

Sustainable Neuroprotection Via Proteomics And Integrative NGS- CADD Analysis Of 6CVP Variant To Improve Future Neurodegenerative Health Outcomes

Vineeta Johri¹, Uma Kumari*¹

¹Research Scholar, Department Of Biotechnology, NIMS Institute Of Allied Medical Science And Technology, NIMS University, Jaipur, India

¹*Senior Scientist, Department Of Bioinformatics, Bioinformatics Project And Research Institute, Noida, India

Abstract

Alzheimer's, Parkinson's, and Huntington's diseases, classified as neurodegenerative disorders, are major global health concerns because of significant neuronal degeneration and the lack of effective treatments. Advancements in high-throughput biotechnology and bioinformatics techniques provide new insights into the molecular mechanisms of neurodegeneration. This study presents a sustainable proteomics methodology combining next-generation sequencing (NGS) with computer-aided drug design (CADD) to explore neuroprotective mechanisms related to structural variants, exemplified by a protein model. The approach uses NGS data to pinpoint genetic variants involved in neurodegenerative processes, which are then mapped onto the 3D structure of the target protein, using the Protein Data Bank entry 6CVP, enabling analysis of structural changes and their effects. Techniques such as protein interaction analysis, pathway enrichment, and system-level network mapping assist in understanding how these proteins influence neuronal survival and neuroprotection. Additionally, CADD methods like molecular docking and virtual screening are applied to find potential neuroprotective compounds that may stabilize or alter the mutated protein's structure. This integrated strategy aims to identify molecular signatures that connect genetic variations with protein dysfunction and neurodegenerative pathways. By merging NGS mutation profiling, proteomic network analysis, and computational drug discovery, this platform supports the identification of new therapeutic targets and neuroprotective agents. Overall, this multi-omics approach underscores the significance of integrating genomic, proteomic, and structural bioinformatics to better understand neurodegenerative disease mechanisms. These insights could lead to earlier diagnosis, personalized treatments, and the development of future neuroprotective therapies, ultimately improving long-term neurological health and advancing personalized medicine for these disorders.

Keywords: Neurodegenerative diseases, sustainable proteomics, Next-generation sequencing (NGS), Computer-aided drug design (CADD), Structural bioinformatics, Molecular docking, Protein stability analysis, Neuroprotection

Date of Submission: 26-04-2026

Date of Acceptance: 06-05-2026

I. Introduction

Neurodegenerative diseases are among the most pressing health issues of the current century. Alzheimer's, Parkinson's, and Huntington's disease involve progressive loss of neuronal function, protein misfolding, and slowly worsening cognitive or motor skills. These disorders are complex, resulting from intricate interactions among genetic mutations, environmental influences, and molecular signalling pathways. [1,2] As global life expectancy continues to rise, so does the prevalence of these conditions, underscoring the urgent need for improved diagnostics and treatment options. Advancements in genomic technology, particularly Next-Generation Sequencing (NGS), have greatly enhanced our ability to detect genetic mutations associated with neurodegenerative diseases. NGS enables high-throughput identification of genetic variants, including SNPs, insertions, deletions, and structural changes, across the genome. These mutations can affect protein structure and function, contributing to neuronal decline. To fully understand their biological significance, it is essential to integrate genetic data with structural and functional analyses of proteins. Structural bioinformatics is crucial for understanding how mutations affect protein structure and stability. The three-dimensional structures stored in the Protein Data Bank provide valuable insights into molecular interactions and key functional areas. In this study, PDB 6CVP was selected as a representative model to examine mutation-induced changes related to neurodegeneration. Structural analysis allows investigation of binding pockets, conformational flexibility, and potential drug-binding sites. [3] Proteomics has become an essential tool for studying the functional effects

of genomic changes. Sustainable proteomics combines high-throughput protein identification with systems biology to analyze protein interactions, signalling pathways, and regulatory networks. These techniques aid in identifying biomarkers and therapeutic targets for neuroprotection. Protein interaction maps, created using databases like STRING, help understand pathways involved in neuronal survival and synaptic function. [4,5]

