Homology Modeling and Drug Designing Approach for Prospective Malignant Brain Cancer

Shubhi Bindal[1], Uma Kumari[2], Pinki Raj Sahu[3]
[1]*Project trainee at BPRI*
[2]*Senior Bioinformatics Scientist, BPRI, Noida*
[3]*Assistant Professor Marwari College Ranchi*

***Abstract*—Brain tumors are a heterogeneous group of malignancies that begins in cells of the central nervous system. A variety of models in biological sample for preclinical studies have been developed to identified mutation in human brain tumor using bioinformatics software allowing us to understand the underlying pathobiology and explore potential treatments. Many promising therapeutic strategies identified using preclinical models have shown limited efficacy or failed at the clinical trial stage, so drug designing started. Therapeutic strategies develop to significantly improve survival rates in patients highlight the compelling to revisit the design of currently available animal models and explore about how new models that allow us to bridge the gap between promising preclinical findings and identification of clinical translation. In review the current strategies used to model glioblastoma, the most malignant brain tumor in adults and highlight the shortcomings of specific models create using biological sample that must be development of innovative therapeutic strategies. Data analysis computational brain tumor models have the potential to provide experimental tumor biologists and cost-efficient tools to generate test hypotheses on tumor progression and fundamental operating principles governing bidirectional signal propagation in multicellular cancer systems. By using software and databases highlights the modeling objectives of and challenges with developing such in-silico brain tumor models to identify target selection.**

***Keywords* -Brain tumor model, Drug designing, Glioblastoma, Glioma, Neuro-oncology, Sequence analysis, and Therapeutic development**.

I.INTRODUCTION

Brain cancer, a commanding type of cancer that causes death in both children and grown-ups, was diagnosed in about 300,000 new cases and caused 241,000 deaths encyclopaedically in 2018(Bray F, Ferlay J, et al, 2018). More lately, mortality numbers of brain and other nervous system cancers in the United States caused an estimated 23,890 deaths in 2020 (12,590 men and 10,300 women) (Siegel RL, Miller KD, 2020). As a miscellaneous disease, unrestrained cell growth in brain cancer has complex molecular mechanisms, which may be caused by promoter methylation, deregulated gene expression, and/or genetically altered tumour-suppressor genes and oncogenes(Binder H, Willscher E, et al,2019). Glioblastoma(GBM) is the most fatal and frequent primary malignant brain tumour. Cases diagnosed with the disease and treated with state-of-the-art remedies(uttermost safe surgery, radiotherapy and chemotherapy) have a median survival of only around 14 months(Stupp Retal. 2005). The 5-year survival rate for cases with glioblastoma(also known as glioblastoma multiforme, or GBM) is only 5.4%, and the 10-year survival rate is only 2.7%(Gittleman H, Boscia A, et al, 2018). These tumours are located behind the blood–brain barrier(BBB) — a system of tight junctions and transport proteins that safeguard delicate neural tissues from exposure to factors in the general circulation, therefore also impeding exposure to systemic chemotherapy(Phoenix TN, etal. 2016, Gerstner ER, Fine RL 2007). Historically, the drug discovery process has depended on experimental high-throughput screening (HTS) to identify biologically active combinations (Congreve,M. etal. 2005). Despite advances in mechanization methodologies for HTS this approach remains extremely laborious, costly and has repeatedly failed to identify potent lead series (Kubinyi,H. 2001, Liu,B. etal. 2004). correlative in silico methodologies like structure-based drug design(SBDD), incorporate the knowledge from high-resolution 3D protein structures to probe structure – function relationships(Hillisch,A. etal. 2004), identify and opt therapeutically applicable targets(assess druggability), study the molecular base of ligand protein interactions(Blundell,T.L. etal. 2006),

characterize binding pockets (Halgren,T.A. 2009), develop target-specific compound libraries(Orry,A.J. etal. 2006, Schnur,D.M. 2008), identify hits by high-throughput docking(HTD)(Cavasotto,C.N. and Orry,A.J. 2007, Cavasotto,C.N. and Singh,N. 2008), and optimize lead compounds(Lundstrom,K. 2007), all of which can be used to attribute, increase effectiveness, speed and cost- effectiveness of the drug discovery process(Anderson,A.C. 2003, Weigelt,J. etal. 2008). Likewise, in this research, we employed homology modelling and computer-aided drug design methodologies to identify a molecule with high affinity for the target 1OZ3 that can potentially serve as a medication in future.

