Journal of Cardiovascular Disease Research

ISSN: 0975-3583, 0976-2833 VOL14, ISSUE 12, 2023

PROTEOMICS HOMOLOGY MODELING AND NGS ANALYSIS OF ACUTE MYELOID LEUKEMIA

Imama Ghani Ansari[1] Uma Kumari[1*] Anoushka Prabhu[1] Motukuri Naveen Kumar[2] Mahadevan Swamy[2]

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

[1*]*Senior Bioinformatics Scientist, Bioinformatics Project and Research Institute, Noida - 201301, India*

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

[2]*Project Trainee, Bioinformatics Project and Research Institute,Noida-201301,India*

[2]*Project Trainee, Bioinformatics Project and Research Institute,Noida-201301,India*

Corresponding Author Uma kumari[1]*

ABSTRACT

The research employed an integrated approach utilizing various bioinformatics tools and databases to comprehensively analyze the molecular and structural aspects of the studied protein. PubChem was utilized for extracting information about the drugs used in the study, showcasing its significance in providing access to biological activities of small molecules. The protein structure, identified by the PDB ID 6U9N, was retrieved from the Protein Data Bank (PDB), emphasizing the importance of this repository in offering three-dimensional structural data crucial for diverse scientific applications. Rasmol and PyMOL were employed for visualizing and analyzing the protein structure, demonstrating their roles in facilitating molecular visualization and structural analysis. Sequence similarity and phylogenetic analysis were performed using BLAST and COBALT, respectively, showcasing their utility in exploring biological sequences and conducting multiple sequence alignments. KEGG software was employed for pathway analysis, highlighting its importance in understanding complex biological pathways. For protein-drug docking, CB Dock2 was used, emphasizing its user-friendly approach to blind docking. Further analysis of protein properties utilized ExPASyProtParam and PDBePISA, demonstrating their roles in computing physicochemical properties and analyzing macromolecular interfaces. Phyre2 was employed for secondary structure prediction and sequence similarity, showcasing its utility in predicting protein structures. STRING was used to predict protein-protein interactions and functional associations. ERRAT and SAVES were employed for assessing the quality of protein structures, emphasizing their significance in structural validation. The comprehensive use of these tools and databases provides a thorough understanding of the protein's structural, functional, and interaction characteristics, contributing valuable insights to the field of molecular and structural biology.

Keywords:Homology Modeling,Phyre2,Molecular Docking,KEGG,Verify protein structure,

INTRODUCTION

Advancements in our comprehension of the pathophysiology of acute myeloid leukemia have not yet resulted in significant enhancements in disease-free and overall survival rates among adult patients. It is worth noting that only approximately one-third of individuals aged 18-60 diagnosed with AML (Acute myeloid leukemia) have the possibility of achieving a cure. It is the predominant type of acute leukemia found in adults and is responsible for the highest number of leukemia-related fatalities in the United States each year. Furthermore, attaining disease-free survival is particularly uncommon, and the present therapy options have severe

ISSN: 0975-3583, 0976-2833 VOL14, ISSUE 12, 2023

consequences for older adults [1]. Acute myeloid leukemia is a diverse group of leukemias that occur when hematopoietic precursors undergo clonal transformation due to chromosomal rearrangements and multiple gene mutations [2]. The term acute refers to the rapid advancement of the disease. This form of leukemia is categorized as myelogenous because it impacts a specific set of white blood cells known as myeloid cells. These cells typically mature into diverse blood cell types, including red and white blood cells and platelets. Numerous recent scientific discoveries have contributed valuable understanding to classifying acute myeloid leukemia. While the impact on disease treatment is presently limited, ongoing investigations into different strategies hold promise for enhancing patient outcomes.