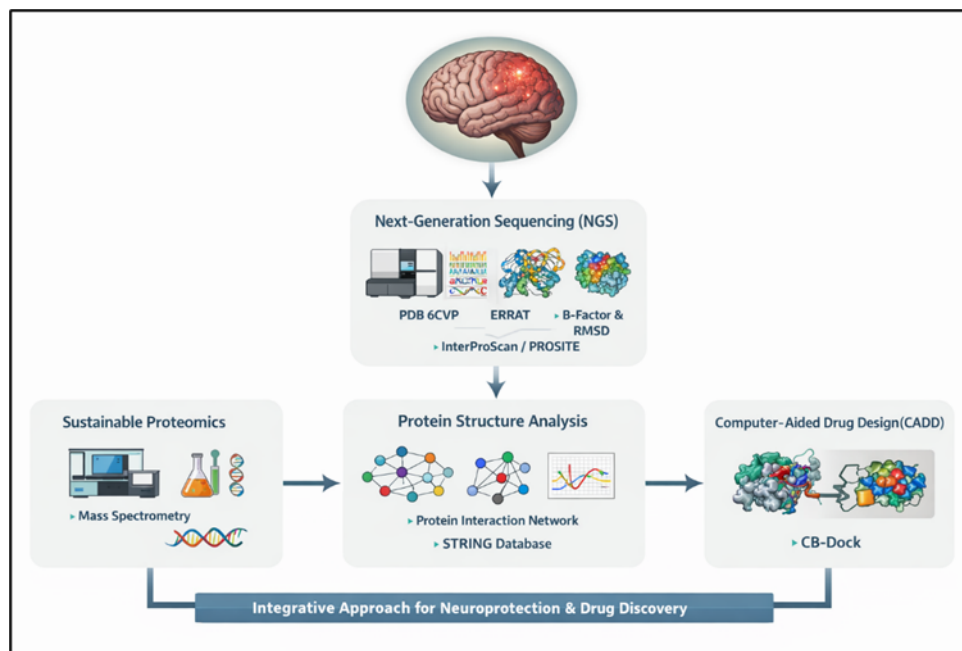


Figure 1: Integrative approach for Neuroprotection and Drug Discovery

Validating protein structures is another fundamental step in computational analysis. Tools like ERRAT evaluate model quality by examining non-bonded atomic interactions. Indicators of structural stability, such as B-factor analysis and RMSD, are used to assess protein flexibility and conformational integrity. High B-factor values often point to regions of increased flexibility or instability, which can impact ligand binding and activity. [6,7] Predicting functional domains offers additional understanding of protein functions. Tools such as InterProScan and PROSITE detect conserved motifs and domains that are key to catalytic activity, ligand binding, or maintaining structural stability. Identifying these regions is essential for grasping how mutations could interfere with important functional areas of the protein. [8,9]

Computer-aided drug design (CADD) improves the process of discovering therapeutic compounds targeting neurodegenerative pathways. Tools like CB-Dock automatically locate binding sites and assess ligand-protein interactions. Integrating structural data with docking simulations helps identify and refine potential neuroprotective drugs. [10,11] The combination of NGS, proteomics, and structural bioinformatics offers a comprehensive multidisciplinary strategy for studying neurodegenerative diseases. This approach enables mutation detection, analysis of structural effects, and discovery of therapeutic candidates. The goal of this study is to develop a sustainable computational framework that combines mutation profiling, structural validation, functional annotation, and docking analysis, focusing on the protein with PDB ID 6CVP. [12] Such an integrated method can enrich our understanding of the molecular mechanisms behind neurodegeneration and support the development of neuroprotective strategies. Ultimately, integrating genomic, structural, and proteomic data can accelerate personalized treatments and improve neurological health outcomes. [13]

II. Methodology

The present study employed an integrative computational framework combining genomics, structural bioinformatics, proteomics, and computer-aided drug design to investigate mutation-induced structural and functional changes associated with neurodegenerative diseases.

1. Retrieval of Protein Structure

Our target protein's three-dimensional structure was obtained from the Protein Data Bank under the ID 6CVP. It was downloaded in PDB format and used as the main template for both structural and docking analyses.

2. Mutation Identification through NGS Data

Genomic mutation data related to neurodegenerative diseases were sourced from published NGS datasets. These mutations were mapped onto the protein structure to examine their locations and possible functional effects. Protein structural data and conserved domain arrangements were examined using the Molecular Modelling Database. MMDB offered comprehensive details on protein folds, structural alignments, and domain organization.

3. Protein Structure Validation

Structural validation involved using the ERRAT server to assess non-bonded atomic interactions and identify possible structural errors. The overall quality score helped determine the reliability of the protein model.

5. Functional Domain Prediction

Protein sequence analysis was performed using:

- Sequence Scanning
- PROSITE

These tools were used to identify conserved domains, motifs, and functional regions within the protein sequence.

6. Structural Stability Analysis

Protein flexibility and stability were analysed using B-factor analysis, which measures atomic displacement in the protein structure and highlights flexible regions that could affect ligand binding. RMSD (Root Mean Square Deviation) calculations were used to assess conformational differences between the native and mutated protein structures.

7. Protein Interaction Network Analysis

Protein-protein interaction networks were generated using the STRING database. The network analysis helped identify interacting proteins involved in neurodegenerative signalling pathways and neuronal survival mechanisms.

8. Molecular Docking Analysis

Structure-based virtual screening was conducted using server Docking. The docking process included:

1. Detection of binding cavities
2. Docking of selected neuroprotective compounds
3. Calculation of binding affinity scores
4. Visualisation of ligand-protein interactions

Docking results were evaluated based on binding energy, hydrogen bonding interactions, and ligand orientation within the binding pocket.

9. Data Integration

All computational results from NGS mutation mapping, structural validation, functional domain prediction, and docking simulations were integrated to identify potential neuroprotective targets and therapeutic candidates.

