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

INTEGRATING COMPUTER AIDED DRUG DESIGN AND NEXT-GENERATION SEQUENCING TO OPTIMIZE TARGETED THERAPIES FOR PI3KA H1047R MUTANT IN CANCER

Uma Kumari[1*], Diksha Gangurde[1], Aiswarya Nair[2], Nupur Phirke[2]

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

Abstract: Cancer is a genetic disease that arises from genetic or epigenetic mutations defined by the abnormal growth of cells in the body, and is also responsible for one of the highest burdens of mortality worldwide. The genes that contribute to cancer development can generally be divided into two groups, the oncogenes and the tumour suppressor genes. The H1047R mutation in the PI3Kα catalytic subunit (PIK3CA) is one of the most common oncogenic alterations seen in multiple cancers, promoting abnormal PI3K signalling and adding to therapeutic resistance. In this research we will pursue precision therapeutics targeting the cryo-EM structure of the cancer-specific PI3Kα H1047R mutant via computer-aided drug design (CADD) to aid structure-based virtual screening and optimization. The 3D structural analysis and visualization are performed using structural bioinformatics tools to identify binding interactions and conformational dynamics. Multiple sequence alignment will be conducted with Clustal Omega to consider the evolutionary conservation of the H1047 residue in PI3K homologs to help localize mutation hotspots and druggable pockets. In this structural and sequence context, web server docking, a cavity-detection-based docking tool, will be used for blind virtual screening of small-molecule libraries to identify new inhibitors with good binding affinity towards the target mutant protein. This integrative computational pipeline is fundamentally based on structural biology and molecular modelling, has the goal of finding and optimizing novel small-molecule inhibitors with improved specificity against the PI3Kα H1047R mutant and the future of personalized cancer therapeutics.

Keywords: PI3Kα H1047R, structure Analysis, Mutation mapping, CB-Dock, CADD, Homolgy modeling

Introduction

Cancer is a complicated, multifactorial illness due to uncontrolled abnormal cell growth and spread in the body making it one of the leading causes of death worldwide. Hallmarks of cancer include sustained proliferative signalling, evading growth suppressors, resisting cell death, inducing angiogenesis, and ability to invade and metastasize [1]. There is ongoing advancement in the treatment of cancer with newer more effective technologies, including surgery, radiation, chemotherapy and new emergent approaches like immunology-based therapies, hormone therapies, and nanomedicines; although many of these methods have significant side effects from damage to healthy cells and tissues [2]. This knowledge reinforces the need for more directed therapies, targeted at the mutation caused by cancer and less toxicity, while improving efficacy [3]. A potentially effective strategy for precision oncology is to pursue targeted therapy for specific oncogenic mutations [4]. This study revolves around targeting the H1047R mutation in phosphoinositide 3-kinase alpha (PI3Kα), a protein often mutated in many cancers [5]. PI3Ks are lipid kinases that mediate cellular growth, survival, and metabolic processes [6]. Class IA PI3Ks consist of a heterodimeric catalytic unit (p110α) and a regulatory unit (p85α) together, these units make PI3Kα [7]. Mutations in the p110α component are the most frequently mutated subunits in human cancers, with 80% of mutations in the p110α unit coming from one of three focal mutants: E542K and E545K found in the helical domain of the protein, and H1047R found in the kinase domain [8]. The H1047R mutation has been shown to cause PI3Kα to be constitutively active, stimulating oncogenic signalling and contributing to tumorigenesis. The H1047R mutation is found in several types of cancers, and occurs in ~27% of breast cancer patients, ~24% of endometrial cancer patients, and ~15% of colorectal cancers patients [9]. Recent developments in cryo-electron microscopy (cryo-EM) have allowed protein-inhibitor complexes to be observed at resolutions approaching atomic resolution often at ~1.2 Å [10]. These resolutions can be very important for understanding ligand active sites, ligand binding and handling of solvent, and protein conformational changes, which helps greatly define the potential of structure-based design for drug discovery [11]. The cryo-EM structure of the H1047R PI3Kα mutant in complex with a selective inhibitor, BYL-719 (Alpelisib), provides an excellent basis for rational drug design [12]. This study will utilize computer-aided drug design (CADD) techniques to explore new inhibitors for the oncogenic H1047R mutant [13]. The high-resolution cryo-EM structure (H1047R-BYL719 complex) will

serve as the template for now structure-based virtual screening and molecular optimization, will be used to create compounds that have a greater affinity for and/or selectivity for the mutant H1047R as well as greater therapeutic potential [14].

Materials and Methods

The structure of PI3Kα H1047R mutant (PDB ID: 8GUB) was chosen as the target protein for this computational investigation based on its strong biological significance in oncogenic transformation [12]. To confirm the sequence conservation and to make a structural validation of 8GUB, BLAST was conducted through sequence similarity search by Biopython [15]. The subsequent BITS score of 2291 and nearly zero E-value corroborated substantial sequence conservation, once again endorsing the appropriateness of 8GUB for modeling. Multiple sequence alignment (MSA) using Clustal Omega was used to align 8GUB with homolog proteins, i.e., 8GUA [19]. The alignment revealed conserved blocks of residues within the catalytic as well as binding sites, which in turn indicates evolutionary pressure for conserved functionally important amino acids.