The three primary types of blood and bone marrow cancer in the workplace are leukemia, lymphoma, and myeloma. Leukemia is a kind of blood cancer that occured in the blood and bone marrow. It occurs when the body produces an excessive amount of abnormal white blood cells, which disrupts the bone marrow's ability to generate red blood cells and platelets. Non-Hodgkin lymphoma is a blood cancer that develops in the lymphatic system from lymphocytes, a type of white blood cell that aids in fighting infections. Hodgkin lymphoma is a blood cancer that develops in the lymphatic system from lymphocytes. Hodgkin lymphoma is distinguished by the presence of an abnormal lymphocyte known as the Reed-Sternberg cell. Multiple myeloma is a blood cancer that initiates in the plasma cells of the blood, a specific type of white blood cell produced in the bone marrow. There are also rare forms of blood and bone marrow cancers and associated disorders that we should be aware of. Some of these include Myelodysplastic syndromes are rare conditions that may occur due to damage to blood-forming cells in the bone marrow. Myeloproliferative neoplasms are rare blood cancers that arise when the body produces an excess of white blood cells, red blood cells, or platelets. The main subcategories are essential thrombocythemia, myelofibrosis, and polycythemiavera. Amyloidosis is a rare disorder characterized by the accumulation of an abnormal protein called amyloid. Although not cancer, it is closely associated with multiple myeloma. Waldenstrommacroglobulinemia is a rare type of non-Hodgkin lymphoma that originates in B cells. Common symptoms of early-stage acute myelogenousleukemia can resemble flu-like symptoms or other typical ailments. Typical signs and symptoms associated with acute myelogenousleukemia encompass, fever, bone discomfort, fatigue, difficulty breathing, paleness of the skin, frequent infections, easy bruising, and abnormal bleeding, such as frequent nosebleeds and gum bleeding. Acute myelogenousleukemia occurs when a bone marrow cell undergoes mutations in its genetic material or DNA. The DNA of a cell contains the instructions that regulate its functions. Typically, the DNA instructs the cell to grow and die at specific rates. In AML, however, these mutations cause the bone marrow cell to continue growing and dividing uncontrollably. As a result, there is an excessive production of immature cells, which develop into leukemic white blood cells called myeloblasts. These abnormal cells are unable to carry out their intended functions and can overwhelm healthy cells. While the exact cause of the DNA mutations that lead to leukemia is not known, medical professionals have identified factors that increase the risk. Factors that may contribute to an individual's heightened likelihood of developing acute myelogenousleukemia include various elements such as advancing age, with the risk being most prevalent among adults aged 65 and above. Additionally, this condition is more commonly observed in males compared to females. Previous exposure to cancer treatment, specifically certain types of chemotherapy and radiation therapy, may also elevate the risk of acquiring AML. Furthermore, individuals who have been exposed to high levels of radiation, such as survivors of nuclear reactor accidents, face an increased susceptibility to AML. Exposure to hazardous chemicals, including benzene, is linked to a greater risk of developing this

ISSN: 0975-3583, 0976-2833 VOL14, ISSUE 12, 2023

condition. The act of smoking has also been associated with AML due to the presence of benzene, as well as other carcinogenic substances present in cigarette smoke. Individuals who have previously experienced other blood disorders, such as myelodysplasia, myelofibrosis, polycythemiavera, or thrombocythemia, have a higher likelihood of developing AML. Lastly, certain genetic disorders, such as Down syndrome, have been associated with an elevated risk of AML. In individuals diagnosed with acute myelogenousleukemia, there are often imbalances in blood cell levels, including an excess of white blood cells, a deficiency of red blood cells, and a shortage of platelets. However, there are cases where the white blood cell count is too low. Additionally, the presence of blast cells, which are immature and typically found within the bone marrow and not within the bloodstream, can serve as an indicator of acute myelogenousleukemia. While a blood test can provide indications of leukemia, a bone marrow test is typically required to confirm the diagnosis. During a bone marrow biopsy, a needle is utilized to extract a sample of bone marrow, usually obtained from the hipbone. The sample is then sent to a laboratory for further examination. In certain circumstances, a lumbar puncture, also known as a spinal tap, may be necessary to extract and analyze the fluid surrounding the spinal cord for leukemia cells. Laboratory testing is conducted to analyze cancer cells, allowing doctors to gain a better understanding of the specific gene mutations present. This information helps determine prognosis and guide treatment decisions. Chemotherapy is a commonly used treatment method for achieving remission in patients with leukemia. It involves the use of chemicals to eliminate cancer cells in the body. Individuals with AML typically receive chemotherapy while in the hospital, as the drugs can negatively impact normal blood cell count while targeting leukemia cells. In cases where initial chemotherapy cycles do not induce remission, the treatment can be repeated. Targeted therapy is another approach that focuses on addressing specific abnormalities found within cancer cells. By blocking these abnormalities, targeted therapy can lead to the death of cancer cells. A thorough analysis of your leukemia cells will determine if targeted therapy is a viable option for you. It can be used as a standalone treatment or in combination with chemotherapy for induction and consolidation therapy. For consolidating treatment, a bone marrow transplant may be considered. This procedure involves the replacement of unhealthy bone marrow with leukemia-free stem cells. This helps in the regeneration of healthy bone marrow and the establishment of a renewed supply of healthy stem cells. Acute myeloid leukemia is a fast-growing blood cancer that requires prompt diagnosis and targeted treatments to improve patient outcomes. Ongoing advancements in therapeutic approaches, such as targeted therapies and immunotherapies, are being pursued to enhance the prognosis and quality of life for individuals affected by this difficult disease.

