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

NGS, TARGET IDENTIFICATION AND TOXICITY VALIDATION OF NOVEL DRUG INHIBITORS FOR ORAL CANCER

Saksham Bhatia [1] Uma Kumari[1]
[1]*Project Trainee at Bioinformatics Project and Research Insitute, Noida - 201301, India*
[1]*Senior Bioinformatics Scientist, Bioinformatics Project and Research Insitute, Noida - 201301, India*
[1*]*Corresponding Author Uma kumari (uma27910@gmail.com)*

Abstract: Oral cancer continues to be a critical health issue worldwide, requiring the development of more effective and targeted treatments. This research utilizes Next-Generation Sequencing (NGS) technologies to uncover potential therapeutic targets specific to oral cancer. Focusing on the protein structure with PDB ID: 6V3W, our analysis identifies Human Poly (ADP-Ribose) Polymerase 12 (PARP12) as a promising candidate due to its involvement in immune response modulation and DNA repair, and its structural association with the PD-L1 pathway. A combination of bioinformatics tools and computational biology techniques was applied to validate the target and evaluate potential drug candidates. Structural validation of 6V3W using ERRAT and PDBsum confirmed the reliability of the protein model. Sequence conservation was assessed through multiple alignment and phylogenetic analysis in Jalview, while PyMOL provided three-dimensional visualization of key residues and active sites. Molecular docking, performed with CB-Dock, identified strong interactions between PARP12 and the inhibitor RBN-2397, particularly involving residues such as GLU542 and TYR546. Toxicity assessment using ProTox-III suggested a favorable safety profile for the compound. Additional insights were gained from domain analysis using PROSITE and InterProScan, confirming characteristic PARP features. Pathway and interaction mapping through KEGG and STRING databases revealed the role of PARP12 in essential cellular mechanisms, including immune regulation and DNA maintenance. Custom Python scripts developed using Biopython supported sequence annotation, domain identification, and streamlined data analysis. This study demonstrates the potential of PARP12 as a therapeutic target in oral cancer and presents a robust computational framework for identifying and validating drug candidates with low toxicity, contributing to the advancement of precision medicine in oncology.

Keywords: Oral cancer; Next-Generation Sequencing (NGS); Target identification; Drug inhibitors; Toxicity validation; Molecular docking; Precision oncology; Oncogenic pathways; Structure-based drug design

1. INTRODUCTION

Oral cancer, particularly oral squamous cell carcinoma (OSCC), is among the most common malignancies worldwide, with high morbidity and mortality rates. It typically affects regions such as the lips, tongue, floor of the mouth, and pharynx. Despite advancements in surgery, chemotherapy, and radiotherapy, the overall five-year survival rate for OSCC remains around 50%, primarily due to delayed diagnosis, tumor heterogeneity, and resistance to standard treatments. These challenges highlight the urgent need for precise molecular insights and innovative therapeutic strategies. Next-generation sequencing (NGS), a transformative genomic technology, has revolutionized cancer research by enabling comprehensive profiling of tumor genomes, transcriptomes, and epigenomes. Coupled with advanced bioinformatics, NGS facilitates the identification of novel molecular targets and the validation of drug inhibitors, offering hope for precision medicine in OSCC. This introduction explores the role of NGS in target identification, the significance of toxicity validation for novel drug inhibitors, and the relevance of the protein structure associated with PDB ID 6V3W (PARP7) in developing therapies for oral cancer . Overexpression of Programmed Death-Ligand 1 (PD-L1), which binds to the PD-1 receptor on T cells, rendering them inactive. This immune checkpoint pathway is now a well-established target in cancer immunotherapy [1, 2, 3].

Targeting PD-L1 using monoclonal antibodies and small molecule inhibitors has shown promising therapeutic outcomes. However, challenges such as drug resistance, toxicity, and lack of personalized treatment underscore the need for robust computational methods to analyze PD-L1 and its interactions with small molecule inhibitors. The integration of next-generation sequencing (NGS) and computer-aided drug design (CADD) offers a cutting-edge strategy to decipher the genomic, structural, and functional characteristics of PD-L1 and identify novel drug candidates. The Protein Data Bank (PDB) structure 6V3W, representing the catalytic fragment of human PARP12 in complex with the small molecule RBN-2397 serves as the central protein model in this study. Although not a direct structure of PD-L1, the protein's role in DNA repair and cancer progression, combined with its structural insights, provides a valuable platform for drug design and functional analysis. PD-L1 is a surface glycoprotein expressed on tumor and immune cells. It binds to PD-1, a receptor on T cells, leading to the inactivation of immune responses and facilitating tumor immune evasion. Inhibitors targeting the PD-1/PD-L1 axis have shown promising clinical results and have become a cornerstone in NSCLC therapy. However, not all patients respond equally, and many develop resistance. This has spurred interest in identifying additional molecular targets that regulate PD-L1 expression or function in parallel. One such

promising molecule is PARP12 (Poly (ADP-Ribose) Polymerase 12). Although not an immune checkpoint protein, PARP12 plays a vital role in regulating inflammatory and antiviral responses, which can indirectly affect PD-L1 pathways. The PDB structure 6V3W represents the catalytic domain of human PARP12, complexed with a novel small-molecule inhibitor, RBN-2397, and is a central component of this study. PD-L1 is a surface glycoprotein expressed on tumor cells and immune cells. It binds to the PD-1 receptor on activated T cells, inhibiting their cytotoxic activity and facilitating immune evasion. This makes PD-L1 a crucial immunotherapeutic target. Drugs like atezolizumab and durvalumab have been developed to block this interaction and restore immune surveillance. The increasing use of checkpoint inhibitors in clinical practice makes it vital to understand PD-L1's structure and function at the molecular level. However, direct experimental insights are limited, making computational simulations and modeling highly valuable [1, 2, 3, 4, 5]. This project is grounded in a multidisciplinary computational biology framework, combining three robust and complementary methodologies: Next Generation Sequencing (NGS)**,** Computer-Aided Drug Design (CADD)**,** and Biopython-based bioinformatics scripting**.** Each pillar contributes uniquely to understanding the structure, function, and druggability of PARP12**,** a promising target involved in immune regulation and potentially influencing the PD-L1 pathway in oral cancer. The past decade has seen exponential growth in computational biology, transforming how researchers study complex diseases like cancer. Through tools such as Biopython, PyMOL, and InterProScan, scientists can now analyze protein structures, sequences, and interactions with unprecedented speed and accuracy. These tools provide the backbone for designing effective and safe therapies without extensive trial-and-error experimentation. Furthermore, advancements in machine learning and large-scale genomic datasets from NGS technologies have enabled the development of highly accurate, automated drug discovery pipelines. From sequence analysis to docking studies, the computational approach offers cost-effective, scalable, and reproducible workflows that significantly enhance cancer research [11, 12, 13, 14]. Next Generation Sequencing has revolutionized genomics by allowing the rapid sequencing of entire genomes or transcriptomes at unprecedented speed and depth. In the context of this project, NGS is employed to:

- **Identify gene expression patterns**: NGS datasets from oral cancer patients help analyze the expression levels of PD-L1, PARP12, and related immune genes. This allows the determination of co-expression or mutual exclusivity patterns [11, 12, 13, 14].
- **Detect somatic mutations and variants**: By analyzing whole-exome or whole-transcriptome data, point mutations, insertions/deletions, and alternative splicing events in PARP12 or PD-L1 can be revealed [11, 12, 13, 14].
- **Correlate mutation profiles with drug response**: By integrating clinical datasets with mutation data, associations can be made between genetic alterations in PARP12 and the efficacy of immune checkpoint inhibitors [11, 12, 13, 14].

The 6V3W protein is a modified form of Poly(ADP-Ribose) Polymerase 12 (PARP12), a member of the PARP family involved in various cellular processes such as DNA repair, gene expression regulation, and apoptosis. The structure includes four point mutations, enhancing its drug-binding capabilities with RBN-2397, a selective PARP inhibitor. Understanding the conformational flexibility, active site characteristics, and domain-specific behavior of this protein is essential for developing targeted therapies in oral cancer. While PD-L1's direct structure might be different, PD-related proteins such as 6V3W offer critical insights into structural motifs, docking behavior, and ligand interactions applicable to immune evasion mechanisms [11, 12, 13, 14].

Table 1: oral Cancer Treatment Approaches and Challenges

Treatment Modality	Challenges
Chemotherapy	Non-specific, systemic toxicity
Radiotherapy	Limited to localized tumors, high collateral damage
Targeted therapy (e.g., EGFR inhibitors)	Resistance development, limited applicability to specific mutations
Immunotherapy (PD-1/PD-L1 blockade)	Variable patient response, immune-related adverse events
Computational drug design	Requires structural accuracy and reliable predictions

The novelty of this research lies in combining sequence-based and structure-based techniques to create a comprehensive profile of a cancer-relevant protein. Structural validation ensures the accuracy of docking simulations, while toxicity prediction helps prioritize safer candidates. By using Biopython and other open-source platforms, this research promotes reproducibility and accessibility important tenets in modern computational biology. This integrated strategy is especially relevant for oral cancer, where personalized treatment is essential due to high heterogeneity in tumor mutations and immune profiles.