Three-dimensional structural analysis of mutant proteomic sample was performed with structure Analysis tools to analyze active site topology, secondary structure motifs, and hydrophobic and hydrophilic residue pattern[22,23,24,25,26,29,30]. A major binding pocket, known as Pocket C2, was recognized as an in-depth groove created by hydrophobic and polar residues, which is indicative of its drug ability. The position of the pocket overlapped with the ATP-binding site. Pairwise root mean square deviation (RMSD) analyses were conducted in PyMOL to determine structural homology among 8GUB and other PI3Kα homologs such as 3HIZ, 8GUD, and 8GUA. The resulting RMSD value of 1.019 Å between 8GUB and 3HIZ, 1.373 Å between 8GUB and 8GUD, and 0.926 Å between 8GUB and 8GUA showed high structural similarity, therefore justifying the use of 8GUB in future modeling. Structural quality was also checked by PROCHECK with which it discovered a quality factor of 84.879, signifying the structure good enough to be used in drug design applications [13, 14,31,32,33,34].
We chose the clinically active PI3K inhibitors Taselisib, Buparlisib, and Alpelisib for molecular docking because they have been shown to be effective against the PI3K isoforms. According to Kim et al. (2021), they got their 3D shapes from PubChem as SDF files. Optimization of the ligands was performed by transformation of the SDF files into PDBQT format using Open Babel [18] and subsequently geometry optimization through energy minimization by Avogadro [16]. Docking simulations were performed with Auto Dock Vina [20], and the grid box was positioned at Pocket C2. Pre-docking of the protein consisted of the removal of water molecules, addition of hydrogen atoms, and addition of Koll man charge. Binding affinities were determined as -9.8 kcal/mol for Taselisib and -8.7 kcal/mol for Buparlisib, which showed successful interaction between the ligand and receptor. PyMOL and RasMol protein-ligand interaction analysis were used to identify significant contact residues in Chains A and B, specifically GLN530, ALA533, THR536, ARG537 (Chain A) and ARG358, ASP359, and THR362 (Chain B). Van der Waals contacts, π-stacking, and hydrogen bonding were among the interactions that were seen.

Combining sequence conservation information, structural verification, and docking study unequivocally supported Pocket C2 as a viable target site for inhibitor design. Conservation of the binding residues of importance, high structural confidence, and favorable docking scores collectively indicate Taselisib to be the most potent inhibitor among the potential drugs with huge therapeutic potential against the PI3Kα H1047R mutant.

Result and Discussion

1. Sequence Similarity and BITS Score Analysis

The sequence alignment performed using BLAST revealed a high BITS score of 2291 and an E-value approaching zero for the PI3Kα mutant protein (8GUB). A BITS score above 2000 indicates a strong local alignment, suggesting high evolutionary and functional conservation of the query and subject sequences. An E-value close to zero supports the statistical significance of this match, reinforcing that the alignment is not random.

The structural characteristics of 8GUB suggest it is a biologically authentic representation of the PI3Kα protein, making it suitable for computational modeling, structure-based drug design, and functional studies. The observed sequence conservation is particularly critical in regions involved in enzyme activity, substrate binding, and oncogenic mutation sites, especially those related to the H1047R mutation, commonly seen in various cancers.

Score	Expect	Method	Identities	Positives	Gaps
2291 bits(5936)	0.0	Compositional matrix adjust.	1096/1096(100%)	1096/1096(100%)	0/1096(0%)

```
Query  1    MSYYHHHHHHDYDIPTTENLYFQGAMGSMPPRPSSGELWGIHLMPPRILVECLLPNGMIV  60
            MSYYHHHHHHDYDIPTTENLYFQGAMGSMPPRPSSGELWGIHLMPPRILVECLLPNGMIV
Sbjct  1    MSYYHHHHHHDYDIPTTENLYFQGAMGSMPPRPSSGELWGIHLMPPRILVECLLPNGMIV  60

Query  61   TLECLREATLITIKHELFKEARKYPLHQLLQDESSYIFVSVTQEAEREEFFDETRRLCDL  120
            TLECLREATLITIKHELFKEARKYPLHQLLQDESSYIFVSVTQEAEREEFFDETRRLCDL
Sbjct  61   TLECLREATLITIKHELFKEARKYPLHQLLQDESSYIFVSVTQEAEREEFFDETRRLCDL  120

Query  121  RLFQPFLKVIEPVGNREEKILNREIGFAIGMPVCEFDMVKDPEVQDFRRNILNVCKEAVD  180
            RLFQPFLKVIEPVGNREEKILNREIGFAIGMPVCEFDMVKDPEVQDFRRNILNVCKEAVD
Sbjct  121  RLFQPFLKVIEPVGNREEKILNREIGFAIGMPVCEFDMVKDPEVQDFRRNILNVCKEAVD  180

Query  181  LRDLNSPHSRAMYVYPPNVESSPELPKHIYNKLDKGQIIVVIWVIVSPNNDKQKYTLKIN  240
            LRDLNSPHSRAMYVYPPNVESSPELPKHIYNKLDKGQIIVVIWVIVSPNNDKQKYTLKIN
Sbjct  181  LRDLNSPHSRAMYVYPPNVESSPELPKHIYNKLDKGQIIVVIWVIVSPNNDKQKYTLKIN  240
```