Acute myeloid leukemia (AML) is a type of blood cancer that involves the rapid growth and buildup of immature, unusually developed blood cells in the bone marrow and bloodstream [5]. Around 18,500 new cases occur in Europe, and approximately 20,000 new cases are reported in the USA, making this the most prevalent type of acute leukemia[6]. Recognized as a condition linked to aging, individuals diagnosed with AML typically have an average age of 67-68, and virtually all fatalities related to AML occur in adults. The advancing age of patients is also identified as a prognostic factor in AML. The condition is presently acknowledged as an exceptionally diverse entity. Research into chromosomal abnormalities through cytogenetics, coupled with a molecular genetics approach, has played a crucial role in unraveling the heterogeneity of AML and offering profound insights into leukemia biology. Identifying recurrent chromosomal and molecular defects at the time of diagnosis has yielded valuable prognostic information, including insights into responses to induction chemotherapy, the risk of relapse, and overall survival rates.

ISSN: 0975-3583, 0976-2833 VOL14, ISSUE 12, 2023

AML is presently categorized into subtypes, taking into account cellular lineage and appearance, expression of cell-surface or cytoplasmic markers, chromosome abnormalities, and gene mutations. This classification utilizes either the French American British (FAB) system [7] or the World Health Organization (WHO) classification system [8]. Factors that increase the risk of AML involve exposure to certain substances, such as benzene and chemotherapy drugs, as well as radiation. Other risk factors include having conditions like myelodysplastic syndrome or myeloproliferative neoplasms. Additionally, there is an elevated risk if there are genetic mutations in family-linked AML genes (like CEBPA, DDX4, or RUNX1) or if there are mutations in genes associated with clonal hematopoiesis (such as DNMT3A, TET2, or ASXL1) [9]. Individuals with inherited genetic syndromes like Fanconi'sanemia, Bloom syndrome, Down syndrome, and others face a higher likelihood of developing AML [10]. The most impactful role of cytogenetic abnormalities in directing AML treatment is evident in acute promyelocyticleukemia (APL), specifically the M3 subtype of AML. APL is identified by the t (15;17) q (22;12) cytogenetic anomaly, leading to the formation of the promyelocyticleukemia-retinoic acid receptor alpha (PML-RARA) fusion protein [11]. The condition was observed to positively react to all-trans-retinoic acid (ATRA) and arsenic trioxide (ATO) [12]. ATRA helps transform leukemic promyelocytes into mature cells, while ATO speeds up the breakdown of PML-RARA. This breakthrough not only increased the chances of beating APL from 30% to over 90%, but it also stands out as the pioneering use of targeted therapy at the molecular level— a monumental advancement in the history of AML treatment [13,14,15]. AML patients can be grouped into three risk categories based on cytogenetics: (1) a favorable risk group showing relatively positive outcomes with chemotherapy; (2) an unfavorable risk group experiencing significantly poorer outcomes, often considered for allogeneic stem cell transplantation; and (3) an intermediate risk group comprising patients not falling into the favorable or unfavorable categories. The definition of the intermediate risk group can be ambiguous and varies among medical professionals. The United Kingdom Medical Research Council (MRC-C) prognostic classification system identifies the intermediate risk group as patients lacking identifiable cytogenetic abnormalities of favorable or unfavorable groups [16,17,18]. In contrast, the European Leukemia Net (ELN-C) system utilizes common genetic mutation information (NPM1 and FLT3 ITD) to further classify the intermediate group into intermediate-1 and intermediate-2 groups with higher and lower risks, respectively[19,20]. In the past decade, our grasp of the genetic and epigenetic terrain in AML has advanced significantly, largely owing to the progress made in next-generation sequencing techniques.

MATERIAL AND METHOD

The drugs that have been used is extracted from the pubchem. PubChem is an invaluable resource in the realm of chemistry and the life sciences. It's a vast database maintained by the National Center for Biotechnology Information (NCBI), a part of the United States National Library of Medicine (NLM). The primary aim of PubChem is to provide free access to information on the biological activities of small molecules [21,22]. The protein structure has been obtained from PDB database. The PDB id of the protein is 6U9N. The Protein Data Bank (PDB) is an essential repository for three-dimensional structural data of biological macromolecules, primarily focusing on proteins and nucleic acids. The data archived in the PDB are freely accessible to researchers worldwide, enabling them to explore the structures of proteins, nucleic acids, and complex assemblies. Scientists use this information for a multitude of purposes, including understanding biological functions, drug discovery, protein engineering, and molecular modelling [23,24,25]. Further, to study the protein structure rasmol software has been utilized and then to analyze the ligand-protein complex interaction pymol has been used. RasMol is an influential and pioneering molecular visualization