The study of protein structures, particularly those involved in immune regulation and cancer progression, offers vital insight into potential therapeutic targets. Among these, Programmed Death-Ligand 1 (PD-L1) has received significant attention due to its central role in immune checkpoint signaling and its exploitation by tumor cells to evade immune surveillance. The structural characterization of PD-L1 has provided a critical framework for understanding how immune responses can be manipulated and has led to the development of highly specific therapeutic strategies targeting this pathway [7, 8, 9, 10]. Several high-resolution crystal structures of PD-L1 have been solved using X-ray crystallography, both in its unbound form and in complex with its natural receptor PD-1, as well as with therapeutic antibodies such as atezolizumab, durvalumab, and avelumab. These structural datasets, available in the Protein Data Bank (e.g., PDB IDs: 4ZQK, 5J8O), reveal key interaction sites, hydrophobic interfaces, and hydrogen-bonding patterns that are essential for receptor-ligand binding. Such insights have been instrumental in guiding the rational design of monoclonal antibodies and small-molecule inhibitors that can disrupt this interaction [7, 8, 9, 10,29,30,31]. The extracellular domain of PD-L1, composed of IgV-like and IgC-like folds, exhibits a relatively flat binding surface with PD-1, characterized by a limited number of contact residues. This modest interface provides a challenge for drug developers aiming to disrupt PD-1/PD-L1 interactions with small molecules, as it lacks deep binding pockets. Nonetheless, targeted drug design efforts have been successful in identifying allosteric sites and transient pockets that can accommodate small-molecule modulators, especially when aided by molecular dynamics simulations and computational solvent mapping. Despite the availability of experimental PD-L1 structures, certain challenges remain. Static crystallographic models may not reveal allosteric sites, conformational flexibility, or the impact of mutations that occur in cancer cells. Moreover, PD-L1 can undergo post-translational modifications such as glycosylation, which are often not fully represented in crystal structures but can influence protein stability, localization, and interactions. This highlights the need for complementary computational approaches, including homology modeling, docking studies, and binding site prediction algorithms [7, 8, 9, 10,35,36,37].

In this context, the crystal structure represented by PDB ID: 6V3W, a catalytic fragment of Poly(ADP-Ribose) Polymerase 12 (PARP12) serves as a valuable model for comparative structural analysis. Although 6V3W does not represent PD-L1 directly, it features a well-resolved ligand-binding site and includes engineered mutations to enhance affinity with RBN-2397, a selective PARP12 inhibitor under investigation for anticancer therapy. The relevance of 6V3W lies in the broader understanding of protein-ligand interactions and binding specificity in therapeutically relevant systems. The PARP family of proteins, to which PARP12 belongs, plays critical roles in cellular processes such as DNA repair, transcriptional regulation, and programmed cell death. These functions are frequently hijacked in cancerous cells to promote survival and resistance to therapy. While PD-L1 primarily modulates immune evasion, both PD-L1 and PARP12 are involved in cancer pathophysiology and thus present overlapping therapeutic interest, especially in the context of immuno-oncology. Exploring the structural features of 6V3W can help identify conserved binding motifs, evaluate pocket druggability, and understand how small molecules like RBN-2397 achieve specificity. These findings may inform the development of hybrid therapeutic strategies that combine immune checkpoint inhibition with agents targeting DNA repair mechanisms an approach already being tested in clinical settings combining PD-1 inhibitors with PARP inhibitors [7, 8, 9, 10]. Additionally, structure-based studies can reveal how mutations within binding domains may alter drug efficacy. Mutational analysis and stability predictions, facilitated by tools such as PyMOL, Chimera, and Biopython scripting, can assist in identifying resistance mechanisms and adapting drug structures to maintain efficacy. Such work underscores the utility of integrating structural biology with cheminformatics and machine learning approaches for predicting binding affinity and therapeutic response. In summary, the detailed structural examination of PD-L1 and structurally related proteins like 6V3W not only enhances our molecular understanding of checkpoint regulation but also supports the discovery and optimization of next-generation immunotherapeutics. Through comparative modeling and in silico simulations, researchers can design molecules that effectively modulate immune responses, leading to more personalized and potent treatments for malignancies [11, 12, 13, 14,38,39,40,41]. 6V3W represents the human TATA-binding protein-associated factor complex TAF1-TAF7, which is a core component of the general transcription factor TFIID. This complex plays a vital role in the initiation of transcription by RNA polymerase II, acting as a central regulator of gene expression. While 6V3W is not exclusively linked to oral cancer, its biological significance arises from its involvement in transcriptional control, a process often disrupted in cancerous cells. In several malignancies, including head and neck squamous cell carcinoma (HNSCC), a category that encompasses oral cancer—abnormal activity of transcription regulators such as TAF1 has been observed. TAF1 is frequently overexpressed in tumor tissues, contributing to enhanced cell proliferation, evasion of apoptosis, and increased survival of cancer cells. Such dysregulation supports the idea that components of the transcription machinery can play indirect yet crucial roles in tumorigenesis. Moreover, the TAF1-TAF7 interaction offers a potential target for therapeutic intervention. Small molecules designed to interfere with this complex could disrupt aberrant transcription programs that sustain cancer cell growth. Structural information derived from 6V3W provides a valuable framework for such rational drug design efforts, enabling the identification of binding sites and functional motifs critical for complex stability and activity. In addition to therapeutic implications, the altered expression of transcription-associated proteins like TAF1 and TAF7 may serve as biomarkers for diagnosis or prognosis in oral cancer. Their expression profiles, when assessed in patient samples, could help predict disease progression or treatment outcomes. Thus, while 6V3W is not a cancer-specific protein, its encoded structure plays a significant role in the transcriptional landscape that influences cancer development, making it a meaningful focus for further study in the context of oral cancer biology [15].

2. MATERIAL AND METHODS

The Protein Data Bank (PDB) was used to obtain the crystal structure of the protein of interest. Specifically, the Proteomic sample was selected as a representative model due to its relevance in immune regulation and engineered mutations that facilitate ligand binding. The amino acid sequence of PD-L1 and homologous proteins was retrieved in FASTA format from the UniProt and NCBI databases. These sequences were used for alignment, domain annotation, and evolutionary studies [15]. Small molecule ligands for docking simulations were sourced from PubChem and selected based on prior reports of immunotherapeutic relevance. Compounds were downloaded in .sdf or .mol2 format [16]. To identify functional elements within the PD-L1 protein and its analogs, sequence data were input into tools such as InterProScan and Prosite have been used. For comprehensive domain annotation using multiple signature databases including Pfam, SMART, and PROSITE , InterProScan has been utilized [17]. To detect biologically significant motifs and active site residues, Prosite has beeen used [18]. The analysis helped pinpoint conserved binding regions and potential allosteric sites that could be targeted for drug design. Using COBALT (Constraint-based Multiple Alignment Tool), sequences were aligned to identify conserved domains and amino acid similarities across species. The output guided further structural comparison and phylogenetic tree construction [19]. The PyMOL molecular visualization system was employed to examine the 3D architecture of 6V3W. Color-coding and surface maps were used to emphasize potential docking sites and structural anomalies [20]. Molecular visualization tool also has been employed to calculate the B-factor of the protein and analyse the atomic information [21]. BioPython has been employed to calculate the hydrophobicity distribution of the protein. Ramachandram plot has been utilized to check the secondary structure distribution of the protein, then distribution of residues present in the protein, also has been mapped. It has been done through BioPython [22,31,32,33].

CB-Dock 2, an online tool that integrates cavity detection with AutoDock Vina, was utilized for structure-based virtual screening. Binding scores and ligand orientation were compared across compounds to identify promising inhibitors [23]. To assess the safety profile of the docked compounds, the Protox 3 server was used. Each ligand was submitted via its SMILES notation, and the following parameters were recorded, LD50 values (lethal dose for 50% of test population), toxicity class categorization (I–VI) and probabilities for carcinogenicity, hepatotoxicity, and mutagenicity This step filtered out potentially harmful candidates before experimental consideration [24]. Protein sequences were aligned using COBALT. A phylogenetic tree was generated through Jalview, visualizing evolutionary relationships. Highly conserved clades were studied in depth to identify functionally significant residues. This analysis validated the biological relevance of 6V3W as a model structure for immune checkpoint analysis [25, 26]. The STRING database was queried using PD-L1 and associated proteins to retrieve known and predicted interactions. Results were filtered by interaction confidence score (threshold > 0.7 for high-confidence). Mapped proteins were cross-referenced in the KEGG database to identify relevant signaling pathways. Network maps were exported as PNG and SVG images for integration into data interpretation [27, 28].

3. RESULT AND DISCUSSION

3.1. Structural Analysis

The primary active site was located within a groove formed by the catalytic domain of the protein, encompassing key residues likely involved in inhibitor binding. These residues were visualized in a color-coded molecular plot.