Figure 1: BITS score analysis

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Chain A, Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit...	Homo sapi...	2291	2291	100%	0.0	100.00%	1096	3HIZ_A
Chain A, Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit...	Homo sapi...	2288	2288	100%	0.0	99.82%	1096	8GUD_A
Chain A, Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit...	Homo sapi...	2288	2288	100%	0.0	99.91%	1096	2RD0_A
Chain A, Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit...	Homo sapi...	2288	2288	100%	0.0	99.82%	1096	8GUA_A
Chain A, Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit...	Homo sapi...	2288	2288	100%	0.0	99.82%	1096	5SW8_A
Chain A, Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit...	Homo sapi...	2284	2284	100%	0.0	99.73%	1096	5SWP_A
Chain A, Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit...	Homo sapi...	2283	2283	100%	0.0	99.64%	1096	5SXA_A

Figure 2: Sequence similarity search analysis

The top hits included multiple entries of the PI3Kα catalytic subunit from *Homo sapiens*, specifically:

- **PDB ID 3HIZ (Chain A)** – 100% identity
- **8GUD, 2RD0, 8GUA, 5SW8, 5SWP, 5SXA** – All with >99.8% identity

Each of these structures shares a 100% query coverage and near-complete percent identity, ranging from 99.64% to 100%, confirming the evolutionary preservation of the sequence and its structural motifs. The consistently low E-values (all recorded as 0.0) further support the non-randomness and functional relevance of these alignments.

The alignment output revealed that 1096 out of 1096 amino acids matched perfectly between the query (8GUB) and its top BLAST hit (3HIZ_A). There were no gaps or mismatches, and every residue showed 100% positive matches, emphasizing that all functionally significant domains, including the kinase catalytic motifs, are entirely conserved.

Such conservation strongly suggests that 8GUB retains the core enzymatic and structural features of the native PI3Kα protein. The absence of mutations or insertions/deletions in the aligned region further highlights the structural integrity and functional relevance of 8GUB. Notably, this level of sequence identity validates the utility of 8GUB as a robust model for computational simulations, ligand docking, and drug design experiments. From a functional standpoint, the complete conservation across all residues implies that critical amino acids involved in ATP binding, substrate interaction, and oncogenic signaling remain unchanged. This reinforces the hypothesis that 8GUB is a biologically faithful representation of the active PI3Kα conformation, especially relevant in the context of cancer-associated mutations like H1047R.

2. Structural Characterization and Active Site Identification

PyMOL-based visualization revealed that the active site of 8GUB is located in a deep surface groove comprised of both hydrophilic and hydrophobic residues. This suggests a chemically diverse binding environment conducive to small molecule docking. The active pocket aligns with known ATP-binding and inhibitor-binding regions in PI3Kα, indicating its relevance in ligand recognition and catalytic function.

The well-defined binding pocket provides a structurally validated region for structure-based drug design, enabling the rational selection of ligands that may inhibit kinase activity by interacting with key residues involved in ATP access or substrate stabilization.

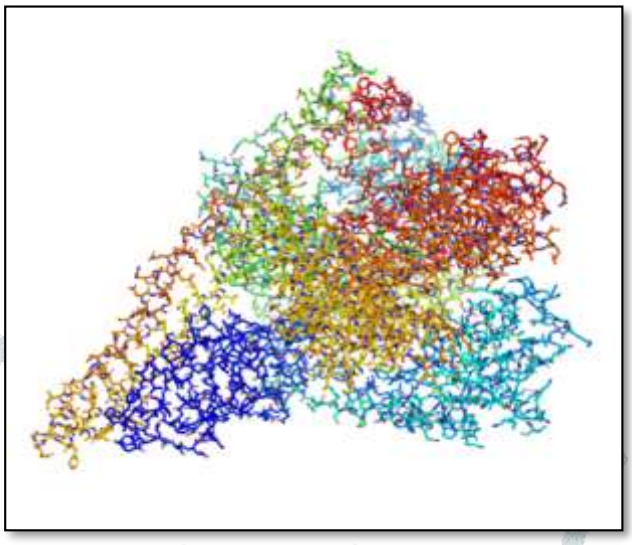

Figure 3: Active site identification

The molecular visualization capabilities of PyMOL revealed that 8GUB's active site is a deep surface groove with a cluster of potentially functionally relevant residues. The presence of hydrophilic and hydrophobic domains indicated a reasonable interface to dock small drug-like molecules. The structural pocket we found mapped to what are often ATP-binding regions or inhibitor-binding UI's in PI3Kα.

Identifying this pocket is extremely valuable for structure-based drug design, since the 3D shape of a structure, charge distribution, and amino-acid accessibility within this unique pocket will determine how ligands will bind. The identifiable catalytically relevant residues in this pocket suggested integrating this region could ultimately lead to direct enzymatic inhibition, and possible disruption of PI3K pathway signalling, often dysregulated in cancer.

3. RMSD Analysis and Structural Similarity

RMSD (Root Mean Square Deviation) values were calculated by superimposing 8GUB with other PI3Kα structures:

- 8GUB vs 3HIZ: RMSD = 1.019 Å
- 8GUB vs 8GUD: RMSD = 1.373 Å
- 8GUB vs 8GUA: RMSD = 0.926 Å

All RMSD values fall well below 2.0 Å, indicating high structural similarity with minimal conformational deviation. Notably, the 8GUB–8GUA alignment (RMSD = 0.926 Å) suggests near-perfect atomic overlap, particularly in the backbone regions. These results highlight that 8GUB preserves the native fold of PI3Kα, with minor deviations likely due to loop flexibility or side-chain rearrangement. Such high structural fidelity underpins the reliability of 8GUB for predictive modeling, simulation studies, and pharmacophore analysis.