II. MATERIALS AND METHODS

*A. Target Selection:*The literature at first studied, during this analysis to decide on the acceptable target. Following that, 1OZ3 was chosen as the target protein, and its data was derived from PDB (Protein Data Bank). Protein Data Bank archive-information regarding the 3D shapes of proteins, nucleic acids, and complicated assemblies that help students and researchers perceive all aspects of biomedicine and agriculture, from macromolecule synthesis to health and disease. (https://www.rcsb.org/)

*B. Homology Modelling:*The structure of 1OZ3 was derived from MMDB (Molecular Modeling Database). Then, it is studied in RasMOL where we visualized molecular graphics in different type of representation. Further, we run the target protein as a query in BLASTp which uses statistical theory to produce a bit score and expect value (E-value) for each alignment pair (query to hit). Then, for Multiple Sequence Analysis we used COBALT (Constraint-based Multiple Alignment Tool) that finds a collection of pairwise constraints derived from conserved domain database, protein motif database, and sequence similarity, using RPS-BLAST, BLASTP, and PHI-BLAST.

C. Validation of 3D-model: The structure that is studied through homology modeling generally needs to be validated and remedied before docking procedure. The reliability and stability of the model were validated by Ramachandran plot generation, that was studied in PDBsum, which is a pictorial database that provides an at-a-glance summary of the contents of each 3D structure deposited within the Protein Data Bank (PDB).

III. RESULTS

A.Homology Modelling:

BLASTp finds regions of similarity between biological sequences. It compares protein sequences to sequence databases and calculates the applied mathematics. Lower the E-value, or the nearer it is to zero, the more "significant" the match. Here different protein of different species shows the significant similarity withthe query sequence

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
lethal(3)malignant brain tumor-like protein 1 isoform X5 [Pan troglodytes]	Pan troglodytes	689	689	100%	0.0	98.49%	842	XP_016793431.1
lethal(3)malignant brain tumor-like protein 1 isoform X3 [Pan troglodytes]	Pan troglodytes	689	689	100%	0.0	98.49%	864	XP_016793429.1
lethal(3)malignant brain tumor-like protein 1 isoform 4 [Homo sapiens]	Homo sapiens	689	689	100%	0.0	98.49%	821	NP_001364235.1
lethal(3)malignant brain tumor-like protein 1 [Pan paniscus]	Pan paniscus	689	689	100%	0.0	98.49%	842	XP_008949541.2
lethal(3)malignant brain tumor-like protein 1 isoform X1 [Pan troglodytes]	Pan troglodytes	689	689	100%	0.0	98.49%	885	XP_016793426.1
lethal(3)malignant brain tumor-like protein 1 isoform X2 [Pan troglodytes]	Pan troglodytes	689	689	100%	0.0	98.49%	873	XP_016793428.1
lethal(3)malignant brain tumor-like protein 1 isoform 3 [Homo sapiens]	Homo sapiens	689	689	100%	0.0	98.49%	862	NP_001364232.1

Figure 1. represents the similarity sequences with the query sequence. Here different organism like *Pan troglodytes*, *Pan paniscus* and different protein of *Homo sapiens* shows significant similarity with the query protein i.e. lethal malignant brain tumour-like protein in *Homo sapiens.*

Now for multiple sequence alignment we used COBALT specifically BLOSUM method to analyse the multiple sequence similarity. BLOSUM method uses well-known substitution matrices to display the degree of match of residues relative to each alignment position/ column. When an anchor row is set, the colouring in the column shows the match score to the residue on the anchor sequence.