ISSN: 0975-3583, 0976-2833 VOL14, ISSUE 12, 2023

program that has significantly contributed to the field of structural biology.RasMol facilitated the visualization of molecular structures by providing features to rotate, translate, and zoom into the three-dimensional representations of molecules. Users could manipulate these structures to examine different angles, surface properties, and structural details, aiding in the understanding of molecular interactions, folding patterns, and active sites [26,27]. PyMOL offers a vast array of tools for structural analysis and exploration. It allows users to measure distances, angles, and dihedral angles, perform alignments between structures, analyze electrostatic potentials, generate molecular surfaces, and visualize molecular dynamics trajectories [28,29].For further sequence similarity and phylogenetic analysis of protein BLAST and COBALT has been used, respectively. BLAST, which stands for Basic Local Alignment Search Tool, is a fundamental and widely used bioinformatics algorithm and software tool. It's designed to compare biological sequences, such as DNA, RNA, or protein sequences, against vast databases to identify similarities and infer functional, structural, or evolutionary relationships between sequences. BLAST outputs results in a format that includes statistical measures indicating the significance of matches found. This information helps researchers assess the likelihood that a match occurred by chance and allows them to prioritize and further investigate the most relevant matches [30,31,32]. COBALT employs a constraint-based approach for multiple sequence alignment. It gathers a set of pairwise constraints derived from various sources, including database searches, sequence similarity data, and user input. These constraints serve as guiding principles or rules that inform the alignment process [33]. For pathway analysis the KEGG software has been used. The Kyoto Encyclopedia of Genes and Genomes (KEGG) is a comprehensive resource and database widely used in bioinformatics and systems biology. Created by the Kanehisa Laboratories in Japan, KEGG provides a wealth of information about biological pathways, molecular interactions, diseases, drugs, and genomic information [34,35]. For the docking of protein and drug, CB Dock2 has been used. CB-Dock2 is a user-friendly web server designed specifically for blind docking, meaning it predicts binding modes without prior information about binding sites. CB-Dock2 showcased a notable success rate of around 85% for binding pose prediction, with root-mean-square deviation (RMSD) values less than 2.0 Å [36]. Further analysis of protein has been performed using ExpasyProtParam and PDBePISA.ExPASyProtParam is a web-based tool provided by the Expert Protein Analysis System (ExPASy), developed by the Swiss Institute of Bioinformatics (SIB). This tool is designed to compute various physicochemical properties and analyze protein sequences [37].PDBePISA (Proteins, Interfaces, Structures, and Assemblies) is a freely available online resource provided by the Protein Data Bank in Europe (PDBe). It's a web server designed for the analysis of macromolecular interfaces, quaternary structures, and assemblies. PDBePISA identifies and analyzes interfaces between macromolecular complexes, providing insights into the interactions between proteins, nucleic acids, and ligands [38]. Further Phyre2 tool has been used for secondary structure prediction and sequence similarity. Phyre2 (Protein Homology/analogY Recognition Engine V 2.0) is a web-based tool developed for protein structure prediction and analysis. It provides a user-friendly interface to predict protein structure, function, and sequence-structure relationships by employing homology modeling, ab initio folding, and protein threading techniques [39]. Then we have calculated protein interaction and try to find gene ontology. For that STRING tool has been used. STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) is a bioinformatics database and web-based tool that consolidates and predicts protein-protein interactions (PPIs) and functional associations across various organisms. It integrates known and predicted interactions, offering a comprehensive resource for analyzing protein interactions and their functional implications [40]. To check the quality of protein ERRAT and SAVES have been used. ERRAT is a bioinformatics tool designed for the evaluation of protein structures based

on atomic coordinates. It assesses the statistics of non-bonded atom-atom interactions, highlighting potential errors or irregularities within the structure. ERRAT computes a quality factor indicating the overall quality of the model by analyzing the distribution of atomic interactions compared to high-resolution structures. On the other hand, SAVES (Structural Analysis and Verification Server) is a suite of tools that collectively assess the accuracy and reliability of protein structures [41,42].

RESULT AND DISCUSSION

First we have to verify that the structure 6U9N which we are using is good enough or not to perform the research. For that we are using ERRAT2 and SAVES (Procheck) software.