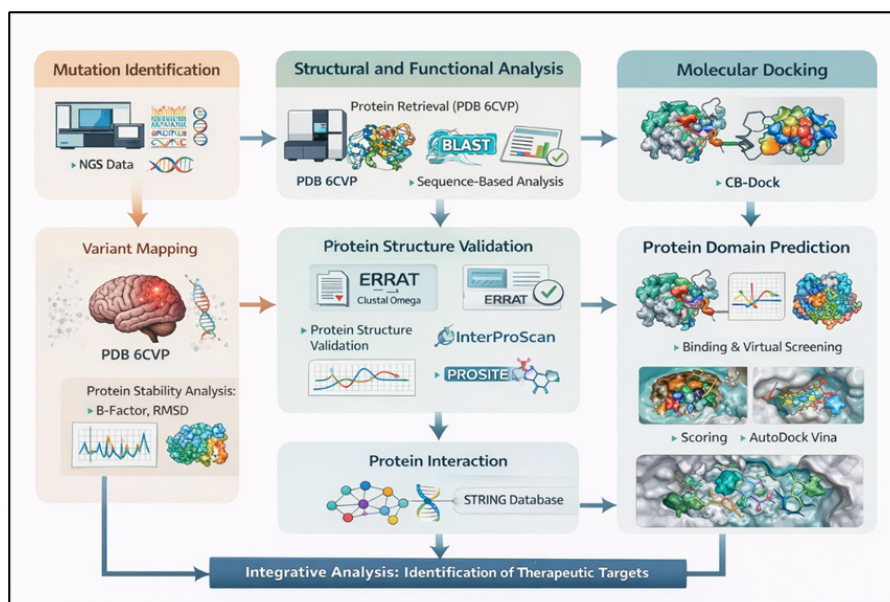


Figure 2 Integrative Analysis of Sample 6CVP

III. Results

Proteomic Sample

6CVP | pdb_00006cvp

Human Aprataxin (Aptx) R199H bound to RNA-DNA, AMP and Zn product complex

- Classification: HYDROLASE/DNA/RNA

- Organism(s): Homo sapiens

- Mutation(s): Yes

PDB: 6CVP_A

>pdb|6CVP|A Chain A, Aprataxin

GSHMGHWSQGLKISMQDPKMQVYKDEQVVVIKDKYPKAHYHWLVLPWTSISLKA VAREHLELLKH
 MHTVGEKVIVDFAGSSKLRFRLLGYHAIPSMHVHLHVISQDFDPSCLKNKKHWNSFNTEYFLESQAVIE
 MVQEAGRVTVRDGMPELLKPLRCHECQQLLPSIPQLKEHLRKHWTQ

ERRAT

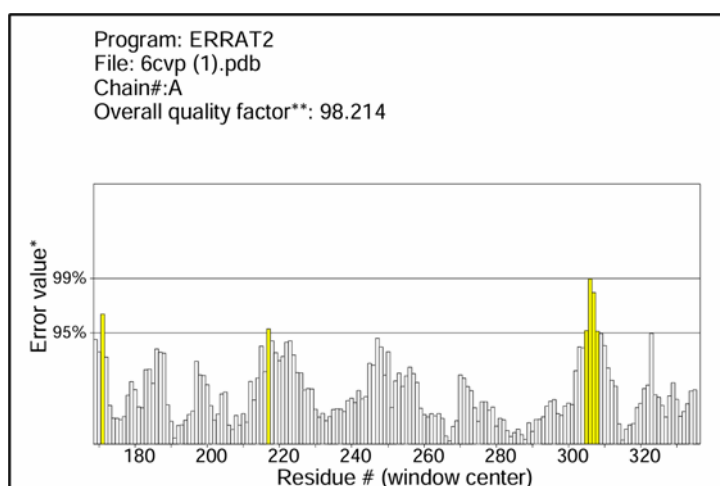


Figure 3: - Structure validation of sample 6CVP, which shows an overall quality factor 98.21%

Root means square Deviation analysis

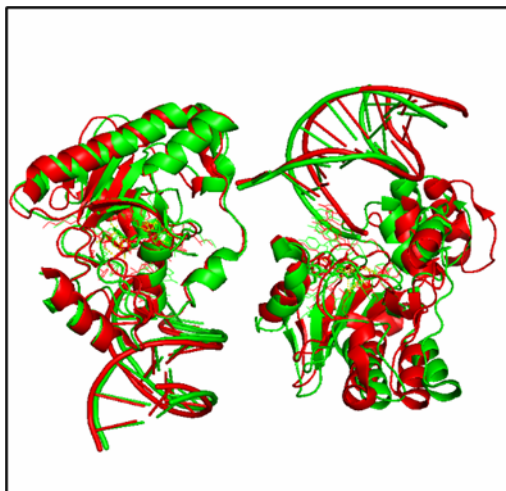


Figure 4: - RMSD result with pymol red colour shows sample 6cvp and green colour shows sample 6cvt score 2.910

Interpretation: An RMSD score of 2.910 typically indicates moderate structural similarity between two molecules, suggesting a protein model.