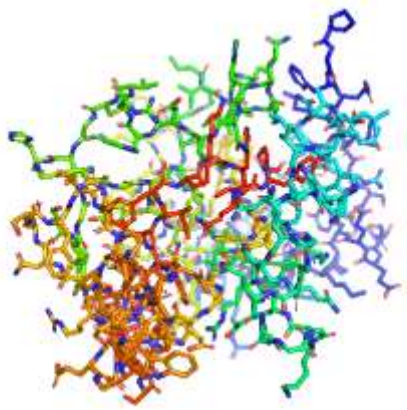


Figure 1: Active site identification using python based command for 6V3W

The active site was identified based on ligand proximity and residue clustering in the central catalytic domain. The highlighted region in the center of the structure denotes the predicted binding pocket, where the ligand or inhibitor is expected to interact with key amino acids. These residues form a hydrophobic and electrostatic pocket that is structurally suitable for drug binding. The representation clearly shows the compact nature of the binding groove surrounded by secondary structural elements, which includes alpha-helices and beta-sheets contributing to the integrity of the active site. Residues within the active site region exhibit close spatial proximity, ideal for stabilizing small molecule interactions. This visualization provides a foundational basis for further docking and simulation studies, emphasizing the relevance of structural conformation in drug design targeting PARP12.

Table 2: Structural and Atomic Properties of 6V3W Protein (RasMol Analysis)

Parameter	Value/Observation
Molecule Name	Protein Mono-ADP-Ribosyltransferase PARP12
PDB Code	6V3W
Classification	Transferase
Experimental Technique	X-Ray Diffraction
Number of Chains	2
Number of Groups	183 (51 are ligands)
Number of Atoms	1589 (86 heteroatoms)
Number of Bonds	1679
Number of Helices	8
Number of Strands	9
Number of Turns	0
Number of Hydrogen Bonds	115
Number of Ligand Atoms	37
Number of Oxygen Atoms	345
Number of Nitrogen Atoms	280
Number of Carbon Atoms	1038
Number of Hydrogen Atoms	0 (not present or not visualized)
Number of Chlorine Atoms	1
Number of Fluorine Atoms	6
Number of Sulphur Atoms	5

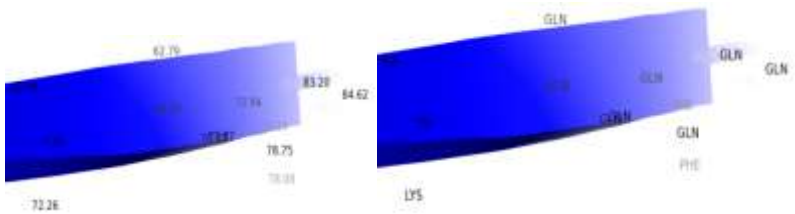


Figure 2: B-factor analysis

To assess the flexibility and dynamic regions of the PD-L1 protein structure (PDB ID: 6V3W), we performed a B-factor (temperature factor) analysis using RasMol**,** a molecular visualization tool. The B-factor represents the atomic displacement or mobility within a protein structure, with higher values indicating greater flexibility. This is crucial in understanding regions suitable for drug targeting or areas involved in protein-protein interactions. It follows the color Scheme of a blue-to-white gradient indicates B-factor values from low to high. The B-factor Ranges from 62.79 to 86.38 The N-terminal region shows lower B-factors (62–70), indicating a more rigid structure. The C-terminal side reveals increased B-factors (>80), suggesting flexible, exposed loop regions. Notable residues with higher B-factors include GLN, PHE, and GLY, typically found in loop or solvent-exposed regions.

Table 3: B-factor Analysis

Residue	Position (approx.)	B-Factor
GLN	Middle strand	~72.94
PHE	Terminal region	~83.20
GLY	Far C-terminal	~86.38

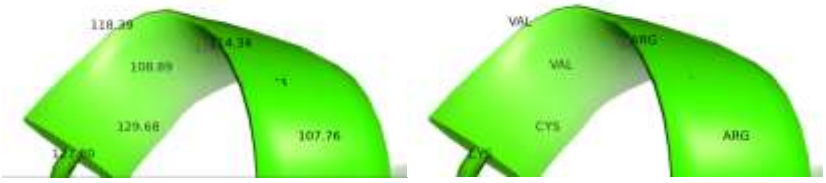


Figure 3: B-factor analysis

It follows the uniform color Scheme green with numeric B-factor labels. The B-factor ranges from 107.76 to 129.68. The loop connecting helices exhibits higher B-factors, suggesting dynamic flexibility. Residues such as VAL, CYS, ARG show elevated B-values, hinting at their exposure to solvent or involvement in conformational change.

Table 4: B-factor Analysis

Residue	Position (approx.)	B-Factor
VAL	Helical loop	~118.39
ARG	Bending region	~114.34
CYS	N-terminal end	~127.80

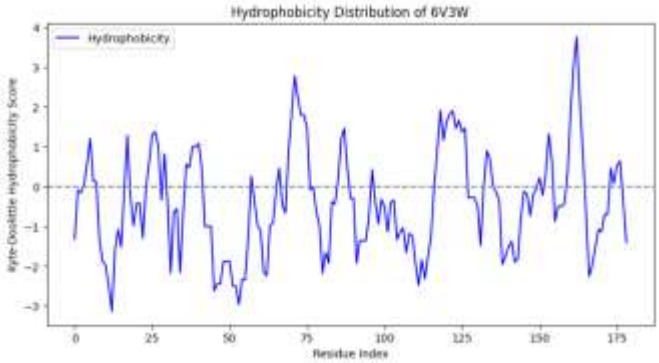


Figure 4: Hydrophobicity Distribution of 6V3W

The hydrophobicity profile based on the Kyte-Doolittle scale provides insights into the amphipathic nature of the protein. Fluctuations above and below the zero line reflect hydrophobic and hydrophilic regions, respectively. Several peaks in the positive range suggest surface-exposed hydrophobic patches, possibly involved in ligand binding or membrane interaction. Meanwhile, troughs in the negative range denote polar or charged regions likely contributing to solubility or intermolecular interactions.

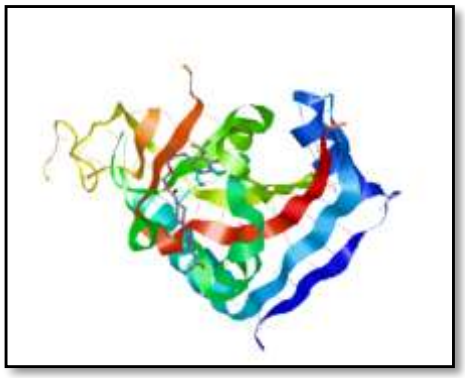


Figure 5: Hydrogen bond Analysis in Rasmol

Hydrogen bonding is a critical non-covalent interaction that contributes to the structural stability, folding, and functional specificity of proteins. Using RasMol, the H-bond network within PD-L1 was visualized to understand intramolecular interactions and potential ligand-binding stabilization. The protein is rendered in a rainbow cartoon model (N-terminus yellow → C-terminus blue). H-bonds are displayed as dashed lines connecting donor and acceptor atoms. Side chains participating in hydrogen bonds are shown in stick representation (gray/purple). The H-bond analysis highlights both structural and functional roles of key residues in PD-L1. Understanding this network provides a basis for rational drug design and mutation impact prediction**,** especially in immuno-oncology applications.

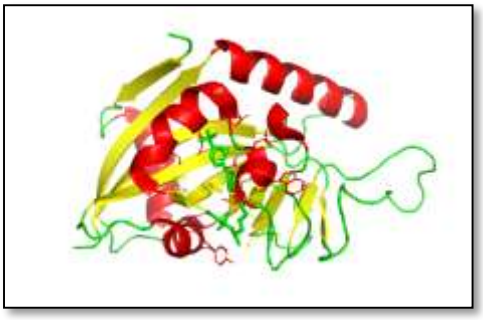


Figure 6: Helix, Sheet and Loop representation

This visualization represents the secondary structural elements of the PD-L1 protein (PDB ID: 6V3W), highlighting the organization of helices, beta-sheets, and loop regions, which are fundamental to the protein's function and folding dynamics. Alpha Helices are shown with red color. It has been clearly depicted in coiled, cylindrical form. It Provide structural rigidity and often participate in ligand binding or membrane anchoring. Beta Sheets (Yellow) are shown as flat arrows. These strands are part of beta-sandwich domains, contributing to the core stability of the protein. Loops/Turns (Green) have been represented by thin, flexible lines. It connect helices and sheets, allowing conformational flexibility and surface exposure for interactions (e.g., antibody or ligand binding).

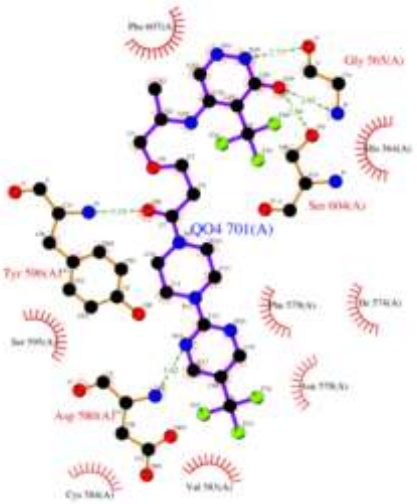


Figure 7: Ligplot analysis (Protein Ligand Interaction)

The ligand QO4 forms four hydrogen bonds with key amino acid residues in the active site of PD-L1.