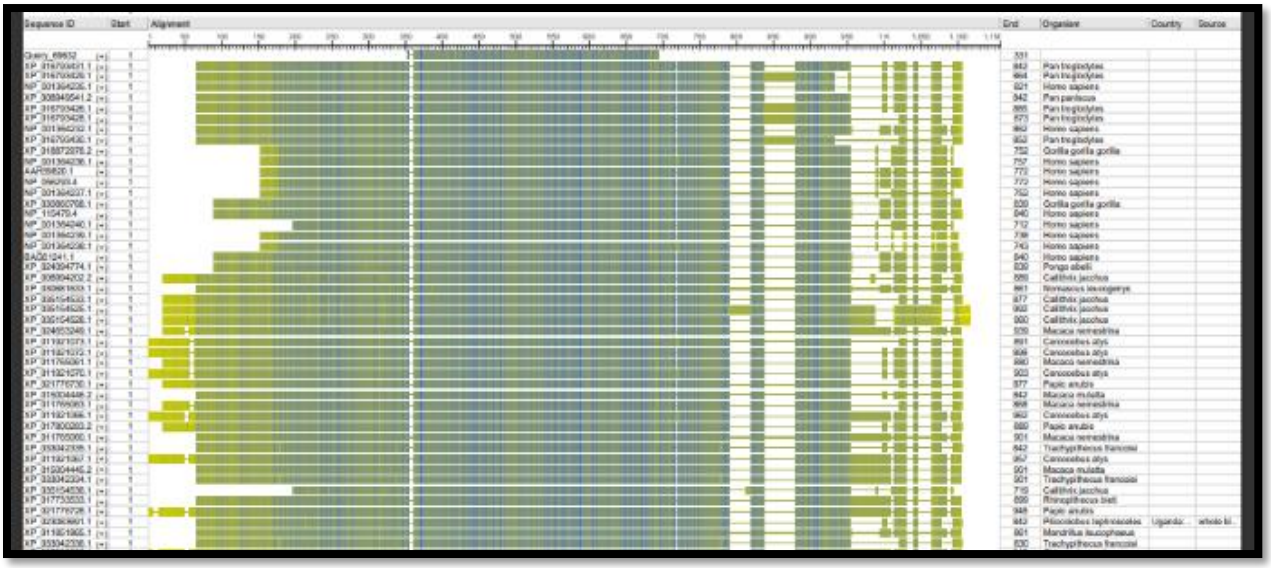


Figure 2. represents the multiple sequence alignment where Blue colour represents better match and green colour represents worse match. In this figure accession number, sequence derived from organism their country origin and derived source is mentioned as well.

B. Validation of 3D model:

Further we analysed the protein structure through Ramachandran Plot in PDBsum.