InterPro annotation Tools

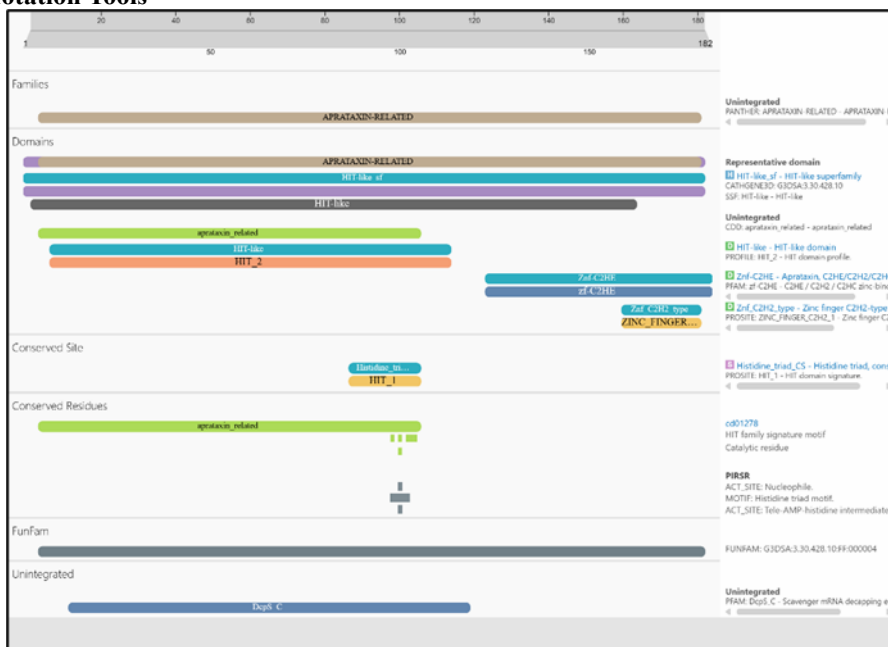


Figure 5: - Interproscan result of sample 6cvp

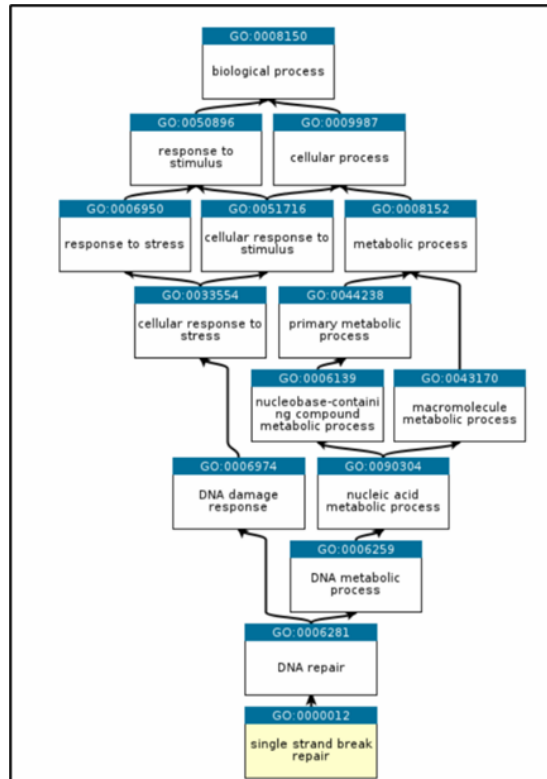


Figure 6: Ancestral chart analysis for single-strand break repair mediated by the enzyme system

Interpretation: In this protein signature analysis, protein sample 6CVP shares a common ancestor, exhibits similar function, and shows sequence similarities with certain protein families and domain residues, such as the APRATAXIN-related domain (residues 5-179), the HIT-like domain (3-162), the HIT-2 domain profile (8-113), the C2HE/C2H2/C2HC zinc-binding region (residues 123-182), and the zinc finger C2H2-type domain signature (159-179). The sample also displays conserved sites like HIT-1 (residues 87-105), with conserved residues at positions 98, 100, and 103. An ACT_SITE nucleophile is found at residue 100, along with the His triad motif (residues 98-102). Additionally, an unintegrated residue, the scavenger mRNA decapping enzyme C-terminal residue, is present between residues 13-118.

Protein motif database

Hits for all PROSITE (release 2025_02) motifs on
1 FASTA sequence(s): pdb-6CVP-A
 Note: Scan performed locally.
 Found: 3 hits in 1 sequence

PDB-6CVP-A (182 aa)

```

GSHMGHWLGGGLKISMGDPKMQVYKDEQVVVTKDKYKPAHYHLLVLPKITSISLKVAREHLELLKH
HHTVGERVVDVFGSSKLFRLGYPHALPSKSHNLHVLSQDFDPSPLKNKHKHWSFNTEYFLESQA
VIEHWQAGRVTVRDGPPELLKPLRCHECQQLLPSIPQLKEHLRHWIQ
    
```

Legend:
 [] disulfide bridge • active site — other 'ranges' + other sites

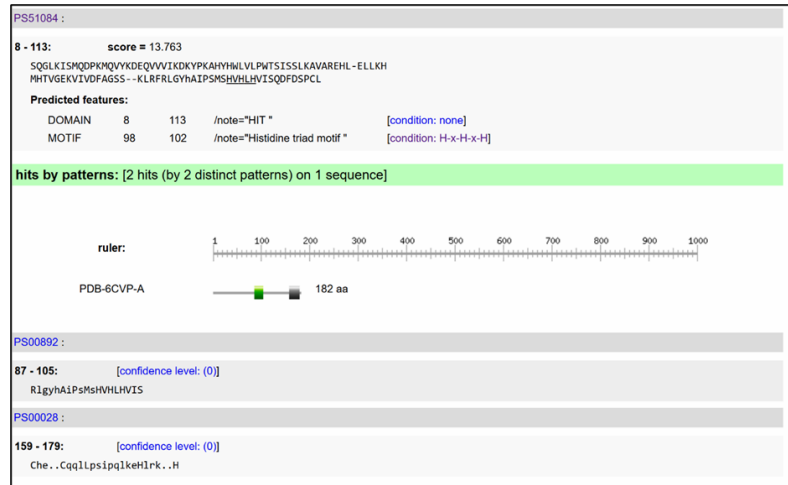
Please note that the graphical representations of domains displayed hereafter are for illustrative purposes only, and that their colors and shapes are not intended to indicate homology or shared function.
 For more information about how these graphical representations are constructed, go to <https://prosite.expasy.org/mydomains/>.

