaginal microecology. *Front Cell Infect Microbiol*. 2024;14:1325500. Published 2024 Jan 25. doi:10.3389/fcimb.2024.1325500
10. Volesky-Avellaneda KD, Laurie C, Tsyruk-Romano O, El-Zein M, Franco EL. Human Papillomavirus Detectability and Cervical Cancer Prognosis: A Systematic Review and Meta-analysis. *Obstet Gynecol*. 2023;142(5):1055-1067. doi:10.1097/AOG.0000000000005370
11. Winer RL, Lin J, Anderson ML, et al. Strategies to Increase Cervical Cancer Screening With Mailed Human Papillomavirus Self-Sampling Kits: A Randomized Clinical Trial. *JAMA*. 2023;330(20):1971-1981. doi:10.1001/jama.2023.21471
12. Sravani AB, Ghate V, Lewis S. Human papillomavirus infection, cervical cancer and the less explored role of trace elements. *Biol Trace Elem Res*. 2023;201(3):1026-1050. doi:10.1007/s12011-022-03226-2
13. Andújar-Sánchez M. Role of human papillomavirus genotyping in cervical cancer screening. *Enferm Infecc Microbiol Clin (Engl Ed)*. 2023;41(5):259-261. doi:10.1016/j.eimce.2022.11.022
14. Piña-Sánchez P. Human Papillomavirus: Challenges and Opportunities for the Control of Cervical Cancer. *Arch Med Res*. 2022;53(8):753-769. doi:10.1016/j.arcmed.2022.11.009
15. Shao N. Research progress on human papillomavirus-negative cervical cancer: A review. *Medicine (Baltimore)*. 2024;103(41):e39957. doi:10.1097/MD.0000000000039957
16. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000;28(1):235-242. doi:10.1093/nar/28.1.235
17. Kim S. Getting the most out of PubChem for virtual screening. *Expert Opin Drug Discov*. 2016;11(9):843-855. doi:10.1080/17460441.2016.1216967
18. Mooers BHM. Shortcuts for faster image creation in PyMOL. *Protein Sci*. 2020;29(1):268-276. doi:10.1002/pro.3781
19. Sayle RA, Milner-White EJ. RASMOL: biomolecular graphics for all. *Trends Biochem Sci*. 1995;20(9):374. doi:10.1016/s0968-0004(00)89080-5
20. Mottarella SE, Rosa M, Bangura A, Bernstein HJ, Craig PA. Conscript: RasMol to PyMOL script converter. *Biochem Mol Biol Educ*. 2010;38(6):419-422. doi:10.1002/bmb.20450
21. Verma R, Pradhan D, Nayek A, Singh H, Jain AK, Khan LA. Target-based drug repurposing against Candida albicans-A computational modeling, docking, and molecular dynamic simulations study. *J Cell Biochem*. 2022;123(2):289-305. doi:10.1002/jcb.30163
22. Sneha S, Pandey DM. In silico structural and functional characterization of Antheraea mylitta cocoonase. *J Genet Eng Biotechnol*. 2022;20(1):102. Published 2022 Jul 11. doi:10.1186/s43141-022-00367-8
23. Kukreja, Vinita & Kumari, Uma. (2023). Data Analysis of Brain Cancer with Biopython. 8. 2146-2154
24. Quevillon E, Silventoinen V, Pillai S, et al. InterProScan: protein domains identifier. *Nucleic Acids Res*. 2005;33(Web Server issue):W116-W120. doi:10.1093/nar/gki442
25. Papadopoulos JS, Agarwala R. COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics*. 2007;23(9):1073-1079. doi:10.1093/bioinformatics/btm076
26. Liu Y, Yang X, Gan J, Chen S, Xiao ZX, Cao Y. CB-Dock2: improved protein-ligand blind docking by integrating cavity detection, docking and homologous template fitting. *Nucleic Acids Res*. 2022;50(W1):W159-W164. doi:10.1093/nar/gkac394
27. Uma kumari, Devanshi Gupta, Computational analysis of Glioma Transcriptome and proteome with Bio python ;2023 Volume-13, Issue-1, Pp 1-14,IJBTR
28. 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.
29. Uma kumari,Meenakshi Pradhan,Saptarshi Mukherjee,Sreyashi Chakrabarti;,Ngs Analysis approach for neurodegenerative disease with Biopython",2023 ,JETIR,volume 10,issue 9 , http://doi.one/10.1729/Journal.36043
30. 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
31. 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
32. Kumari Uma & Agrawal, Nidhi. (2023). NGS and Mutational Profile Analysis of Non-Small-Cell Lung Carcinoma (NSCLC). International Journal for Research in Applied Science and Engineering Technology. 11. 3090-3094. 10.22214/ijraset.2023.50880.
33. Kumari, Uma & Gupta, Shruti. (2023). NGS and Sequence Analysis with Biopython for Prospective Brain Cancer Therapeutic Studies. International Journal for Research in Applied Science and Engineering Technology. 11. 10.22214/ijraset.2023.50885.
34. Uma Kumari, Vineeta Johri, Swarali Dhopate, Tijil Jha Structure Based Drug Designing for the prediction of epitope for targetin Malignant brain Tumor, 2024 JETIR July 2024, Volume 11, Issue 7 www.jetir.org (ISSN-2349-5162)
35. Tanmay Bandbe , Juri Saikia , Uma Kumari. (2025). NGS Analysis Human Papillomavirus Type 18 E2 DNA-Binding Domain Bound to its DNA Target with Biopython. *South Eastern European Journal of Public Health*, 3781–3792. https://doi.org/10.70135/seejph.vi.5856,SEEJPH Volume XXVI, S2
36. Uma Kumari,Taruni Bajaj "Ngs Data Analysis And Active Site Identification Alpha-1-Acid Glycoprotein Bound To Potent Anti Tumor Compound UCN-01 In Malignant Brain Tumor", International Journal of Emerging Technologies and Innovative Research ,Vol.12, Issue 3, page no.g107-g116, http://doi.one/10.1729/Journal.44259
37. Uma Kumari,Renu et al "Structure Analysis and molecular Docking of Mesothelin-207 fragment in human cancer", Jetir,Vol.12, Issue 3, page no.g98-g106, http://doi.one/10.1729/Journal.44261
38. Uma Kumari,Gaurav Verma ,Murugesan AnbuMegala "IN SILICO DRUG DESIGN IN HUMAN APOPTOSIS INDUCING FACTOR (AIF) IN LUNG CANCER", International Journal of Emerging Technologies and Innovative ISSN:2349-5162, Vol.12, Issue 2, page no. ppd139-d148, http://doi.one/10.1729/Journal.43606