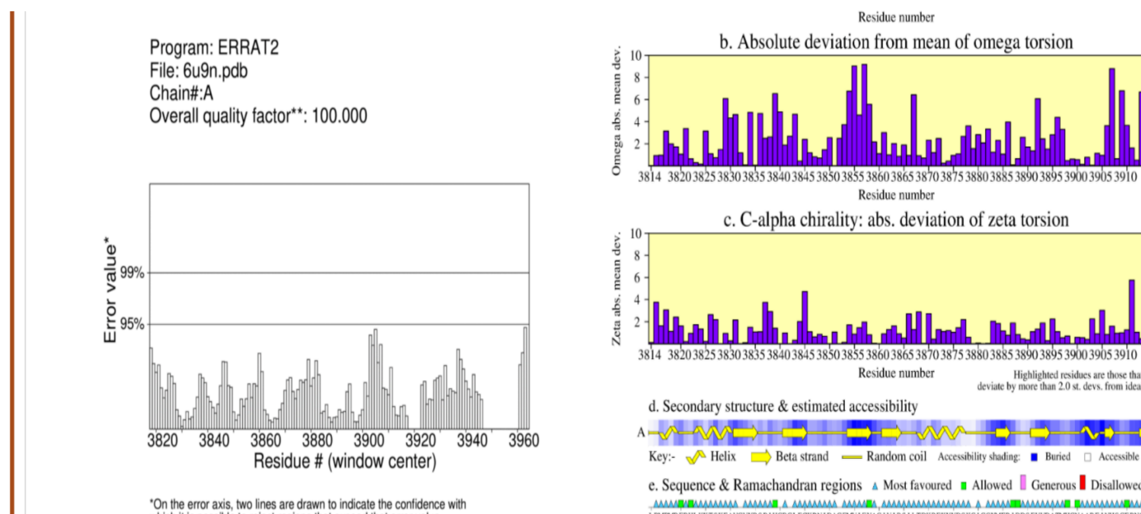


Figure1 : Validation of modelled protein Figure 2: Model validation by using SAVESstructure via ERRAT

The result is showing good ERRAT quality factor score and the SAVES result(Ramachandran plot) is showing 88.2% residues are in the most favoured region.

Further, structural analysis of protein has been performed using PyMol and RasMol.Surface representations aid in identifying potential binding sites or active sites on the molecule. By visualizing the surface, researchers can locate cavities, pockets, or regions likely to interact with other molecules, which is crucial for drug design or understanding molecular interactions.

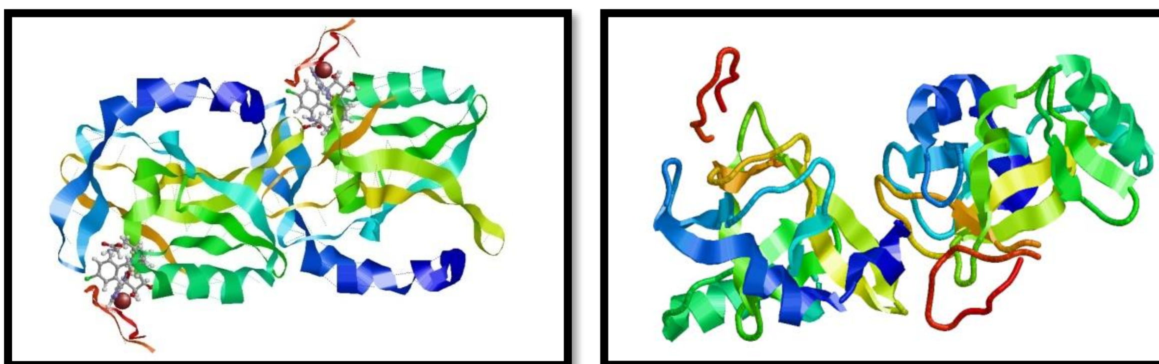


Figure 3 :Strucutur ereprsenting H-bond Figure 4 : Representing N and C terminal on (175)

ISSN: 0975-3583, 0976-2833 VOL14, ISSUE 12, 2023

Both results have been obtained using RasMol.

Visual representations of hydrogen bonds on residue 175 facilitate clearer communication of structural interactions within the molecule. Hydrogen bonds contribute significantly to the stability of protein structures. Displaying these bonds, especially concerning residue 175, helps in assessing the stability and rigidity of that region within the molecule.

Displaying N (Blue) and C (Red) loops accentuates the flexible and dynamic regions of the protein structure. Loops often play a significant role in protein function, such as in binding sites or regions critical for molecular recognition.

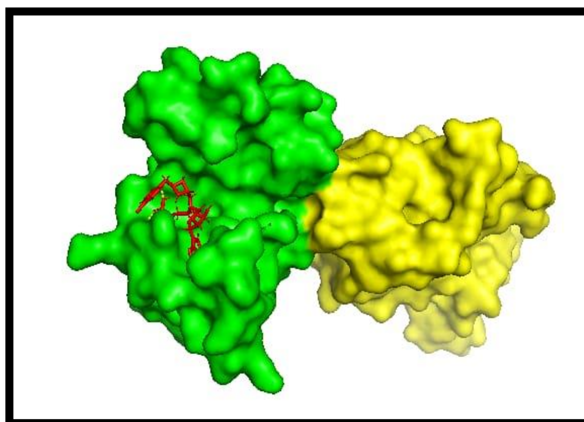


Figure 5 :Representation of chain A (yellow) 1192 atoms and chain B (green) 1191 atoms, Q34 ligand (red) using PyMol

Displaying different chains (A in yellow, B in green) helps identify and distinguish between distinct components or subunits within the macromolecular structure. Comparing the atom count between chain A (1192 atoms) and chain B (1191 atoms) provides insights into their structural differences or similarities, potentially indicating variations in conformation or content.