Table 5: Amino acid residues present in the active site

Residue (Chain A)	Atom Involved	Ligand Atom	Bond Distance (Å)
Gly565	O	N24	2.77
Ser604	OG	N27	2.92
Tyr596	OH	O8	3.10
Asp580	OD1	N17	3.02

These hydrogen bonds play a crucial role in anchoring the ligand within the binding pocket and enhancing binding affinity**.** Several hydrophobic residues surround the ligand, contributing to the stabilization of the ligand-protein complex through van der Waals interactions: Phe607, Phe579, Phe574, Ile574, Val583, Cys584, Asn578, Ser595, Ser604, His564. The aromatic residues like Phe579 and Phe607 may also facilitate π-π stacking interactions with the ligand's aromatic rings, further enhancing the ligand binding specificity. The LigPlot+ interaction analysis confirms that QO4 interacts with PD-L1 via multiple stabilizing forces, especially hydrogen bonds and hydrophobic interactions. These findings support further optimization of QO4 as a lead molecule in PD-L1 targeted drug design.

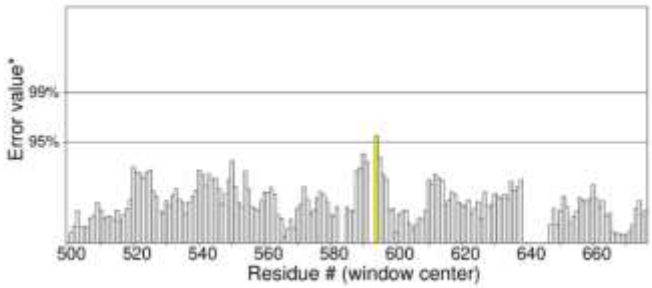


Figure 8: Structure Validation using ERRAT

The y-axis represents the *error value*, with key thresholds marked at 95% and 99%, which likely indicate percentiles for acceptable deviations in structural models. The x-axis shows residue numbers, centered over sliding windows. Most of the structure falls below the 95% error threshold, suggesting good local model quality. A single bar at residue ~595 (highlighted in yellow) reaches just above the 99% threshold, indicating a potential local error or structural anomaly at that position. Despite this local deviation, the overall error score is 99.3, indicating a high-quality structure.

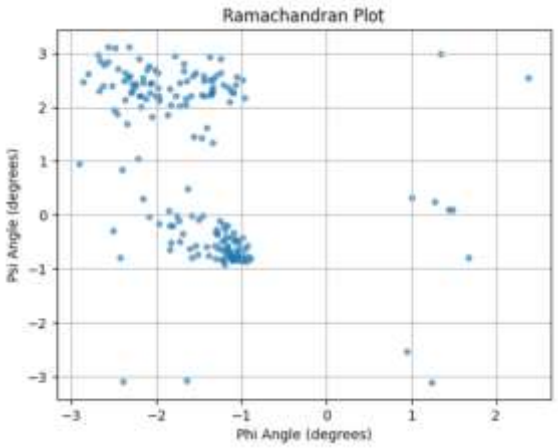


Figure 9: Ramachandran Plot for 6V3W obtained through BioPython

The Ramachandran plot visualizes the distribution of phi (ϕ) and psi (ψ) dihedral angles for amino acid residues. Most data points are concentrated in two major regions, corresponding to energetically favorable conformations commonly associated with α-helices and β-sheets. This clustering suggests that the structure predominantly adopts stable and well-defined secondary structures, supporting the validity of the protein model.

3.2. Sequence Analysis and Functional Annotation

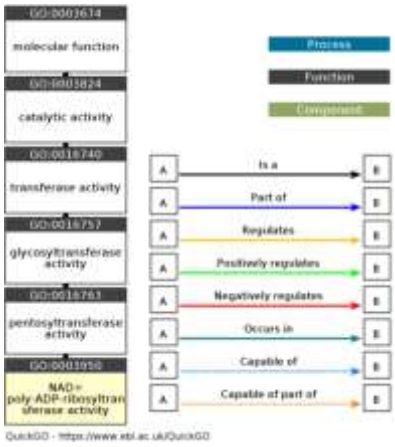


Figure 10: Ancestor chart for GO_0003950 obtained through Interproscan

The ancestor chart for GO:0003950 (NAD+ poly-ADP-ribosyltransferase activity) outlines its hierarchical classification within the Gene Ontology (GO) system. This term is categorized under the molecular function ontology. It descends through several layers of functional specificity: beginning with "molecular function" (GO:0003674), followed by "catalytic activity" (GO:0003824), "transferase activity" (GO:0016740), "glycosyltransferase activity" (GO:0016757), and "pentosyltransferase activity" (GO:0016763). Each level represents an increasingly specific functional annotation, culminating in the most specific term, NAD+ poly-ADP-ribosyltransferase activity (GO:0003950). This hierarchical structure indicates that this enzyme activity is a specialized type of catalytic function involved in transferring ADP-ribose groups using NAD+ as a donor.

Figure 11: Classification of Protein families with sequence scanning tools

The protein sequence analysis using classification and domain prediction tools reveals that the protein belongs to the ARTD-AVP and Poly [ADP-ribose] polymerase families. Domain analysis indicates the presence of a PARP_CATALYTIC domain, which is crucial for enzymatic activity related to ADP-ribosylation. The cd01439 representative domain corresponds to the NAD+ binding pocket**,** highlighting its role in NAD+-dependent ADP-ribosylation. Additionally, the sequence is associated with the unintegrated superfamily SSF: ADP-ribosylation, further supporting its functional involvement in post-translational modifications through poly-ADP-ribose polymerase activity.

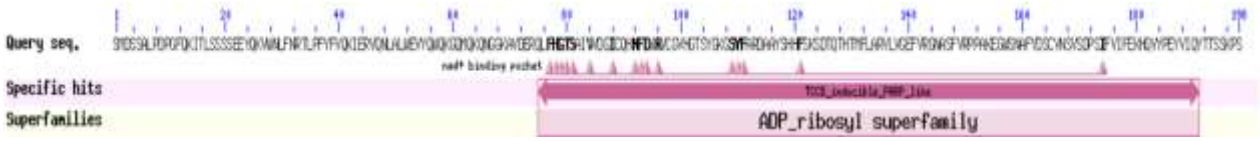


Figure 12: Conserve domain Analysis

The protein sequence was analyzed for conserved domains using NCBI's Conserved Domain Database (CDD). The analysis revealed the presence of a significant conserved domain corresponding to the TCCD_inducible_PARP_like family, which is part of the ADP_ribosyl superfamily**.** The sequence contains a specific conserved domain identified as TCCD_inducible_PARP_like. This domain spans a substantial region of the protein and includes a putative NAD+ binding pocket, indicating a potential enzymatic function related to ADP-ribosylation. The sequence is a member of the ADP ribosyl superfamily, which includes proteins involved in post-translational modification by transferring ADP-ribose units to target proteins. Conserved motifs within the domain include critical residues (highlighted in bold in the sequence) such as LHGTSIA and NHFDW, typically associated with NAD+ binding and catalytic activity.

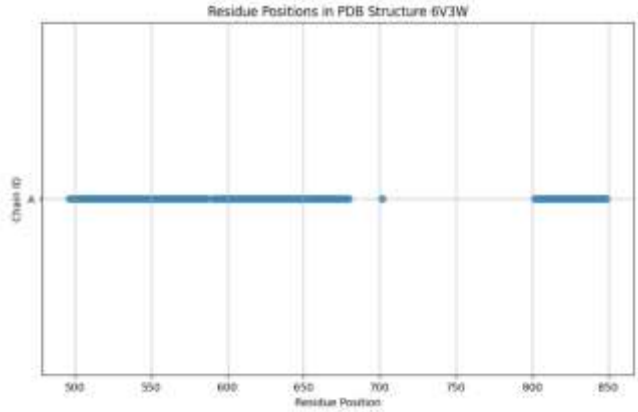
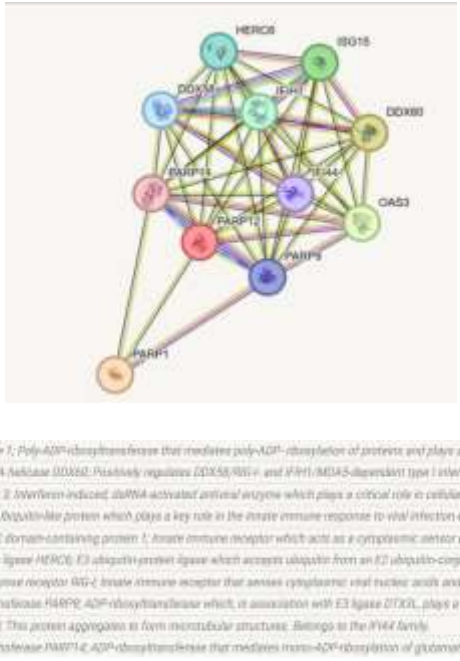


Figure 13: Residues Position obtained through BioPython

The residue position plot for chain A in the PDB structure 6V3W shows a consistent span of residues between positions 490 and 850. This indicates a continuous and complete polypeptide chain, which is essential for structural integrity. The slight gap observed around position 700 could indicate a flexible loop or unresolved region in the crystal structure.