hits by profiles: [1 hit (by 1 profile) on 1 sequence]

Upper case represents match positions, lower case insert positions, and the '-' symbol represents deletions relative to the matching profile.

ruler: 1 100 200 300 400 500 600 700 800 900 1000

PDB-6CVP-A HIT_2 182 aa



Interpretation: In Protein Signature Database, the yellow highlighted sequence marks a region matching a known pattern, representing a conserved domain or signature with established biological functions. In the computational analysis of NGS data, the Prosite scan of 6CVP involved 199 histidine-mutant sequences, each 182 amino acids long, resulting in 2 detected hits corresponding to two different patterns in the first sequence. HIT1, covering residues 8-113, is labeled as the histidine triad domain with a score of 13.763, indicating the strength of the match. Conditions show no match to the standard pattern. HIT2, from residues 98-102, represents the histidine triad motif with a confidence level of zero, reflecting a standard notation in Prosite. The green bar on the graphical ruler marks the HIT2 domain. Histidine does not directly influence DNA but impacts its function by binding metal ions like Zn or Cu, which can lead to aggregation, as observed in neural disorders.

Protein Enrichment Analysis

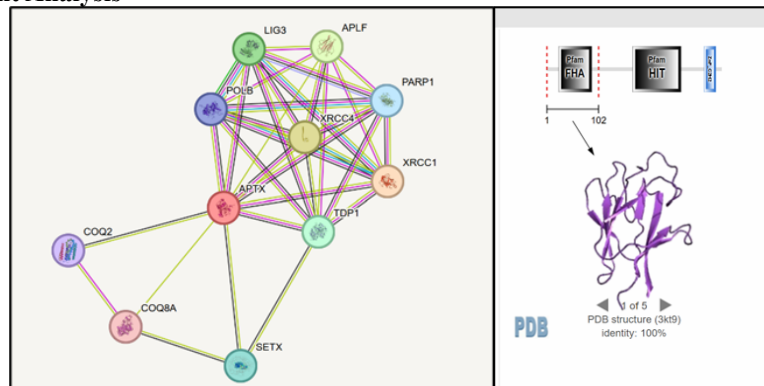


Figure 7: - string result of sample 6cvp with 100% similarity of PDB sample 3KT9

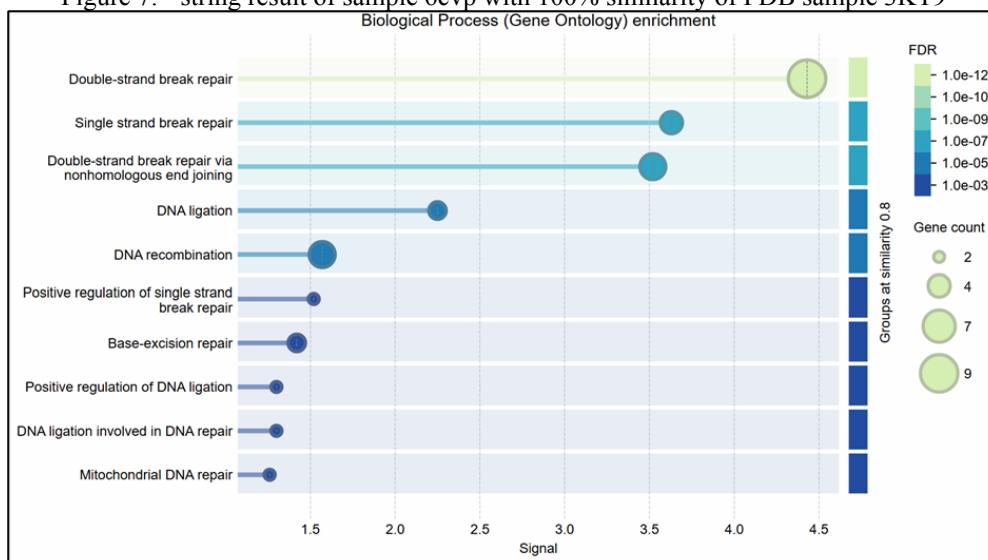


Figure 8: - biological process enrichment graph of sample 6CVP

Interpretation: Dark blue and light blue colours typically indicate levels of statistical significance or gene interaction in bar graphs. Dark blue signifies higher gene counts or stronger significance, whereas light blue indicates lower significance and smaller gene counts. In this graph, the 'Immune response' has a high gene count. In gene ontology analysis, a gene count of 7 is associated with the apoptotic process, while counts of 4 and 9 correspond to different gene ontology functions.