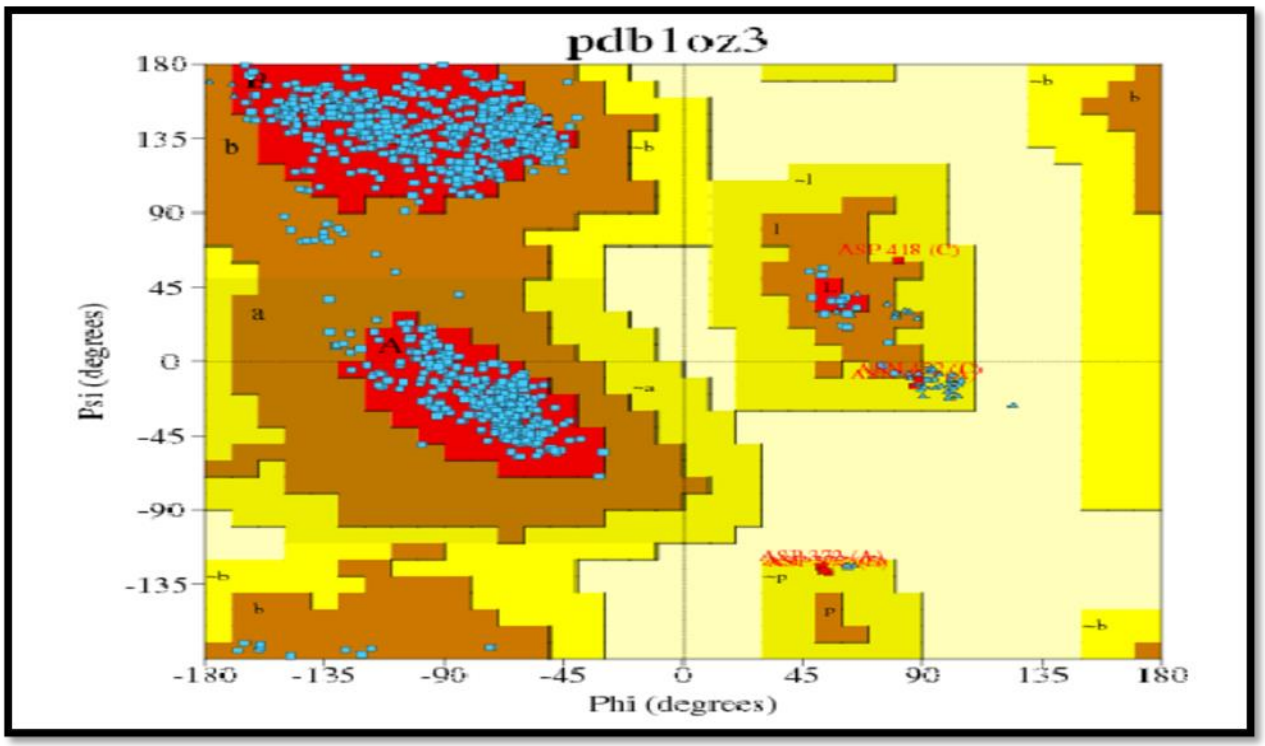


Figure 3.Ramachandran Plot

Table.I. Analysis of the Ramachandran plot

	No. of residues	%- tage
Most favoured regions [A, B, L]	721	90.7%
Additional allowed regions[a, b, l, p]	68	8.6%
Generously allowed regions [~a, ~b, ~l, ~p]	6	0.8%
Disallowed regions	0	0.0%
Non-glycine and non-proline residues	795	100.0%
End residues (excluding Gly and Pro)	6	
Glycine residues	51	
Proline residues	87	
Total no. of residues	939	

C. Pathway analysis of Glioma:

GBM might develop de novo (primary glioblastoma) or by progression from low- grade or anaplastic astrocytoma (secondary glioblastoma).

Primary glioblastomas develop in older patients and typically show genetic alterations [EGFR (Epidermal Growth Factor) amplifications, p16/INK4a (Cyclin dependent kinase inhibitor 2A) deletion, and PTEN (Phosphatase and Tensin homolog) mutations] at frequencies of 24-34%.

Secondary glioblastomas develop in younger patients and frequently show overexpression of PDGF (Platelets-Derived Growth Factor) and CDK4 (Cyclin Dependent Kinase 4) as well as p53 (tumour suppressor gene) mutations (65%) and loss of Rb (Retinoblastoma-associated protein) playing major roles in such transformations. Loss of PTEN has been implicated in both pathways, although it is much more common in the pathogenesis of primary GBM. (source:https://www.genome.jp/pathway/hsa05214).