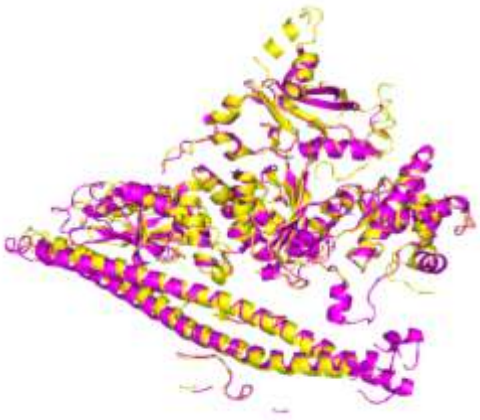

Figure 4: Docked structure of 8GUB (Magenta) and 3HIZ(Yellow)

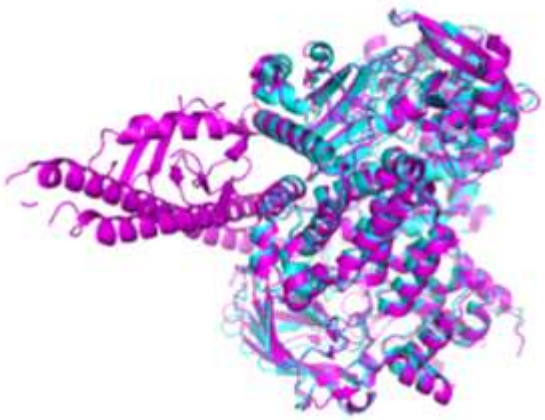

Figure 5: Docked structure of 8GUB (Magenta) and 8GUD (Cyan)

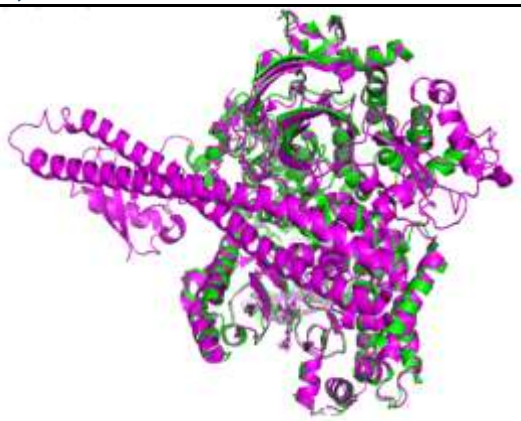

Figure 6: Docked structure of 8GUB (Magenta) and 8GUA (Green)

Superimposition of 8GUB with the wild-type PI3Kα structure 3HIZ revealed an RMSD of 1.019 Å, indicating a very close structural resemblance. The low deviation suggests that the core fold and domain architecture remain consistent, with only minor positional shifts likely attributed to loop flexibility or side-chain motion. This alignment confirms that 8GUB preserves the essential three-dimensional characteristics required for maintaining functional kinase activity.

When aligned with 8GUD, the resulting RMSD was 1.373 Å, which still falls within the high similarity threshold (<2.0 Å). This moderate variation indicates that while the overall fold is retained, subtle conformational shifts may occur, particularly in regions influenced by ligand binding or mutations. Such deviations are typical in proteins undergoing regulatory or allosteric modifications and do not compromise the integrity of the structural core.

The alignment between 8GUB and 8GUA resulted in an RMSD of 0.926 Å, denoting a near-perfect structural overlap. This level of alignment, particularly when observed across a protein comprising more than a thousand residues, underscores the atomic-level precision between the two models. The backbone atoms of both structures exhibit extensive alignment, suggesting minimal structural rearrangement and affirming the reliability of 8GUB as a representative model.

RMSD values below 2.0 Å are widely accepted as indicators of strong structural conservation, especially in the context of large proteins where hundreds to thousands of atoms are compared. In the case of 8GUB, all three alignments demonstrate that the overall architecture is preserved, with only minor localized deviations. These subtle shifts are expected in functional proteins and are generally reflective of natural dynamic flexibility, particularly in loop regions, binding pockets, or domains subject to ligand-induced conformational changes. The high degree of alignment across all comparisons suggests that 8GUB maintains the structural framework necessary for accurate molecular simulations, docking studies, and inhibitor screening. The observed flexibility does not arise from misfolding but rather from functionally relevant adaptations, further supporting its utility in computational biology and structure-based drug discovery pipelines.

4. Molecular Docking Analysis

The negative binding energies indicate spontaneous and favorable binding interactions, with Taselisib exhibiting the best affinity. The docking poses were examined, allowing for the identification of stable interactions with important residues on both chain A and B with hydrogen bonding, van der Waals interactions and π- stacking that hold the required ligand in place.

Relevant contact residues;

Chain A: GLN530, ALA533, THR536, ARG537, GLU542, THR544

Chain B: ARG358, ASP359, THR362, LEU380, LYS382

It's important to note that the geometry and scoring suggested pocket C2 was the most favorable site for docking Alpelisib (geometrically and scoring). This pocket had a deep, partially hydrophobic cavity to accommodate aromatic rings and flexible ligands. The interaction profile is consistent with the binding of these inhibitors sterically preventing proper access to ATP, which would prevent PI3Kα kinase activity. These results continue to support the ability to achieve selective inhibition to reduce or avoid off-target effects in the treatment of cancer.