Atomic Displacement Parameter (ADP)

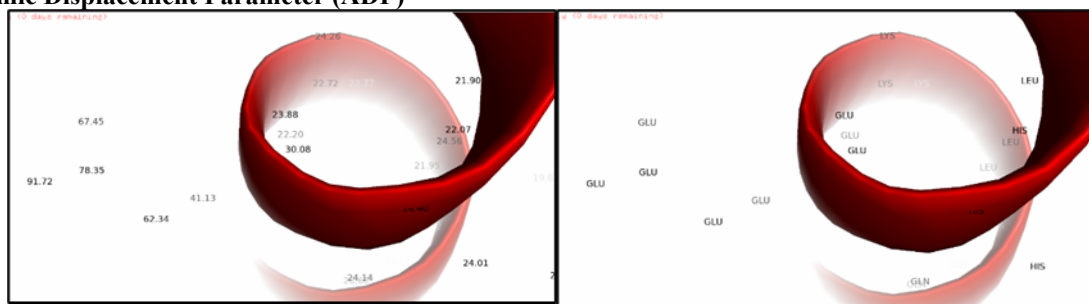


Figure 9: - B-factor analysis of sample 6CVP red colour shows score 91.72 GLU at position 333

Analysis of B-factor shows that in domain A, residue E (333) — a GLU — has a score of 91.72 in the loop region, indicating increased motion disorder.

SAMPLE 6CVP

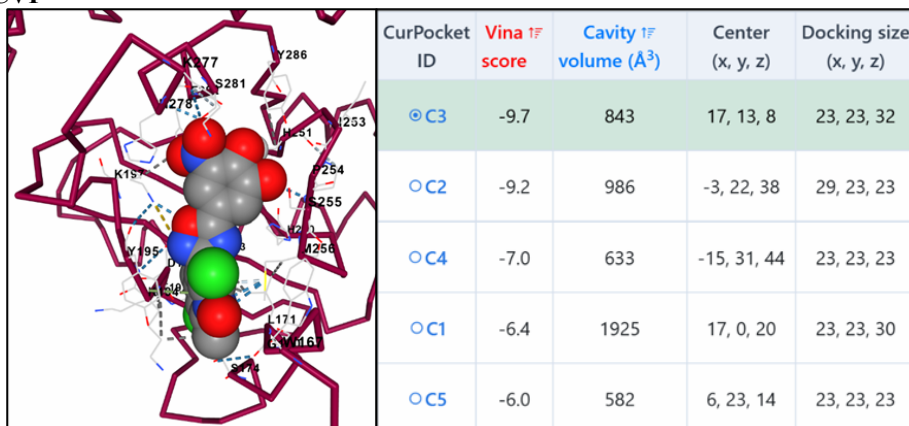


Figure 10: CB dock result of sample 6CVP and Opicapone.

Figure 11: Cur pocket result of sample 6CVP and Opicapone

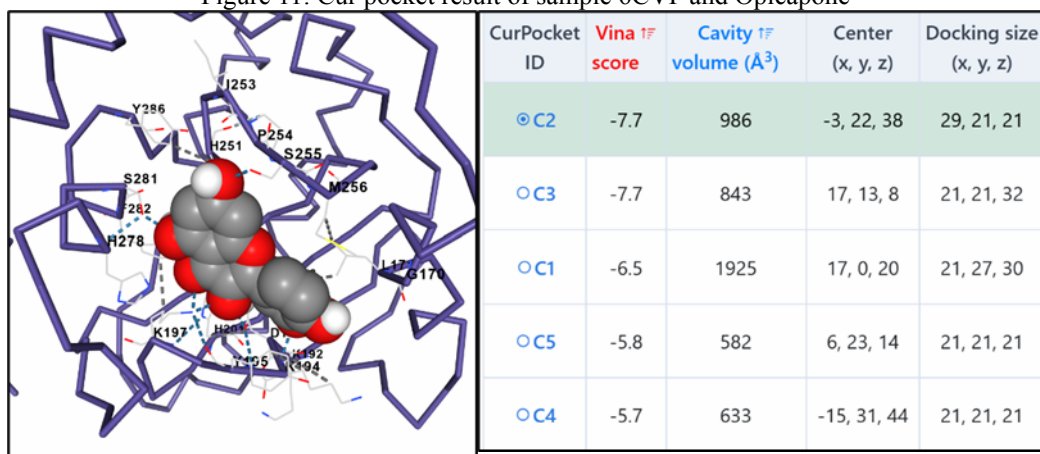


Figure 12: CB dock result of sample 6CVP and Quercetin.