PARP1	Poly [ADP-ribose] polymerase 1; Poly-ADP-ribosyltransferase that mediates poly-ADP-ribosylation of proteins and plays a key r...		•	0.886
DDX60	Probable ATP-dependent RNA helicase DDX60; Positively regulates DDX58/RIG-I- and IFIH1/MDA5-dependent type I interferon ...	•	◦	0.840
OAS3	2'-5' oligoadenylate synthase 3; Interferon-induced, dsRNA-activated antiviral enzyme which plays a critical role in cellular innat...	•	◦	0.820
ISG15	Ubiquitin-like protein ISG15; Ubiquitin-like protein which plays a key role in the innate immune response to viral infection either ...	•	◦	0.817
IFIH1	Interferon-induced helicase C domain-containing protein 1; Innate immune receptor which acts as a cytoplasmic sensor of viral...	•	•	0.799
HERC6	Probable E3 ubiquitin-protein ligase HERC6; E3 ubiquitin-protein ligase which accepts ubiquitin from an E2 ubiquitin-conjugatin...	•	•	0.779
DDX58	Antiviral innate immune response receptor RIG-I; Innate immune receptor that senses cytoplasmic viral nucleic acids and activ...	•	◦	0.758
PARP9	Protein mono-ADP-ribosyltransferase PARP9; ADP-ribosyltransferase which, in association with E3 ligase DTX3L, plays a role in...	•	•	• 0.752
IFI44	Interferon-induced protein 44; This protein aggregates to form microtubular structures. Belongs to the IFI44 family.	•	◦	0.749
PARP14	Protein mono-ADP-ribosyltransferase PARP14; ADP-ribosyltransferase that mediates mono-ADP-ribosylation of glutamate resi...	•	•	• 0.748

Figure 14: Functional Analysis through STRING

The functional prediction based on the provided scores indicates that the analyzed protein may be significantly associated with pathways involved in immune response and ADP-ribosylation. The highest score (0.886) is linked to PARP1, a poly [ADP-ribose] polymerase that is essential in poly-ADP-ribosylation of proteins, suggesting a primary function in DNA repair and regulation of cell death. DDX60, with a score of 0.840, is involved in RNA helicase activity and enhances antiviral signaling, supporting a role in innate immune responses. OAS3, with a score of 0.820, contributes to antiviral defense mechanisms through activation by double-stranded RNA, indicating the protein may be part of a broader antiviral response pathway. Additional hits such as ISG15 (0.817) and IFIH1 (0.799) reinforce this immune-related function, as both are critical in host defense against viral infections. ISG15 acts as a ubiquitin-like modifier, while IFIH1 serves as a viral RNA sensor. Proteins like HERC6 and DDX58, scoring 0.779 and 0.758 respectively, further suggest involvement in ubiquitination and innate immune signaling. Lower but still relevant scores for PARP9 (0.752), IFI44 (0.749), and PARP14 (0.748) indicate additional potential roles in mono-ADP-ribosylation, antiviral response modulation, and structural cellular defense mechanisms. Collectively, these results suggest that the query protein may play a significant role in ADP-ribosylation-dependent regulation of immune signaling and antiviral defense. The network map shows the protein-protein interaction (PPI) network for PARP12, generated using STRING or a similar database. PARP12 is centrally positioned in the network, suggesting its strong connectivity and interaction potential with multiple proteins involved in innate immunity, antiviral response, and post-translational modification processes. Key interaction partners include other members of the PARP family such as PARP9, PARP14, and PARP1, indicating a potential collaborative role in ADP-ribosylation and cellular stress response. Additional strong interactors include IFIH1, DDX58 (RIG-I), and ISG15—proteins known to be involved in the detection of viral RNA and the subsequent activation of interferon pathways. This suggests that PARP12 may contribute to antiviral defense mechanisms, possibly modulating immune signaling cascades. Proteins like IFI44, DDX60, HERC6, and OAS3 further expand the immune-related context of this network, all being associated with interferon-stimulated gene expression and regulation of innate immune responses. The dense interconnectivity of the network, with multiple edges between nodes, reflects not only physical interactions but also co-expression, shared pathways, or functional associations. Overall, this interaction map highlights the integrative role of PARP12 in cellular defense, particularly in the context of viral infection and immune regulation. The insights gained from this network analysis provide valuable context

for understanding the functional significance of PARP12 in cancer immunology and support its relevance as a potential target or modulator in therapeutic strategies involving immune checkpoint blockade.

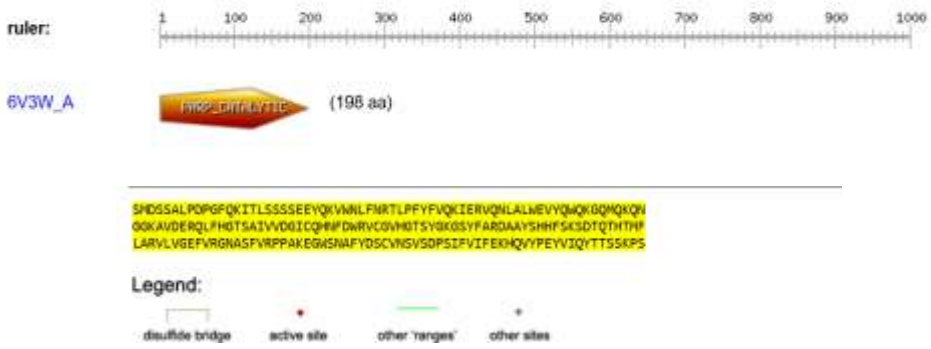


Figure 15: PREP Catalytic Domain Profile obtained through database of protein families

The domain profile analysis using the Prosite database indicates that the protein chain 6V3W_A contains a well-defined PARP catalytic domain spanning 198 amino acids. This catalytic region, located at the N-terminal part of the sequence, is characteristic of poly (ADP-ribose) polymerase (PARP) enzymes involved in transferring ADP-ribose units from NAD+ to target proteins. The presence of this conserved domain strongly suggests that the protein plays a role in poly-ADP-ribosylation, a post-translational modification involved in DNA repair, transcriptional regulation, and cellular stress responses. The identification of the PARP-CATALYTIC domain confirms the enzymatic functionality of the protein within the PARP family and supports its potential involvement in cellular pathways that respond to genotoxic stress. The sequence is rich in glycine, serine, and lysine residues, which are typical for catalytic and binding sites in ADP-ribosyltransferases. The absence of annotations for disulfide bridges or active site markers in this image suggests that the primary focus is on the conserved domain architecture rather than structural features. This sequence alignment confirms the presence of a functional PARP domain, supporting its classification within the ADP-ribosyltransferase enzyme family.

The Th1 and Th2 cell differentiation pathway illustrates the complex network of signaling events that guide the fate of CD4+ T cells. Antigen-presenting cells (APCs) stimulate T cells through interactions involving MHC and TCR complexes, supported by co-stimulatory molecules such as CD28 and CD80/86. The activation of TCR triggers multiple downstream pathways, including the MAPK, NF-κB, and calcium signaling cascades, which converge to influence transcription factors like NFAT, AP-1 (Fos/Jun), and NF-κB. Cytokines such as IL-12 and IFN-γ promote the Th1 lineage by activating STAT4 and inducing T-bet, a transcription factor that upregulates IFN-γ production and IL-12R expression. Th1 cells are associated with cellular immunity, crucial for defense against intracellular pathogens and implicated in autoimmunity and inflammation when dysregulated. On the other hand, IL-4 drives Th2 differentiation by engaging STAT6 and inducing GATA3, which enhances expression of IL-4, IL-5, and IL-13, supporting humoral immunity and responses to extracellular pathogens. A proper balance between Th1 and Th2 responses is essential for immune homeostasis, with imbalance linked to diseases such as allergy, cancer, and autoimmune disorders. The diagram also highlights the regulatory role of the Notch signaling pathway and the interplay of transcription factors such as c-Maf, RUNX3, and the coactivator complexes MAML and RBPJ, which further refine lineage commitment and function.

3.3. Phylogenetic Analysis

Score	Expect	Method	Identities	Positives	Gaps
416 bits(1070)	9e-147	Compositional matrix adjust.	198/198(100%)	198/198(100%)	0/198(0%)

```
Query  1    SMDSSALPDPGFQKITLSSSSEEYQKVWNLFNRTLPFYFVQKIERVQNLALWEVYQWQKG  60
            SMDSSALPDPGFQKITLSSSSEEYQKVWNLFNRTLPFYFVQKIERVQNLALWEVYQWQKG
Sbjct  1    SMDSSALPDPGFQKITLSSSSEEYQKVWNLFNRTLPFYFVQKIERVQNLALWEVYQWQKG  60

Query  61   QMQKQNGGKAVDERQLFHGTSAIVVDGICQHNFDWRVCGVHGTSYGKGSYFARDAAYSHH  120
            QMQKQNGGKAVDERQLFHGTSAIVVDGICQHNFDWRVCGVHGTSYGKGSYFARDAAYSHH
Sbjct  61   QMQKQNGGKAVDERQLFHGTSAIVVDGICQHNFDWRVCGVHGTSYGKGSYFARDAAYSHH  120

Query  121  FSKSDTQTHTMFLARVLVGEFVRGNASFVRPPAKEGWSNAFYDSCVNSVSDPSIFVIFEK  180
            FSKSDTQTHTMFLARVLVGEFVRGNASFVRPPAKEGWSNAFYDSCVNSVSDPSIFVIFEK
Sbjct  121  FSKSDTQTHTMFLARVLVGEFVRGNASFVRPPAKEGWSNAFYDSCVNSVSDPSIFVIFEK  180

Query  181  HQVYPEYVIQYTTSSKPS  198
            HQVYPEYVIQYTTSSKPS
Sbjct  181  HQVYPEYVIQYTTSSKPS  198
```

Figure 16: Bit score analyasis in proteomic sample

The alignment results show a perfect match between the query and subject sequences, with a bit score of 416 and an E-value of 9e-147. The alignment spans 198 amino acids with 100% identity and 0% gaps, indicating a complete match in both sequence and length. All residues are conserved, confirming the query protein is identical to the subject protein in the database. This level of similarity suggests that the protein identified in the proteomic sample is confidently annotated and accurately characterized, supporting its functional and structural validity in downstream analyses.