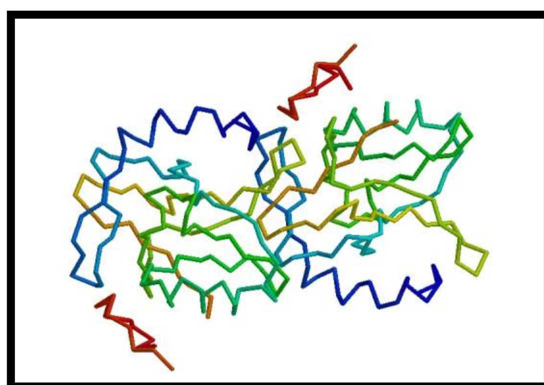


Figure 6 : Presentation of backbone representing to phosphorous nucleic acid using RasMol

Emphasizing the phosphorus atoms in the backbone provides a specific focus on these essential elements of the nucleic acid structure, which are pivotal in its stability and interactions.

ISSN: 0975-3583, 0976-2833 VOL14, ISSUE 12, 2023

By giving commands in the command line window of RasMol, user can explore how many atoms are comprised in the structure. The above data (output) has been found. For structure analysis, the 3D structure can be displayed according to the user and can similarly change the color.

TABLE 1: RasMol result

ATOM	NO OF ATOMS OBSERVED
OXYGEN	601
NITROGEN	421
CARBON	1458
ZINC	2
CALCIUM	NONE
SULFUR	24
HYDROGEN	66

Now, the sequence similarity analysis has been performed using BLAST. It's a statistical approach generating a bit score and E value per alignment score. The maximum score is derived by summing rewards for matched nucleotides and penalties for mismatches and gaps, while the total score represents the sum of alignment scores across all segments in the same subject sequence. Query coverage refers to the percentage of the query length within aligned segments. A lower E value, closer to zero, signifies a more meaningful match.

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Chain A, Histone-lysine N-methyltransferase 2A [Homo sapiens]	Homo sapiens	328	328	100%	3e-113	100.00%	158	5F5E_A
Chain C, Histone-lysine N-methyltransferase 2A [Homo sapiens]	Homo sapiens	330	330	99%	5e-113	100.00%	209	6PWV_C
hypothetical protein CIB84_013164 [Bambusicola thoracicus]	Bambusicola tho...	323	323	99%	4e-111	98.73%	158	POI23088.1
hypothetical protein CB1_000200041 [Camelus ferus]	Camelus ferus	324	324	99%	4e-111	98.73%	193	EPY88054.1
Chain K, Histone-lysine N-methyltransferase 2A [Homo sapiens]	Homo sapiens	325	325	99%	6e-111	98.73%	216	6KIU_K
MLL [Pan paniscus]	Pan paniscus	322	322	99%	6e-111	98.73%	162	ABM54292.1
MLL [Gorilla gorilla]	Gorilla gorilla	329	329	99%	6e-111	98.73%	338	ABM46716.1

Figure 7 :Protein description result in blast

These are the top results shown in the above given figure. These are the proteins that have shown sequence similarity with the 6U9N protein.

Next, for the analysis of protein ExPASyProtParam tool has been used.ProtParam calculates the overall composition of amino acids in a given protein sequence, providing counts and percentages of each amino acid.

Amino acid composition:

Ala	(A)	9	5.7%
Arg	(R)	13	8.2%
Asn	(N)	8	5.1%
Asp	(D)	10	6.3%
Cys	(C)	7	4.4%
Gln	(Q)	1	0.6%
Glu	(E)	9	5.7%
Gly	(G)	11	7.0%
His	(H)	5	3.2%
Ile	(I)	15	9.5%
Leu	(L)	7	4.4%
Lys	(K)	14	8.9%
Met	(M)	6	3.8%
Phe	(F)	7	4.4%
Pro	(P)	6	3.8%
Ser	(S)	9	5.7%
Thr	(T)	4	2.5%
Trp	(W)	0	0.0%
Tyr	(Y)	9	5.7%
Val	(V)	8	5.1%
Pyl	(O)	0	0.0%
Sec	(U)	0	0.0%

Figure 8: ExPasyProt Param result showing amino acid concentration in the protein

- **Total number of negatively charged residues (Asp + Glu):** 19
- **Total number of positively charged residues (Arg + Lys):** 27
- **Estimated half-life:**

- The N-terminal of the sequence considered is S (Ser).

- The estimated half-life is: 1.9 hours (mammalian reticulocytes, in vitro).

>20 hours (yeast, in vivo).

>10 hours (Escherichia coli, in vivo).