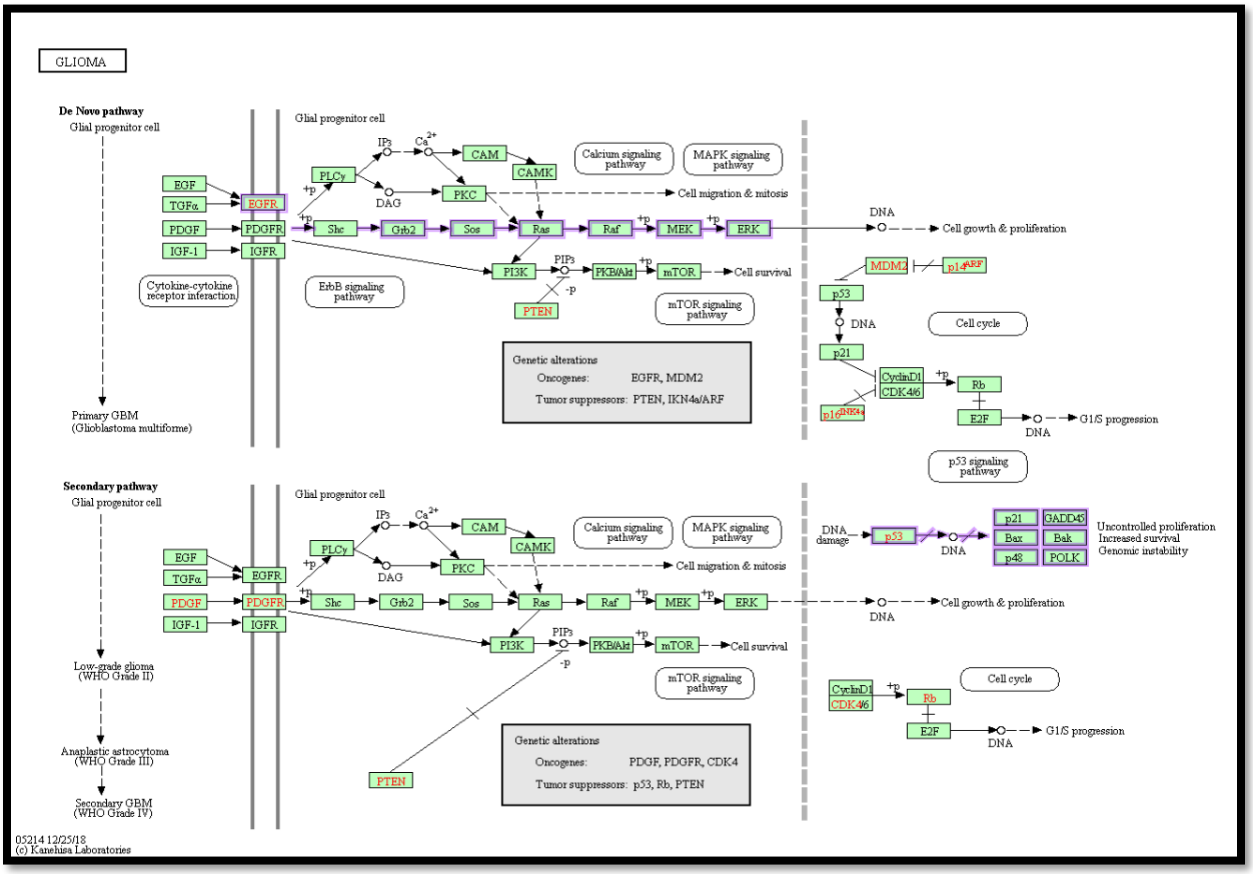


Figure 4. Glioma Pathway

D. Molecular Docking :

After analysing the target protein 1OZ3, it was further visualized in PyMOL and studied according to the density representation (volume rendering visualization) which can be seen as Figure.5.

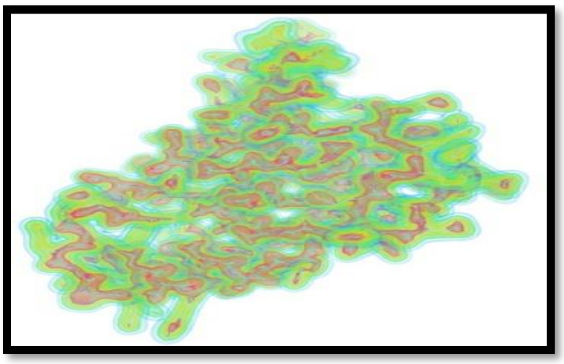


Figure 5. Represents high-density regions (protein) are denoted as opaque red and low-density region (protein) are denoted as transparent blue.

and six ligand viz. are Dacomitinib, Betulinic acid, Selinexor, Paclitaxel C, Etoposide, Larotrectinib Sulfate studied in PubChem. PyMOL was used to prepare the ligands partially (SDF to PDB). After partial preparation of ligands PyMOL was used to take a broad idea of ligand binding to the target site.