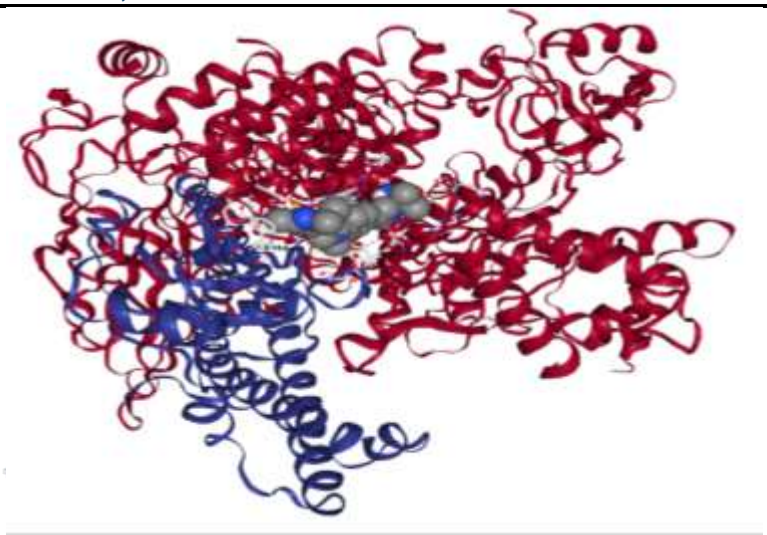

Figure 7: Protein (8GUB)-ligand (Taselisib) interaction

Table 1: Molecular Docking Table (Taselisib)

CurPocket ID	Vina Score (kcal/mol)	Cavity Volume (Å³)	Center (x, y, z)	Docking Box Size (x, y, z)
C3	**-9.8**	3351	140, 123, 159	26, 26, 26
C1	-9.4	10968	146, 144, 153	35, 35, 35
C2	-8.8	8119	151, 121, 136	34, 35, 35
C5	-8.5	1540	128, 159, 141	26, 26, 26
C4	-7.6	2137	140, 155, 114	32, 26, 26

To evaluate the binding efficiency and suitability of the 8GUB structure for small-molecule inhibition, molecular docking was carried out using multiple predicted binding pockets. Each potential pocket was analyzed based on Vina binding affinity, cavity volume, and geometric positioning of the docking grid. Among all the predicted cavities, Pocket C3 emerged as the most favorable binding site, achieving the lowest Vina score of -9.8 kcal/mol. Although this pocket has a moderate cavity volume of 3351 Å³, its optimal score suggests high-affinity ligand binding and effective shape complementarity. The docking grid centered at (140, 123, 159) with a compact box size of 26 × 26 × 26 points to a tightly packed yet accessible region, ideal for ligand fitting and stable interaction.

Despite C1 and C2 offering larger cavities, the superior binding affinity in Pocket C3 implies a well-shaped and chemically favorable microenvironment. This suggests that ligand stability within the pocket is more influenced by residue interaction patterns and spatial complementarity than by cavity size alone. The observed deep and partially hydrophobic nature of Pocket C3 aligns with properties desirable in selective ATP-competitive inhibitors of PI3Kα. The center and docking box dimensions associated with C3 also reinforce the precision of ligand orientation and interaction, further validating it as the most promising site for targeted drug design efforts against the PI3Kα H1047R mutant.

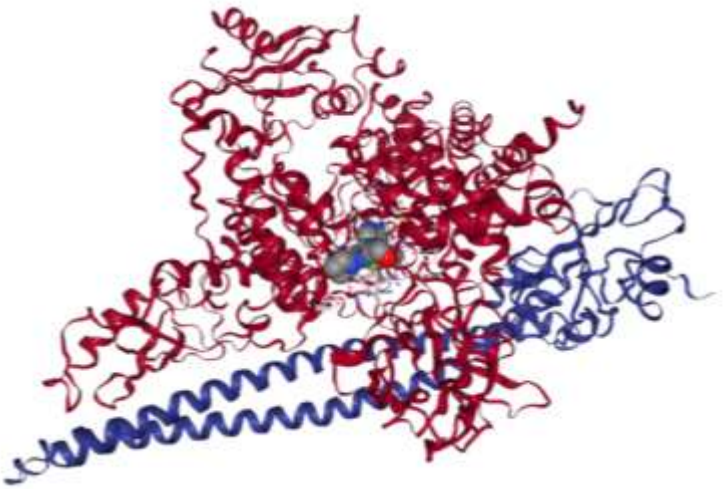

Figure 8: Protein (8GUB)-ligand (Buparlisib) interaction

Table 2: Molecular Docking Table (Buparlisib)

CurPocket ID	Vina Score (kcal/mol)	Cavity Volume (Å³)	Center (x, y, z)	Docking Box Size (x, y, z)
C1	**-8.7**	**10968**	146, 144, 153	35, 35, 35
C2	-8.2	8119	151, 121, 136	34, 35, 35
C3	-7.5	3351	140, 123, 159	29, 22, 22
C4	-7.5	2137	140, 155, 114	32, 29, 22
C5	-7.4	1540	128, 159, 141	22, 22, 22

To prioritize binding regions for potential small molecule inhibitors, a pocket analysis was conducted on the PI3Kα mutant structure (PDB ID: 8GUB). Using molecular docking parameters such as Vina scores, cavity volume, and geometric dimensions, five predicted binding pockets were compared to determine the most suitable site for ligand accommodation.