- **Aliphatic index:** 74.68

- **Grand average of hydropathicity (GRAVY):** -0.446

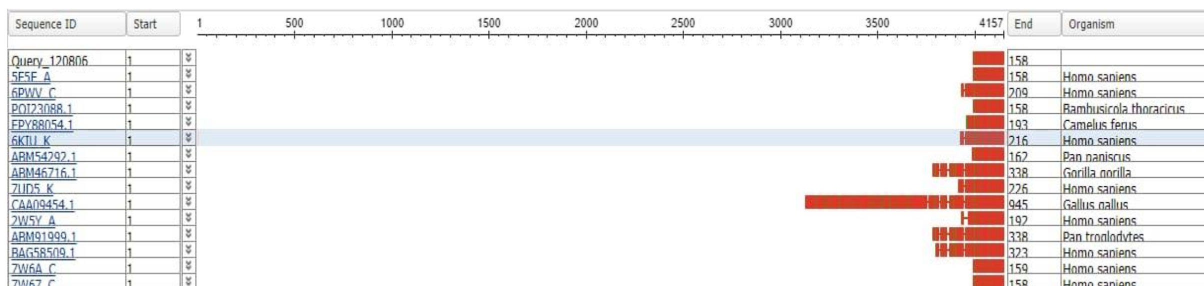
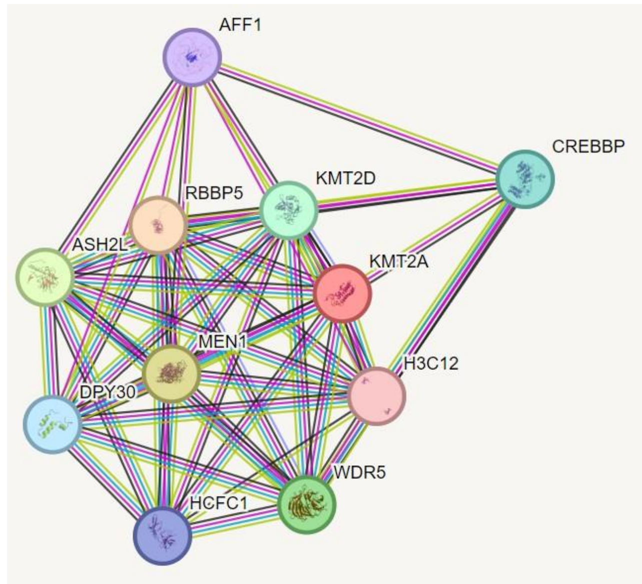


Figure 9 :Membrane preference result in cobalt

The Cobalt Hydropathy Scale evaluates amino acid residues' hydrophobic or hydrophilic characteristics in a protein sequence by assigning numeric values based on their relative tendencies. This method distinguishes between hydrophobic amino acids (depicted in red) and hydrophilic ones (depicted in blue).

Now the functional analysis of the protein is performed using STRING.

ISSN: 0975-3583, 0976-2833 VOL14, ISSUE 12, 2023



number of nodes: 11

number of edges: 48

average node degree: 8.73

avg. local clustering coefficient: 0.89

expected number of edges: 15

PPI enrichment p-value: 6.11e-12

Figure 10 :STRING result to predict protein function

This result suggests that the proteins exhibit a higher level of interactions among themselves compared to what would be anticipated in a random selection of proteins with similar size and distribution taken from the genome. This increased enrichment implies that these proteins, as a collective group, are likely interconnected at least to some extent biologically.

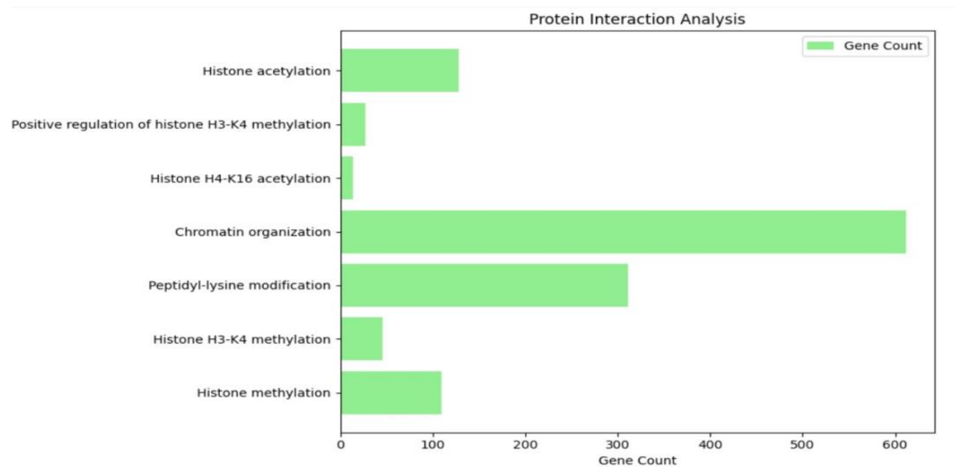


Figure 11: Graph of gene ontology data (Data obtained from STRING)