Figure 13: Cur pocket result of sample 6CVP and Quercetin

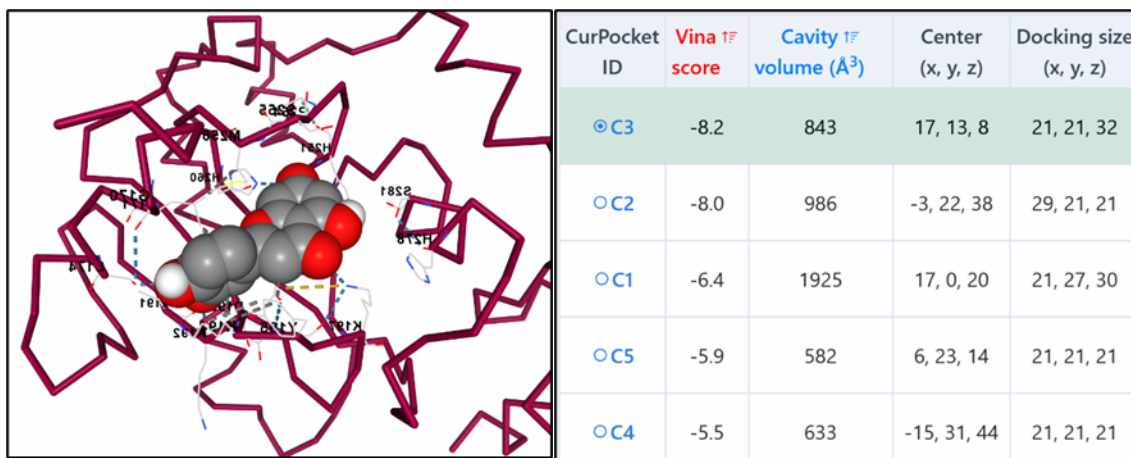


Figure 14: CB dock result of sample 6CVP and Luteolin. Figure 15: Cur pocket of sample 6CVP and Luteolin

Table 1: Interpretation of Docking Result

Sr.No.	Drug name	PDB ID	Docking score and cur pocket id	Remarks
1	Opicapone	6CVP	-9.7 & C3	Samples indicate a strong to the strongest binding affinity.
2	Quercetin	6CVP	-7.7 & C2	Samples indicate a moderate-to-strong binding affinity.
3	Luteolin	6CVP	-8.2 & C3	Samples indicate a strong binding affinity.

IV. Discussion

Analysis of the protein with PDB ID 6CVP revealed a well-defined 3D structure suitable for computational studies. Validation through the ERRAT server indicated a high-quality model with acceptable scores for non-bonded interactions, supporting its reliability for further research. Domain analysis using InterProScan and PROSITE identified conserved motifs associated with stability and molecular interactions, which are essential for maintaining structure and function. Mutations from NGS datasets mapped onto these domains showed several variants within conserved regions, suggesting possible disruptions in protein activity. [14] Flexibility analysis highlighted multiple regions with high B-factor values, indicating dynamic areas that could influence ligand binding and interactions. RMSD comparisons between native and mutant models showed significant structural deviations, implying important conformational changes caused by some mutations. [15,16] Interaction network analysis with STRING identified partners involved in neuronal signaling pathways, pointing to the protein's potential role in neuroprotection and neuronal survival. Molecular docking with CB-Dock found compounds with strong binding affinities, forming stable interactions via hydrogen bonds and hydrophobic contacts within the binding sites. Overall, this comprehensive computational approach indicates that mutations may impact protein stability, interaction networks, and ligand binding, emphasizing the importance of integrating genomic, structural, and proteomic data to better understand neurodegenerative diseases mechanisms. [17,18]

V. Conclusion

This research highlights the value of integrating genomic, proteomic, and structural bioinformatics approaches to investigate molecular mechanisms in neurodegenerative diseases. By combining mutation data from next-generation sequencing with structural studies and molecular docking, we gain detailed insights into protein dysfunction and potential therapies. The protein structure from PDB 6CVP was a central model for analysing mutation effects; ERRAT validation confirmed its accuracy, while tools like InterProScan and PROSITE identified crucial conserved motifs. Mapping variants onto the structure pinpointed mutations in key regions that could impair stability and function. Flexibility was assessed using B-factor values and RMSD, revealing conformational shifts that affect interactions and dynamics. Protein interaction analyses using STRING demonstrated the protein's involvement in neural signalling and cellular homeostasis, with mutations potentially driving disease progression. Drug discovery through CB-Dock identified compounds that could bind active sites, potentially stabilizing or modulating protein activity for neuroprotection. Overall, this comprehensive computational strategy emphasises the importance of integrating genomics, proteomics, and structural bioinformatics to advance understanding and identify therapeutic targets in neurodegenerative diseases.

VI. Future Scope

Future research should aim to expand the use of the integrative multi-omics approach from this study to include transcriptomics, metabolomics, and epigenomics data. These extensive datasets can provide deeper insights into the molecular mechanisms underlying neurodegenerative diseases. Furthermore, validating computational predictions with experimental methods is essential. Performing in vitro and in vivo studies will confirm the functional effects of mutations identified by NGS and evaluate the biological activity of potential therapeutic compounds identified via molecular docking. Advanced molecular dynamics simulations can also improve understanding of protein stability and ligand interactions, helping to analyse long-term structural behaviour and identify stable drug-target complexes. Applying machine learning and AI techniques could enhance predictive models of mutation effects and the progress of drug discovery. Combining AI with structural bioinformatics has the potential to accelerate the development of personalized therapies for neurodegenerative disorders. Ultimately, integrating genomic data, proteomics, and computational drug discovery promises to transform neurodegenerative disease research, enabling early diagnosis, targeted treatments, and improved clinical outcomes.