After that, to analyse binding affinity we used AutoDock tools where firstly, we convert PDB files into PDBQT of both target protein and ligands. For target protein to convert into PDBQT file all water molecules were removed and only polar hydrogen and Kollman Charges were added. The grid was developed to ensure where the ligands may potentially bind. The dimension of grid box were set as x centre- -5.722, Y centre- -9.417, z centre- -7.806. Both the target and ligands were permitted to dock using AutoDock 4.2. The finding were concluded once after the completion of docking.

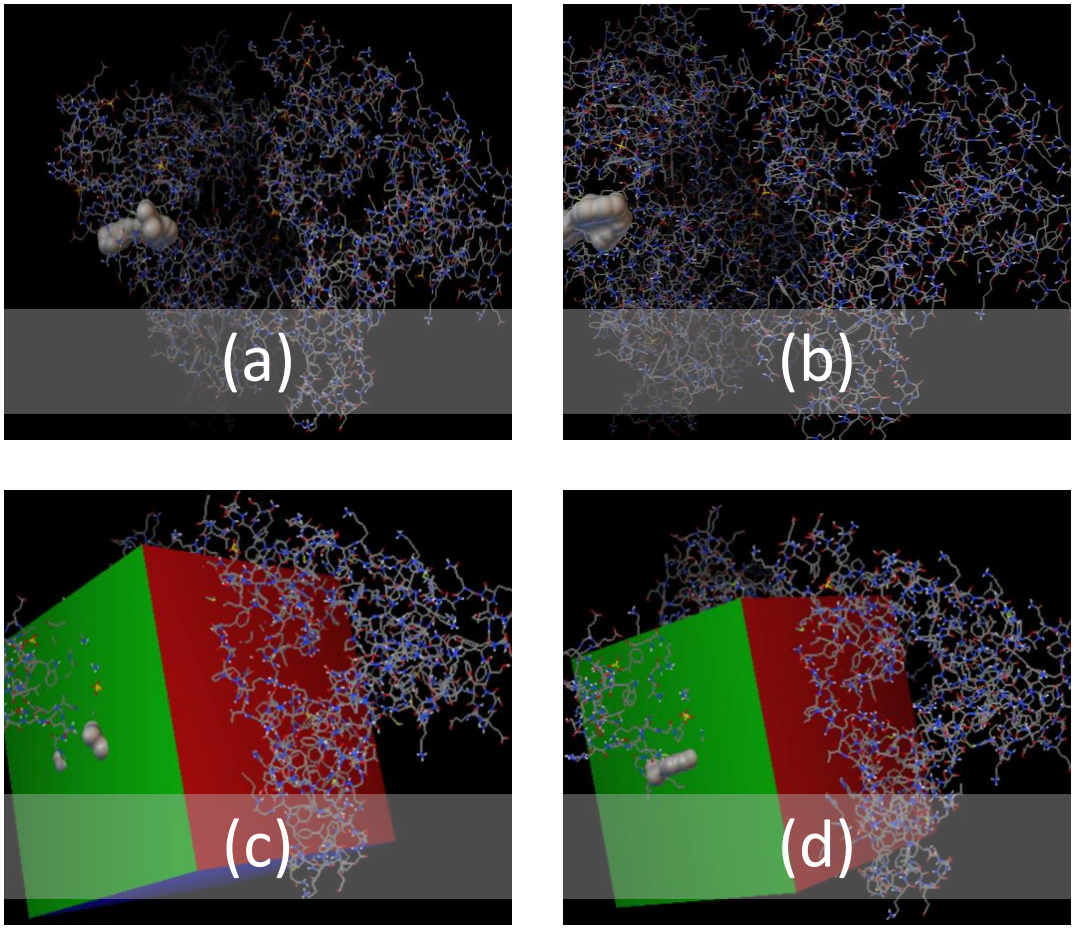


Figure 6. Shows molecular docking. Here, (a)shows molecular docking of protein and ligand (Dacomitinib). (b)shows molecular docking of protein and ligand (Selinexor). (c)Grid box showing blind docking of Dacomitinib to the target site.(d)Grid box showing blind docking of Selinexor to the target site.

Table.II. Shows the resolution for X-RAY crystallographic results.