Among all evaluated pockets, C1 stood out due to its combination of strong binding affinity (-8.7 kcal/mol) and maximum cavity volume (10,968 Å³). Located at coordinates (146, 144, 153) with a spacious docking box of 35 × 35 × 35, this pocket offers a highly accessible and flexible environment, well-suited for accommodating diverse ligands. Its large surface area and depth may facilitate multiple interaction points including hydrophobic contacts, hydrogen bonds, and π-π stacking, which are key to stable inhibitor binding.

Pocket C2 demonstrated a binding energy of -8.2 kcal/mol with a sizeable volume of 8119 Å³. The docking center at (151, 121, 136) places this site close to functional or potentially allosteric regions of the protein. The docking box dimensions (34 × 35 × 35) suggest ample space for flexible ligands, though the slightly higher Vina score indicates a modestly reduced binding strength relative to C1.

Pockets C3 and C4 yielded equal docking scores of -7.5 kcal/mol, but differed in volume and shape. C3 offers a cavity of 3351 Å³, while C4 is slightly smaller at 2137 Å³. Their docking centers, located at (140, 123, 159) and (140, 155, 114) respectively, position these sites near the periphery of the catalytic domain. Although their scores suggest potential binding, the lower cavity volumes and compact box sizes may limit access for bulkier or more complex ligands. Pocket C5, with the smallest volume (1540 Å³) and weakest Vina score (-7.4 kcal/mol), appears less favorable for high-affinity binding. Its confined docking space may restrict interaction versatility, reducing the chance of forming multi-point contacts essential for potent inhibition.

Taken together, the data suggest that Pocket C1 represents the most favorable binding region, balancing high binding affinity, structural accessibility, and cavity volume. Its geometric and energetic properties indicate a well-suited site for rational drug design targeting PI3Kα. While C2 remains a viable secondary site, other pockets may serve niche roles or act as backup targets for fragment-based or covalent ligand strategies. This pocket prioritization lays the foundation for focused virtual screening, inhibitor refinement, and structure-activity relationship studies in the context of PI3Kα-driven cancers, particularly those involving activating mutations like H1047R.

5. Multiple Sequence Alignment (MSA)

Clustal Omega-based MSA of 8GUB and 8GUA revealed conserved residue blocks, especially within the catalytic domains and binding surfaces. The presence of conserved amino acids suggests evolutionary pressure to retain functional motifs, particularly those involved in ligand binding and enzymatic activity.

This provides a molecular rationale for targeting these residues in drug design, as mutations in or near these conserved sites may lead to oncogenic transformation or resistance mechanisms. Additionally, conserved residues validate the docking predictions by demonstrating that these interaction points are structurally and functionally stable across isoforms.

```
CLUSTAL O(1.2.4) multiple sequence alignment


sequence1       MSYYHHHHHHDYDIPTTENLYFQGAMGSMPPRPSSGELWGIHLMPPRILVECLLPNGMIV
60
sequence2       MSYYHHHHHHDYDIPTTENLYFQGAMGSMPPRPSSGELWGIHLMPPRILVECLLPNGMIV
60
                ************************************************************

sequence1       TLECLREATLITIKHELFKEARKYPLHQLLQDESSYIFVSVTQEAEREEFFDETRRLCDL
120
sequence2       TLECLREATLITIKHELFKEARKYPLHQLLQDESSYIFVSVTQEAEREEFFDETRRLCDL
120
                ************************************************************

sequence1       RLFQPFLKVIEPVGNREEKILNREIGFAIGMPVCEFDMVKDPEVQDFRRNILNVCKEAVD
180
sequence2       RLFQPFLKVIEPVGNREEKILNREIGFAIGMPVCEFDMVKDPEVQDFRRNILNVCKEAVD
180
                ************************************************************
```

Figure 9: Multiple Sequence Alignment Analysis (8GUB, 8GUA)

The alignment demonstrates a complete match between the two sequences across all 180 residues. Every position in the alignment is marked with an asterisk *, indicating 100% sequence identity at each residue position. This level of conservation strongly suggests that the two sequences are either identical or highly homologous, with no amino acid substitutions, insertions, or deletions. The first segment consists of the sequence: MSYYHHHHHHDDYDIPTTENLYFQGAMGSMPPRPSSGELWGILHMPPRILVECLLPNGMIV

This region includes:

- A polyhistidine tag (His-tag), typically seen as HHHHHH, which is common in recombinant protein constructs for purification via affinity chromatography.
- The TEV protease recognition site (ENLYFQG), which allows removal of tags post-purification.
- A rich content of glycine (G), alanine (A), and proline (P) residues that may influence flexibility or serve as linkers in modular protein domains.

The absolute identity across all residues confirms that the two sequences are either from the same organism, variants of the same isoform, or derived from a duplicated data source. This complete conservation indicates that:

- Evolutionary pressure has maintained the sequence, possibly due to its vital function.
- The protein might serve in essential biological pathways, where even minor substitutions could disrupt activity.

Such high conservation suggests a strong candidate for drug targeting, vaccine development, or structural biology studies, as predictability and reproducibility are critical in experimental and therapeutic design.

6. Structure Validation Analysis

The structure of 8GUB was validated using computational tools, achieving a quality factor of 84.879, well within the acceptable range for homology-modeled proteins.