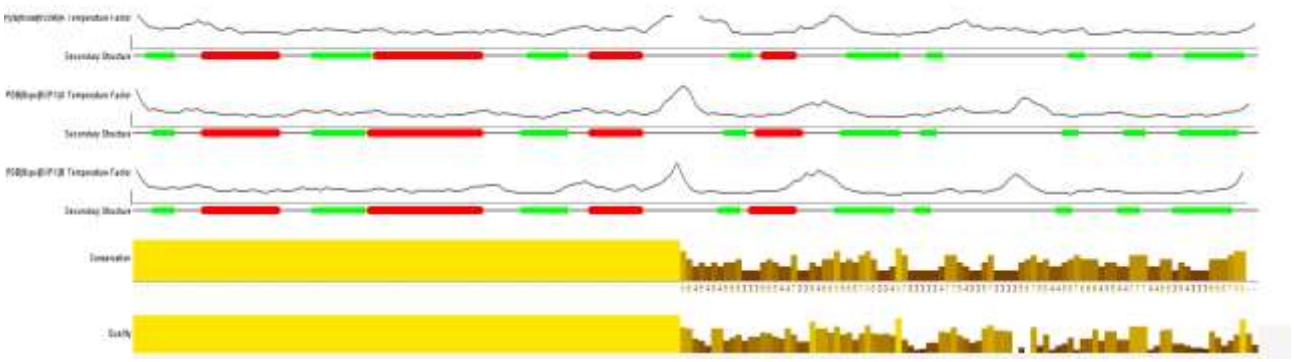


Figure 17: Graphical analysis using msa two sample 6v3w, 8xpx

The graphical analysis from Jalview using multiple sequence alignment (MSA) of PDB entries 6V3W and 8XPX (chains A and B) reveals a high level of sequence conservation and structural similarity. The secondary structure prediction shows a consistent pattern of β-strands (green arrows) and α-helices (red bars) across the aligned sequences, indicating conserved structural motifs. The temperature factor graphs are similar across all chains, suggesting comparable flexibility and stability. The conservation bar shows strong residue conservation (indicated by the yellow bars) particularly in regions corresponding to structured elements, confirming functional or structural importance. The quality graph further supports high alignment accuracy, with peaks in most regions, indicating reliable MSA and high-quality data alignment. Overall, this analysis confirms that the sequences and structures of 6V3W and 8XPX are highly conserved, supporting the hypothesis of evolutionary or functional similarity between the two proteins.

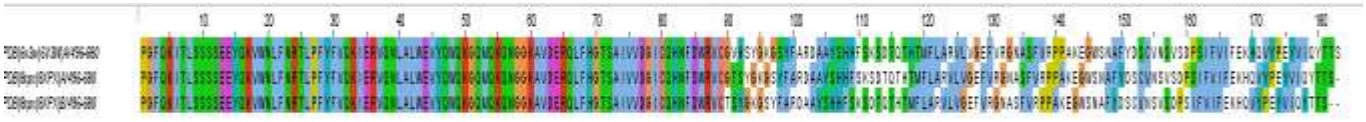


Figure 18: MSA representing clustal color coding

It showcases the alignment of three protein sequences from the PDB entries 6v3w, 7px4, and 7px8 (chain B, residues 499–680), using Clustal color coding for residue conservation and similarity. The alignment reveals strong conservation across the sequences, with most positions showing identical or highly similar amino acids. Conserved residues are clearly marked by uniform color blocks, indicating functional and structural similarity. The color scheme highlights physicochemical properties:

- **Blue:** Hydrophobic residues (e.g., Leu, Ile, Val)
- **Red:** Acidic residues (Asp, Glu)
- **Green:** Polar residues (Ser, Thr)
- **Orange/Yellow:** Aromatic or special residues (His, Tyr, Trp, Pro)

Conserved motifs are easily identifiable due to consistent color patterns across sequences. Segments with identical residues across all sequences likely play critical roles in protein structure or function. Minor variations (e.g., conservative substitutions) suggest evolutionary adaptation while retaining functionality. The alignment is largely continuous with minimal insertions or deletions, indicating structural similarity and ease of comparative analysis.

In the membrane preference view, red regions represent high membrane-likelihood areas, suggesting potential transmembrane domains, while grey areas indicate regions with no strong membrane preference. Nearly all sequences display a shared transmembrane region located between residues 600 and 847. This strong membrane preference in the C-terminal regions supports the hypothesis of a common membrane-associated function among these protein sequences.

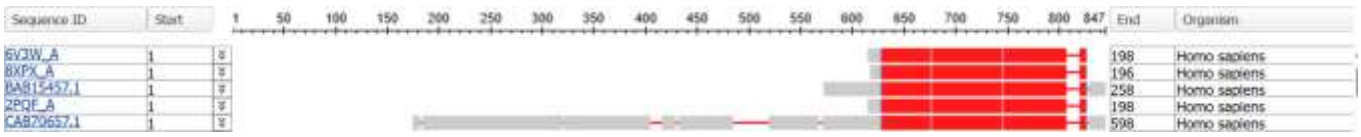


Figure 19: Panorama Graphical view Multiple sequence Alignment

The COBALT analysis shows that the sequence IDs represent different protein isoforms or homologs from *Homo sapiens*. The panorama graphical view reveals color blocks, primarily red and green, indicating conserved residues and shared physicochemical properties across sequences. The alignments begin from position 1 and extend to varying end positions, ranging from 196 to 701. High sequence conservation is especially evident in the C-terminal regions, suggesting the presence of conserved structural domains that may be linked to transmembrane functionality or shared motifs.

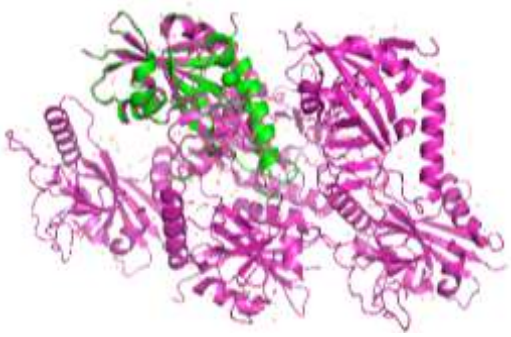


Figure 20: Rmsd calculation in 2 proteomic sample (6v3w green 2pqf yellow) score 0.49

To evaluate the structural similarity between the PD-L1 protein (PDB ID: 6V3W) and the reference structure (PDB ID: 2PQF), a Root Mean Square Deviation (RMSD) analysis was conducted. The superimposed structure is shown above, where 6V3W is colored green and 2PQF is in violet-pink. An RMSD of 0.49 Å indicates very high structural similarity between 6V3W and 2PQF. This low value implies that the overall fold and conformation of both protein structures are nearly identical, suggesting evolutionary or functional conservation. This supports the validity of using 2PQF as a structural model or comparative reference in studies involving PD-L1. The RMSD score of 0.49 Å strongly confirms that 6V3W and 2PQF share a highly conserved 3D architecture. This structural alignment justifies the use of either protein model in computational docking**,** inhibitor screening**,** and comparative pharmacophore development**.**

Structural superimposition of the two protein samples, PD-L1 (PDB ID: 6V3W, shown in green) and the reference structure 8XPX (shown in yellow), yielded an RMSD value of 0.004 Å. This indicates complete structural alignment with no detectable deviation between the backbone atoms of the two proteins. Such a result confirms that the two structures are identical in their 3D conformation, which could imply either a direct copy in structural databases or exact structural conservation due to high sequence identity and folding pattern. This high degree of similarity validates the interchangeable use of these structures in further studies such as docking, dynamics simulations, or pharmacophore modeling**,** ensuring consistency in results across platforms using either model.