Template	Compound name	Method	Resolution
1OZ3	Selinexor	X- RAY	0.375Å
1OZ3	Betulinic acid	X- RAY	0.397Å
1OZ3	Larotrectinib sulfate	X- RAY	0.403Å
1OZ3	Etoposide	X- RAY	0.414Å
1OZ3	Paclitaxel C	X- RAY	0.425Å
1OZ3	Dacomitinib	X- RAY	0.458Å

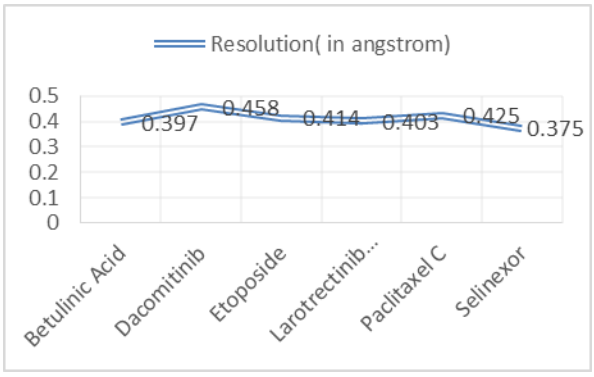
(Protein with resolution < 2 Å are good for docking.)

High numeric values of resolution, such as 4 Å, mean poor resolution, while low numeric values, such as 1.5 Å, mean good resolution. 2.05 Å is the median resolution for x-ray crystallographic results in the PDB.

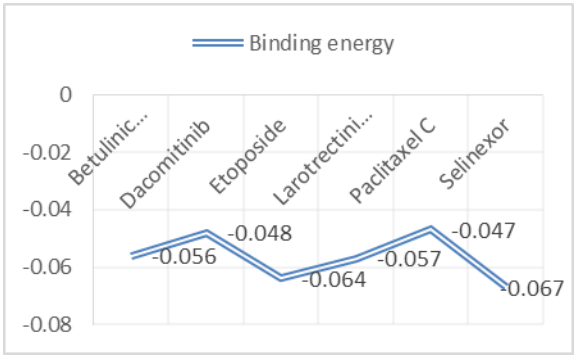
Table.III. Shows the binding energy of ligands

Ligand	Binding Energy
Betulinic Acid	-0.056
Dacomitinib	-0.048
Etoposide	-0.064
Larotrectinib Sulfate	-0.057
Paclitaxel C	-0.047
Selinexor	-0.067

Higher the binding energy, higher the binding affinity.



(a)



(b)

Graphical representation of (a) resolution and (b) binding energy of different ligands with the target site.

IV. CONCLUSIONS

It is critical to find high affinity drugs that bind to proteins and genes that cause malignant brain cancer. Our research results has established that few ligands like Dacomitinib, Selinexor, Betulinic Acid etc. are strong ligands that can bind to 1OZ3 and might possibly be employed as treatments for malignant brain cancer. Additionally, we have done Homology Modelling and sequence analysis with the use of RasMOL, MMDB, BLASTp, and COBALT. We studied Glioma pathway in KEGG Pathway Database and Ramachandran Plot in PDBsum. Further, with the use of PyMOL we analyze the binding of different ligands with the protein sample. We also learnt about different visualization form like volume rendering. Then, we analyzed the resolution and docking affinity of ligands and protein by using AutoDock 4. Additional analysis can be carried out to confirm their therapeutic potential.

ACKNOWLEDGEMENTS

Authors are greatful toBioinformatics Project and Research Institute, Noida, India for providing assistance to carried out this research work**.**

REFERENCES

[1] Anderson, A.C. (2003) The process of structure-based drug design. Chem. Biol. 10, 787–797.

[2] Binder H, Willscher E, Loeffler-Wirth H, Hopp L, Jones DTW, Pfister SM, et al. DNA methylation, transcriptome and genetic copy number signatures of diffuse cerebral WHO grade II/III gliomas resolve cancer heterogeneity and development. Acta NeuropatholCommun. 2019;7(1):59. https://doi.org/10.1186/s40478-019-0704-8.