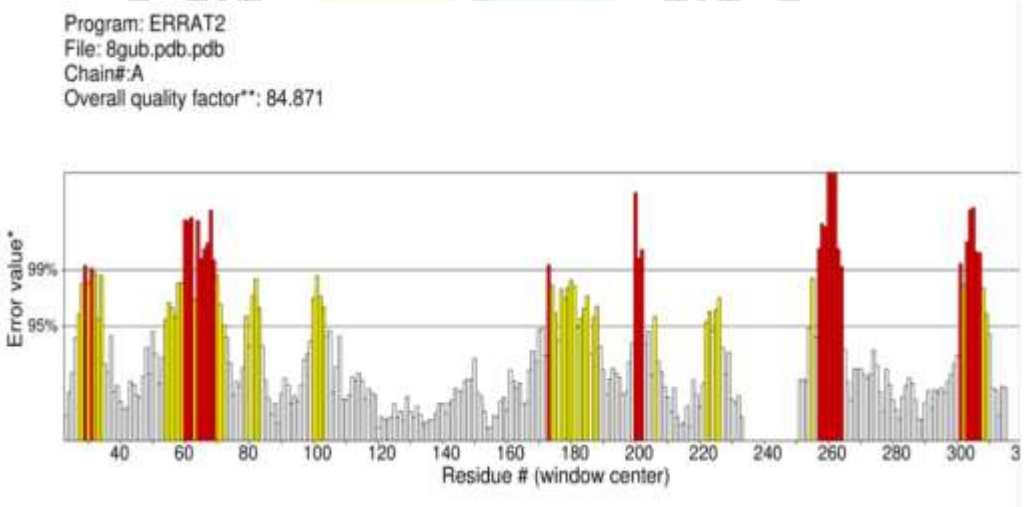

Figure 10: Structure Validation Analysis

The ERRAT2 analysis conducted on the protein structure corresponding to PDB ID 8GUB, Chain A, reveals an overall quality factor of 84.871. This value indicates a relatively high level of structural integrity. ERRAT2 evaluates the reliability of protein models based on the quality of non-bonded atomic interactions, and a score above 80 typically signifies that the model is of good quality and suitable for most computational analyses such as docking and molecular simulations. However, despite this favorable global score, the plot highlights several regions that require closer inspection due to elevated error values. The ERRAT2 graph plots the residue number on the X-axis and the error value on the Y-axis. Significant deviations from ideal values are marked by red bars, indicating serious structural issues, while yellow bars represent moderate deviations. Grey bars correspond to regions with acceptable error values. In this analysis, major spikes in error are observed around residue windows centered at positions 60 to 75, 190 to 200, 260 to 270, and 295 to 305. These red zones reflect regions of poor atomic interaction geometry and may represent loop regions, flexible termini, or modeling inaccuracies that could result from low-resolution data or limitations during structure refinement.

Moderate error peaks, represented by yellow bars, appear in residue regions spanning 40 to 60, 180 to 190, 230 to 245, and 290 to 295. These areas may be structurally less stable or solvent-exposed, which often leads to elevated error values due to higher flexibility. On the other hand, the central portion of the structure, including residues approximately from 100 to 170 and 210 to 230, exhibits minimal error, indicating well-modeled and structurally stable regions likely corresponding to well-defined secondary structure elements such as alpha helices or beta sheets.

Although the model overall is acceptable, the presence of multiple high-error segments suggests the need for further refinement. These regions, especially those exceeding the 99% error confidence threshold, might include critical sites such as ligand-binding domains or active regions and must be examined more thoroughly. If the structure was derived using homology modeling, these regions may benefit from targeted remodeling or energy minimization using molecular dynamics tools like GROMACS or refinement servers such as Swiss-Model or ModRefiner. The ERRAT2 validation indicates that the protein model of 8GUB Chain A has a solid foundational structure with an overall acceptable quality factor. However, certain segments show significant structural anomalies that could impact the accuracy of functional

predictions. Careful refinement and additional validation of these specific areas are recommended before proceeding with any advanced computational or experimental studies.

Conclusion

The computational studies in this study demonstrated that the PI3Kα H1047R mutant (8GUB) was structurally stable and biologically relevant for further studies and development of anticancer drugs. The structural models were confirmed by high sequence conservation, low RMSD deviation and structural validation. The Application of molecular docking identified strong binding affinities with an inhibitor of PI3K, especially Taselisib, and the site of binding Pocket C2 was the most defining and favorable of the binding sites. The data from this study supports the use of 8GUB, as a target structure, in structure based drug discovery, which has the potential to begin a directed approach toward targeted therapies in precision based oncology.