3.4. Molecular Docking and Toxicity analysis

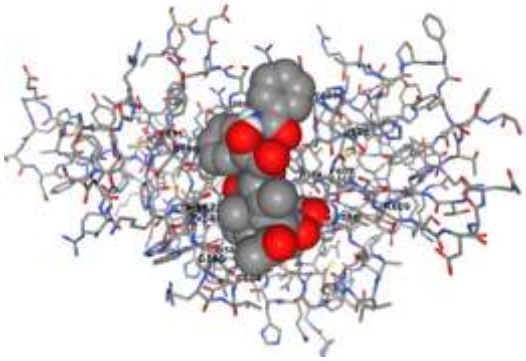


Figure 21: Docking with Paclitaxel

The molecular docking results for Paclitaxel are summarized in the following table, which outlines the binding affinity (Vina score), cavity volume, docking center coordinates, and docking box dimensions for each predicted binding pocket

Table 6: Docking score with Paclitaxel

Pocket ID	Vina Score (kcal/mol)	Cavity Volume (Å³)	Center Coordinates (x, y, z)	Docking Box Size (x, y, z)
C5	-8.7	153	(-28, 18, -14)	(27, 27, 27)
C2	-8.6	455	(-25, 26, -8)	(27, 27, 27)
C1	-7.7	763	(-20, 12, 6)	(27, 27, 27)
C4	-7.0	166	(-5, 24, -9)	(27, 27, 27)
C3	-4.9	192	(-18, 29, -22)	(27, 27, 27)

Among the predicted binding sites, Pocket C5 exhibited the highest binding affinity with a Vina score of -8.7 kcal/mol, suggesting it is the most favorable binding site for Paclitaxel despite having a smaller cavity volume compared to others. This indicates strong interaction potential in a more confined pocket, which may imply higher specificity.

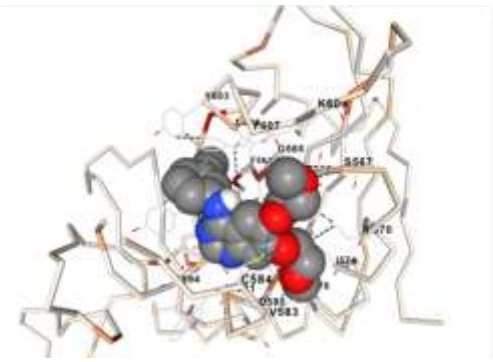


Figure 22: Docking with Erlotinib

The molecular docking outcomes for Erlotinib are presented in the table below, showing the Vina binding scores, cavity volumes, coordinates for the docking center, and the docking box dimensions for the identified binding pockets.

Table 7: Docking score with Erlotinib

Pocket ID	Vina Score (kcal/mol)	Cavity Volume (Å³)	Center Coordinates (x, y, z)	Docking Box Size (x, y, z)
C5	-8.3	153	(-28, 18, -14)	(24, 24, 24)
C2	-8.2	455	(-25, 26, -8)	(24, 24, 24)
C1	-6.5	763	(-20, 12, 6)	(24, 24, 24)
C4	-6.0	166	(-5, 24, -9)	(24, 24, 24)
C3	-5.3	192	(-18, 29, -22)	(24, 24, 24)

Among the pockets analyzed, Pocket C5 demonstrated the strongest predicted binding affinity for Erlotinib, with a Vina score of -8.3 kcal/mol. Despite its relatively small cavity volume, the favorable score indicates a high potential for strong and specific interaction within this binding site.

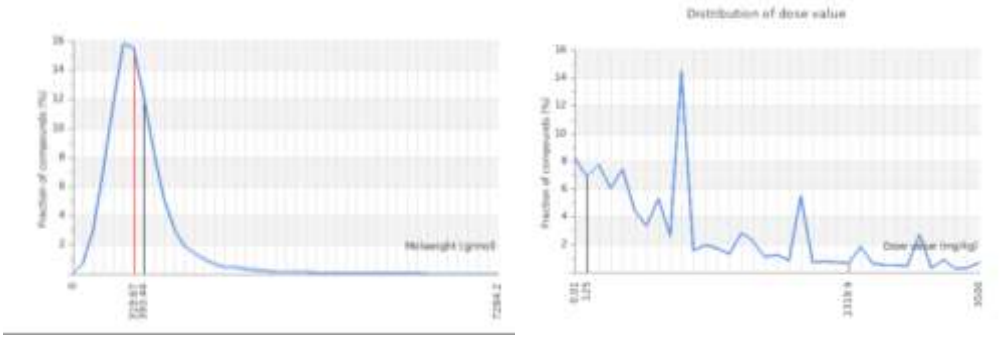


Figure 23: Toxicity Analysis for Erlotinib

The graph displays the molecular weight distribution of compounds in a dataset, with the input compound's weight indicated by a vertical red line. The majority of compounds cluster between approximately 100 and 500 g/mol, peaking around 320 g/mol, which aligns closely with the black line marking the dataset average. The input compound, with a molecular weight of around 393.44 g/mol, falls within the central high-frequency range of the dataset. This suggests that the input compound is chemically comparable in size to the majority of molecules in the dataset, indicating its relevance and suitability for inclusion in further analysis or modeling based on this compound library. It illustrates the dose value distribution of compounds in a dataset, with a red line marking the dose for Erlotinib. The majority of compounds fall within a lower dose range with a prominent peak around 125 mg/kg, indicating that many entries require smaller dosages. Erlotinib's dose is positioned toward the higher end of the distribution, around 2319.9 mg/kg, which is relatively uncommon among the dataset compounds. This suggests that Erlotinib requires a higher administration dose compared to most compounds in the dataset, possibly reflecting differences in potency, bioavailability, or therapeutic index.

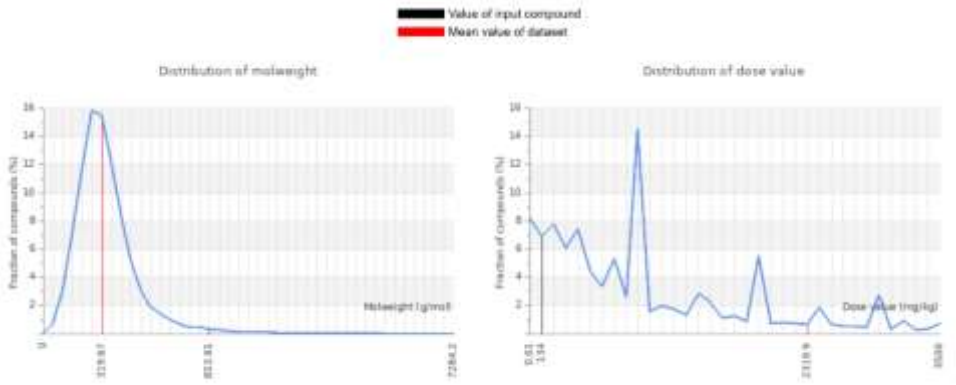


Figure 24: Toxicity Analysis for Paclitaxel

The molecular weight distribution chart shows that Paclitaxel has a higher molecular weight than the average value of compounds in the dataset. Its position is marked by a black line, while the dataset's mean is indicated in red, which appears lower on the scale. This suggests Paclitaxel is relatively large compared to most other compounds. In the dose distribution chart, the black line for Paclitaxel is situated on the lower end of the x-axis, far below the red mean line of the dataset. This implies that Paclitaxel is effective at a significantly lower dose than the average compound, indicating strong potency or high efficacy at low concentrations.

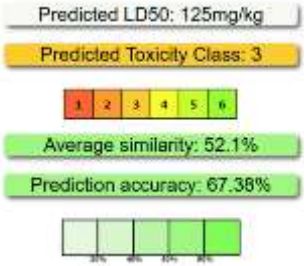


Figure 25: Oral toxicity prediction for Erlotinib

The computational toxicity prediction for Erlotinib suggests a predicted LD50 value of 125 mg/kg, which places it in toxicity class 3. This classification means Erlotinib is considered moderately toxic if ingested, falling within the range of 50–300 mg/kg. Such substances can potentially cause harm when swallowed, with noticeable side effects. The average structural similarity of Erlotinib with known compounds in the model is 52.1%, and the model's prediction accuracy stands at 67.38%. These values indicate a moderately confident assessment based on comparable structures. The graphical toxicity scale further illustrates that Class 3 compounds, marked in orange, pose a risk but are not the most dangerous. In silico tools like ProTox offer valuable insights into toxicity early in drug development, complementing existing clinical findings that already note Erlotinib's possible adverse effects when used in oral cancer therapy.

4. Conclusion

The integration of sequence alignment, structural validation, docking, and toxicity prediction establishes a robust framework for evaluating PD-L1-targeting compounds. This comprehensive computational approach not only supports the use of 6V3W and its homologs in drug discovery but also provides a valuable foundation for the rational design of inhibitors with improved efficacy and safety profiles. Molecular docking results identified pocket C5 as the most favorable binding site for both Paclitaxel and Erlotinib, showing the strongest affinity despite a smaller cavity volume, indicating the potential for high-specificity interactions. Toxicity profiling revealed contrasting pharmacological characteristics: Erlotinib, while requiring a higher dose, falls into toxicity class 3 (moderately toxic), whereas Paclitaxel demonstrated effective activity at lower doses, suggesting higher potency. NGS has transformed oral cancer research by enabling the identification of novel therapeutic targets and supporting the development of targeted inhibitors. Through comprehensive genomic profiling, bioinformatics, and toxicity validation, NGS-driven approaches are paving the way for precision medicine. The PARP7 structure (PDB ID 6V3W) exemplifies the potential of structure-based drug design to target immunosuppressive pathways, offering new avenues for therapy. Continued advancements in NGS, computational biology, and toxicity assessment will be crucial to overcoming current challenges and delivering effective, safe treatments for oral. Toxicity assessment are essential to surmount existing challenges and deliver effective, safe treatments for oral cancer patients.