[3] Blundell, T.L. et al. (2006) Structural biology and bioinformatics in drug design: opportunities and challenges for target identification and lead discovery. Philos. Trans. R. Soc. Lond. B: Biol. Sci. 361, 413–423.

[4] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394–424. https://doi.org/10.3322/caac.21492.

[5] Cavasotto, C.N. and Orry, A.J. (2007) Ligand docking and structure-based virtual screening in drug discovery. Curr. Top. Med. Chem. 7, 1006–1014

[6] Cavasotto, C.N. and Singh, N. (2008) Docking and high throughput docking: successes and the challenge of protein flexibility. Curr. Comput. Aid. Drug Discov. 4, 221–234

[7] Congreve, M. et al. (2005) Structural biology and drug discovery. Drug Discov. Today 10, 895–907.

[8] Gerstner ER, Fine RL. Increased permeability of the blood-brain barrier to chemotherapy in metastatic brain tumors: establishing a treatment paradigm. J. Clin. Oncol. 2007; 25:2306–2312.

[9] Gittleman H, Boscia A, Ostrom QT, Truitt G, Fritz Y, Kruchko C, et al. Survivorship in adults with malignant brain and other central nervous system tumor from 2000–2014. Neuro Oncol. 2018;20(suppl_7): vii6–vii16.

[10] Halgren, T.A. (2009) Identifying and characterizing binding sites and assessing druggability. J. Chem. Inf. Model. 49, 337–389.

[11] Hillisch, A. et al. (2004) Utility of homology models in the drug discovery process. Drug Discov. Today 9, 659–669.

[12] Kubinyi, H. (2001) HTS technologies – IBC informa conference. IDrugs 4, 168–173.

[13] Liu, B. et al. (2004) Technological advances in high-throughput screening. Am. J. Pharmacogenom. 4, 263–276.

[14] Lundstrom, K. (2007) Structural genomics and drug discovery. J. Cell Mol. Med. 11, 224–238.

[15] Morris GM, Goodsell DS, Halliday DS, Huey R, Hart WE, Belew R, Olson AJ (1998) Automated docking using a lamarckian genetic algorithm and and empirical binding free energy function. J Comp Chem 19:1639–1662.

[16] Orry, A.J. et al. (2006) Structure-based development of target-specific compound libraries. Drug Discov. Today 11, 261–266.

[17] Phoenix TN, et al. Medulloblastoma genotype dictates blood brain barrier phenotype. Cancer Cell. 2016; 29:508–522.

[18] Schnur, D.M. (2008) Recent trends in library design: 'rational design' revisited. Curr. Opin. Drug Discov. Dev. 11, 375–380.

[19] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020;70(1):7–30. https://doi.org/10.3322/caac.21590.

[20] Stupp Ret al.2005Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. N. Engl. J. Med. 352, 987-996.

[21] Kumari Uma and Choudhary, Ashok Kumar, Genome Sequence Analysis of SolanumLycopersicum by Applying Sequence Alignment Method to Determine the Statistical Significance of an Alignment (June 30, 2016). International Journal of Bio-Technology and Research (IJBTR) ISSN(P) :2249-6858; ISSN(E):2249-796X Vol. 6, Issue 3, Jun 2016, 9-12, Available at SSRN: https://ssrn.com/ abstract =2838750

[22] Kumari, U., &Choudhary, A. K. (2016). Genome sequence analysis of solanumlycopersicum showing the phylogenetic relationship based on multiple sequence alignment and conserved domain proteins. *INTERNATIONAL JOURNAL OF ADVANCED BIOTECHNOLOGY AND RESEARCH*, *7*(4), 2012-2014.

[23] Uma kumari ,NavjotKaurVirk, Identification of new potential drug for lung adenocarcinoma causing protein RMB10 using computer aided drug design approach, JournalIJBTR,Volume 12, Issue 2, Pages, 1-8, TJPRC Pvt.Ltd

[24] Uma Kumari "Insilico analysis and computer aided drug designing approach for mutant cancer gene (IJBTR) DEC 2021(Impact factor 6.6, ICV 61.5 ,NASS RATING 3.8)