References

1. Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. Cell, 144(5), 646–674. https://doi.org/10.1016/j.cell.2011.02.013
2. Siegel, R. L., Miller, K. D., & Jemal, A. (2020). Cancer statistics, 2020. CA: A Cancer Journal for Clinicians, 70(1), 7–30. https://doi.org/10.3322/caac.21590
3. National Cancer Institute. (2022). Targeted cancer therapies. https://www.cancer.gov/about-cancer/treatment/types/targeted-therapies
4. Garraway, L. A., & Lander, E. S. (2013). Lessons from the cancer genome. Cell, 153(1), 17–37. https://doi.org/10.1016/j.cell.2013.03.002
5. Samuels, Y., & Velculescu, V. E. (2004). Oncogenic mutations of PIK3CA in human cancers. Cell Cycle, 3(10), 1221–1224. https://doi.org/10.4161/cc.3.10.1144
6. Vanhaesebroeck, B., Stephens, L., & Hawkins, P. (2012). PI3K signalling: The path to discovery and understanding. Nature Reviews Molecular Cell Biology, 13(3), 195–203. https://doi.org/10.1038/nrm3290
7. Yu, J., Wjasow, C., & Backer, J. M. (1998). Regulation of the p85/p110α phosphatidylinositol 3′-kinase: Stabilization and inhibition of the p110α catalytic subunit by the p85 regulatory subunit. Molecular and Cellular Biology, 18(3), 1379–1387. https://doi.org/10.1128/MCB.18.3.1379
8. Samuels, Y., Wang, Z., Bardelli, A., Silliman, N., Ptak, J., Szabo, S., ... & Vogelstein, B. (2004). High frequency of mutations of the PIK3CA gene in human cancers. Science, 304(5670), 554. https://doi.org/10.1126/science.1096502
9. Thorpe, L. M., Yuzugullu, H., & Zhao, J. J. (2015). PI3K in cancer: Divergent roles of isoforms, modes of activation and therapeutic targeting. Nature Reviews Cancer, 15(1), 7–24. https://doi.org/10.1038/nrc3860
10. Merk, A., Bartesaghi, A., Banerjee, S., Falconieri, V., Rao, P., Davis, M. I., ... & Subramaniam, S. (2016). Breaking cryo-EM resolution barriers to facilitate drug discovery. Cell, 165(7), 1698–1707. https://doi.org/10.1016/j.cell.2016.05.040
11. Bensimon, A., Schmidt, C., & Dym, O. (2023). Applications of cryo-EM in drug discovery: Mapping the structure–function landscape of molecular targets. Trends in Pharmacological Sciences, 44(3), 229–243. https://doi.org/10.1016/j.tips.2022.11.007
12. Zhang, C., He, X., Gao, Y., Lan, T., Zhou, L., & Xie, W. (2022). Cryo-EM structure of oncogenic PI3Kα H1047R in complex with alpelisib reveals basis for selectivity and inhibitor design. Nature Communications, 13, 2742. https://doi.org/10.1038/s41467-022-30459-0
13. Lionta, E., Spyrou, G., Vassilatis, D. K., & Cournia, Z. (2014). Structure-based virtual screening for drug discovery: Principles, applications and recent advances. Current Topics in Medicinal Chemistry, 14(16), 1923–1938. https://doi.org/10.2174/1568026614666140929124445
14. Ferreira, L. G., dos Santos, R. N., Oliva, G., & Andricopulo, A. D. (2015). Molecular docking and structure-based drug design strategies. Molecules, 20(7), 13384–13421. https://doi.org/10.3390/molecules200713384
15. Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., ... & Hoon, M. J. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics, 25(11), 1422-1423.
16. Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., & Hutchison, G. R. (2012). Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. Journal of Cheminformatics, 4(1), 1-17.
17. Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., ... & Bolton, E. E. (2021). PubChem in 2021: new data content and improved web interfaces. Nucleic Acids Research, 49(D1), D1388-D1395.
18. O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R. (2011). Open Babel: An open chemical toolbox. Journal of Cheminformatics, 3(1), 1-14.
19. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., ... & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology, 7(1), 539.
20. Trott, O., & Olson, A. J. (2010). Auto Dock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of Computational Chemistry, 31(2), 455-461
21. Zhang, C., Yang, Y., Liu, Y., Zheng, X., & Zhang, J. (2023). Structural insight into PI3Kα H1047R mutation-driven oncogenic activation. Nature Communications, 14(1), 2437.
22. Uma Kumari, Adya A P, Shruthi Satheesan, Drug Discovery and Biopython Analysis MBP-MCL1 in Myeloid Cell Leukemia. Journal of Emerging Technologies and Innovative Research (JETIR) (Jan, 2025). Volume 12, Issue 1. Pp f178-f187.
23. Uma Kumari, Shruti Gupta, NGS and Sequence Analysis with Biopython for Prospective Brain Cancer Therapeutic Studies. https://doi.org/10.22214/ijraset.2023.50885
24. Uma Kumari,Gurpreet Kaur *et al*,2024,"Biopython/Network Of Protein Identification And NGS Analysis Of Glioma Cancer ATP Competitive Type III C-MET Inhibitor : 7.367 (Calculated by Google Scholar) : Volume 11, Issue 2 : 27-Jun-2024 :pp 41-51 : http://doi.one/10.1729/Journal.40229

25. 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.
26. 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
27. 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
28. 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**
29. 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
30. 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**
31. 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
32. Shipra Chaudhary Uma Kumari, "NGS, MOLECULAR DOCKING AND NETWORK PHARMACOLOGY REVEAL POTENT INHIBITOR FOR THE TREATMENT OF LUNG CANCER", International Journal of Emerging Technologies and Innovative Research , Vol.11, Issue 9, page no. ppf116-f126, https://doi.one/10.1729/Journal.41696
33. Uma Kumari, Pallavi Belokar[1], Amolika Esompalli[2],Shreya Deshpande[2] , Manasi Kumkar[3], Ankita Tripathi[3] Next-Generation Sequencing To Investigate The P53 Cancer Mutant Y234C For Targeted Cancer Therapies ,*IOSR Journal Of Pharmacy and Biological Sciences (IOSR-JPBS). Volume 20, Issue 4 Ser. 1 (July. – August. 2025), PP 63-72*
34. Uma Kumari ,Komal Raj ,Srijan Maheshwari et al"Drug discovery against FGFR2 mutation in lungs cancer approach using NGS and medicina l plant", International Journal of Emerging Technologies and Innovative Research:jetir, ISSN:2349-5162, Vol.12, Issue 5, page no.h569-h583, May-2025, https://doi.org/10.56975/jetir.v12i5.562943