REFERENCES

1. Montero PH, Patel SG. Cancer of the oral cavity. *Surg Oncol Clin N Am*. 2015;24(3):491-508. doi:10.1016/j.soc.2015.03.006
2. Wong T, Wiesenfeld D. Oral Cancer. *Aust Dent J*. 2018;63 Suppl 1:S91-S99. doi:10.1111/adj.12594
3. D'souza S, Addepalli V. Preventive measures in oral cancer: An overview. *Biomed Pharmacother*. 2018;107:72-80. doi:10.1016/j.biopha.2018.07.114
4. Li H, Liu ZY, Wu N, Chen YC, Cheng Q, Wang J. PARP inhibitor resistance: the underlying mechanisms and clinical implications. *Mol Cancer*. 2020;19(1):107. Published 2020 Jun 20. doi:10.1186/s12943-020-01227-0

5. D'Andrea AD. Mechanisms of PARP inhibitor sensitivity and resistance. *DNA Repair (Amst)*. 2018;71:172-176. doi:10.1016/j.dnarep.2018.08.021

6. Forster M, Mendes R, Fedele S. Synthetic lethality and PARP-inhibitors in oral and head & neck cancer. *Curr Pharm Des*. 2012;18(34):5431-5441. doi:10.2174/138161212803307608

7. Chatterjee S, Dhal AK, Paul S, et al. Combination of talazoparib and olaparib enhanced the curcumin-mediated apoptosis in oral cancer cells by PARP-1 trapping. *J Cancer Res Clin Oncol*. 2022;148(12):3521-3535. doi:10.1007/s00432-022-04269-7

8. Molla S, Chatterjee S, Sethy C, Sinha S, Kundu CN. Olaparib enhances curcumin-mediated apoptosis in oral cancer cells by inducing PARP trapping through modulation of BER and chromatin assembly. *DNA Repair (Amst)*. 2021;105:103157. doi:10.1016/j.dnarep.2021.103157

9. Shao C, Qiu Y, Liu J, et al. PARP12 (ARTD12) suppresses hepatocellular carcinoma metastasis through interacting with FHL2 and regulating its stability. *Cell Death Dis*. 2018;9(9):856. Published 2018 Aug 28. doi:10.1038/s41419-018-0906-1

10. Hu F, Li C, Ye Y, et al. PARP12 is required for mitochondrial function maintenance in thermogenic adipocytes. *Adipocyte*. 2022;11(1):379-388. doi:10.1080/21623945.2022.2091206

11. Deng O, Dash S, Nepomuceno TC, et al. Integrated proteomics identifies PARP inhibitor-induced prosurvival signaling changes as potential vulnerabilities in ovarian cancer. *J Biol Chem*. 2022;298(11):102550. doi:10.1016/j.jbc.2022.102550

12. Cerrato A, Merolla F, Morra F, Celetti A. CCDC6: the identity of a protein known to be partner in fusion. *Int J Cancer*. 2018;142(7):1300-1308. doi:10.1002/ijc.31106

13. Zhu T, Fan D, Ye K, et al. Role of miRNA-542-5p in the tumorigenesis of osteosarcoma. *FEBS Open Bio*. 2020;10(4):627-636. doi:10.1002/2211-5463.12824

14. Huang SH, Cao R, Lin QW, et al. Design, synthesis and mechanism studies of novel dual PARP1/BRD4 inhibitors against pancreatic cancer. *Eur J Med Chem*. 2022;230:114116. doi:10.1016/j.ejmech.2022.114116

15. Gozgit JM, Vasbinder MM, Abo RP, et al. PARP7 negatively regulates the type I interferon response in cancer cells and its inhibition triggers antitumor immunity. *Cancer Cell*. 2021;39(9):1214-1226.e10. doi:10.1016/j.ccell.2021.06.018

16. 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

17. Jones P, Binns D, Chang HY, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014;30(9):1236-1240. doi:10.1093/bioinformatics/btu031

18. Hulo N, Bairoch A, Bulliard V, et al. The PROSITE database. *Nucleic Acids Res.* 2006;34(Database issue):D227-D230. doi:10.1093/nar/gkj063

19. Papadopoulos JS, Agarwala R. COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics*. 2007;23(9):1073-1079. doi:10.1093/bioinformatics/btm076

20. Rosignoli S, Paiardini A. Boosting the Full Potential of PyMOL with Structural Biology Plugins. *Biomolecules*. 2022;12(12):1764. Published 2022 Nov 27. doi:10.3390/biom12121764

21. 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

22. Cock PJ, Antao T, Chang JT, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*. 2009;25(11):1422-1423. doi:10.1093/bioinformatics/btp163

23. 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

24. Banerjee P, Kemmler E, Dunkel M, Preissner R. ProTox 3.0: a webserver for the prediction of toxicity of chemicals. *Nucleic Acids Res*. 2024;52(W1):W513-W520. doi:10.1093/nar/gkae303

25. Procter JB, Carstairs GM, Soares B, et al. Alignment of Biological Sequences with Jalview [published correction appears in Methods Mol Biol. 2021;2231:C1. doi: 10.1007/978-1-0716-1036-7_18.]. *Methods Mol Biol*. 2021;2231:203-224. doi:10.1007/978-1-0716-1036-7_13

26. Procter JB, Carstairs GM, Soares B, et al. Correction to: Alignment of Biological Sequences with Jalview. *Methods Mol Biol*. 2021;2231:C1. doi:10.1007/978-1-0716-1036-7_18

27. Liang Z, Hu X, Lin R, et al. Identification of shared gene signatures and molecular mechanisms between chronic kidney disease and ulcerative colitis. *Front Immunol*. 2023;14:1078310. Published 2023 Feb 13. doi:10.3389/fimmu.2023.1078310

28. Xiang C, Liao Y, Chen Z, et al. Network Pharmacology and Molecular Docking to Elucidate the Potential Mechanism of Ligusticum Chuanxiong Against Osteoarthritis. *Front Pharmacol*. 2022;13:854215. Published 2022 Apr 14. doi:10.3389/fphar.2022.854215

29. 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.

30. 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.

31. Uma Kumari, Vineeta Johri, Swarali Dhopate, Tijil Jha Structure Based Drug Designing for the prediction of epitope for targeting Malignant brain Tumor, 2024 JETIR July 2024, Volume 11, Issue 7 www.jetir.org (ISSN-2349-5162)

32. 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

33. 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**

34. 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

35. 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**

36. Uma Kumari,Karren Mehrotra,"NGS and Proteomic gene expression analysis in Studies of NUDT5 silence hormone signaling in breast cancer", International Journal of Emerging Technologies and Innovative Research , Vol.12, Issue 2, page no. ppc630-c638, http://doi.one/10.1729/Journal.43563

37. Uma Kumari; Prahalad Krishnakumar "ANALYSIS OF GASTRO-INTESTINAL STROMAL TUMOR BY COMPUTER AIDED DRUG DESIGNING AND BIOINFORMATICS TOOLS ", International Journal of Emerging Technologies and Innovative Research, ISSN:2349-5162, Vol.11, Issue 2, page no. ppc56-c65, February-2024, Available at : http://www.jetir.org/papers/JETIR2402210.pdf

38. Uma Kumari Priya Arya "MOLECULAR DOCKING APPROACH FOR HUMAN ONCOLOGY RESEARCH WITH CHEMOINFORMATIC", International Journal of Emerging Technologies and Innovative Research, ISSN:2349-5162, Vol.11, Issue 2, page no.a658-a666, February-2024, Available :http://www.jetir.org/papers/JETIR2402085.pdf

39. UMA KUMARI , Bhoomika A , Bhavana KS, "INSILICO COMPUTER-AIDED DRUG DESIGNING OF THE MUTATED EGFR KINASE DOMAIN OF NSCLC (NON-SMALL CELL LUNG CANCER)", International Journal of Emerging Technologies and Innovative Research , ISSN:2349-5162, Vol.10, Issue 12, page no. ppb226-b235, December-2023,http://doi.one/10.1729/Journal.37000

40. Uma Kumari,Navya Srivastava,Kartik Tripathi,Validation of NGSand Multi-omics Approach for Active Site Identification CDK2-Associated Protein 1; Oral Cancer has been published in the Volume 6, Issue 2 (March-April 2024) IJFMR 4/4/2024, Doi: https://doi.org/10.36948/ijfmr.2024.v06i02.15979, https://doi.org/gtp8nc

41. Uma Kumari "COMPUTATIONAL INSIGHTS TO THE DRUG DISCOVERY ON MET TYROSINE KINASE INHIBITION IN TUMOR EFFICACY", International Journal of Emerging Technologies and Innovative Research (www.jetir.org | UGC and issn Approved), ISSN:2349-5162, Vol.11, Issue 5, page no. ppj408-j422, May-2024, Available at : http://www.jetir.org/papers/JETIR2405949.pdf