References

- [1]. Liu, K., Riaz, B., & Zhong, S. (2025). Investigating The Therapeutic Potential Of The Ketogenic Diet In Modulating Neurodegenerative Pathophysiology: An Interdisciplinary Approach. *Nutrients*, 17(7), 1268.
- [2]. Chiti, F., & Dobson, C. M. (2017). Protein Misfolding, Amyloid Formation, And Human Disease: A Summary Of Progress Over The Last Decade. *Annual Review Of Biochemistry*, 86, 27–68. <https://doi.org/10.1146/annurev-biochem-061516-045115>
- [3]. 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)
- [4]. Uma Kumari, Vineeta Johri "Integrating NGS And CADD For The Identification Of Medicinal Plant Compounds Targeting EGFR Mutants In Non-Small Cell Lung Cancer", *International Journal Of Emerging Technologies And Innovative Research*, Vol. 12, Issue 8, Page No. Ppf365-F378, August-2025, DOI: <https://doi.org/10.56975/jetir.V12i8.568698>
- [5]. Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Von Mering, C. (2019). STRING V11: Protein–Protein Association Networks With Increased Coverage. *Nucleic Acids Research*, 47(D1), D607–D613. <https://doi.org/10.1093/nar/gky1131>
- [6]. Liu, Y., Grimm, M., Dai, W. T., Hou, M. C., Xiao, Z. X., & Cao, Y. (2020). CB-Dock: A Web Server For Cavity Detection-Guided Protein–Ligand Blind Docking. *Acta Pharmacologica Sinica*, 41(1), 138–144. <https://doi.org/10.1038/s41401-019-0228-6>
- [7]. Blum, M., Chang, H. Y., Chuguransky, S., Grego, T., Kandasamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G. A., Williams, L., Bork, P., Bridge, A., Gough, J., Haft, D., Letunic, I., Marchler-Bauer, A., ... Finn, R. D. (2021). The InterPro Protein Families And Domains Database: 20 Years On. *Nucleic Acids Research*, 49(D1), D344–D354. <https://doi.org/10.1093/nar/gkaa977>
- [8]. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S., Ballard, A., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly Accurate Protein Structure Prediction With Alphafold. *Nature*, 596(7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- [9]. Baek, M., Dimairo, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Millán, C., Park, H., Adams, C., Glassman, C. R., Degiovanni, A., Pereira, J. H., Rodrigues, A. V., Van Dijk, A. A., Ebrecht, A. C., ... Baker, D. (2021). Accurate Prediction Of Protein Structures And Interactions Using A Three-Track Neural Network. *Science*, 373(6557), 871–876. <https://doi.org/10.1126/science.abj8754>
- [10]. Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Židek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., ... Velankar, S. (2022). Alphafold Protein Structure Database: Massively Expanding The Structural Coverage Of Protein-Sequence Space. *Nucleic Acids Research*, 50(D1), D439–D444. <https://doi.org/10.1093/nar/gkab1061>
- [11]. Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., ... Bouatra, S. (2022). Drugbank 5.0: A Major Update To The Drugbank Database. *Nucleic Acids Research*, 50(D1), D1268–D1276. <https://doi.org/10.1093/nar/gkab1068>
- [12]. Aebersold, R., & Mann, M. (2023). Mass-Spectrometric Exploration Of Proteome Structure And Function. *Nature*, 613(7943), 257–267. <https://doi.org/10.1038/s41586-022-05564-6>
- [13]. Johri, V., Tyagi, E., & Kumari, U. (2024). Next Generation Sequencing Analysis Of NSP5, NSP6 And NSP7 SARS-Cov-2 Nonstructural Protein With Biopython. *SEEJPH*, Volume XXV(S2).
- [14]. Kumari, U., Johri, V., Sharma, P., & Kumar, D. (2026). NGS Mutation Profiling Of TP53 Gene Analysis In Oral Squamous Cell Carcinoma (OSCC). *International Journal Of Drug Delivery Technology*, 16(14s), 403–415. <https://doi.org/10.25258/ijddt.16.14s.42>
- [15]. Mandra, D., Kumari, U., Singh, N., & Marouthu, A. (2026). Integrated Multi-Omics And Biopython Approaches For Target Identification Of A-Synuclein Polymorph In Parkinson's Disease. *IOSR Journal Of Biotechnology And Biochemistry*, 12(2), 12–21. <https://doi.org/10.9790/264X-1202011221>
- [16]. Kumari, U., & Divakar, S. (2026). Integrative NGS-Based Genomic Profiling And Structure-Guided Toxicity Prediction Of KRAS G12C Inhibition. *International Journal Of Emerging Technologies And Innovative Research*, 13(3), H177–H191. <https://doi.org/10.56975/jetir.V13i3.578071>
- [17]. Kumari, U., & Singh, N. (2026). Multi-Omics-Based Drug Discovery Pipelines Incorporating Machine Learning, AI, And Bio Python In Parkinson's Disease: A Case Study Of 5A Polymorph Of Alpha-Synuclein. *International Journal Of Emerging Technologies And Innovative Research*, 13(2), B617–B634.
- [18]. Kumari, U., Koundal, S., & Katal, S. (2025). Integrated Structural And Network-Based Characterization Of ABL1 As A Central Hub In Leukaemia, Using Protein Interaction And Motif Analysis With Biopython. *International Journal Of Emerging Technologies And Innovative Research*, 12(12), G355–G366. <https://doi.org/10.56975/jetir.V12i12.573916>

- [19]. Kumari, U., Mandra, D., & Sonali. (2025). Computational Analysis Of PD-L1 Structural Variants And Expression Profiles Using Next-Generation Sequencing For Predicting Immunotherapy Response In Non-Small Cell Lung Cancer (NSCLC). *IOSR Journal Of Biotechnology And Biochemistry*, 11(6), 72–79