ISSN: 0975-3583, 0976-2833 VOL14, ISSUE 12, 2023

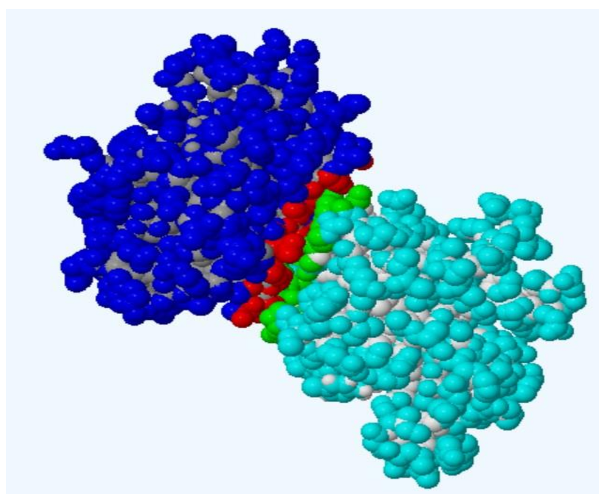


Figure 12: protein structure obtained through PDBePISA

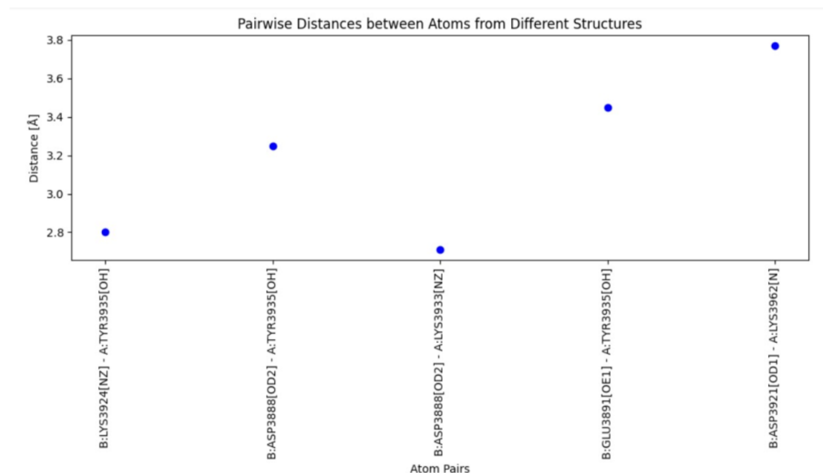
The molecule shown in red and green are the interface molecules.

TABLE 2: Detailes of salt bridge

Structure 1	Dist. [Å]	Structure 2
B:ASP3888[OD2]	2.71	A:LYS3933[NZ]

TABLE 3: Detailes of Hydrogen bonds between chain A and chain B

S.N.	Structure 1	Dist. [Å]	Structure 2
1	B:LYS3924[NZ]	2.80	A:TYR3935[OH]
2	B:ASP3888[OD2]	3.25	A:TYR3935[OH]
3	B:ASP3888[OD2]	2.71	A:LYS3933[NZ]
4	B:GLU3891[OE1]	3.45	A:TYR3935[OH]
5	B:ASP3921[OD1]	3.77	A:LYS3962[N]



ISSN: 0975-3583, 0976-2833 VOL14, ISSUE 12, 2023

Figure 13: Graphical representation of data obtained from PDBePISA

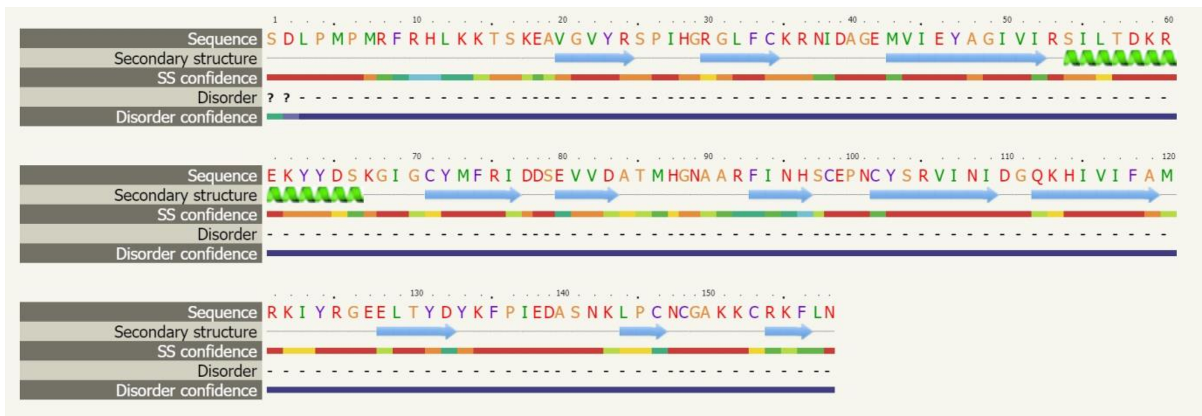


Figure 14: secondary structure prediction using Phyre2

Beta sheets are represented with light blue color and alpha helices are represented with green color.

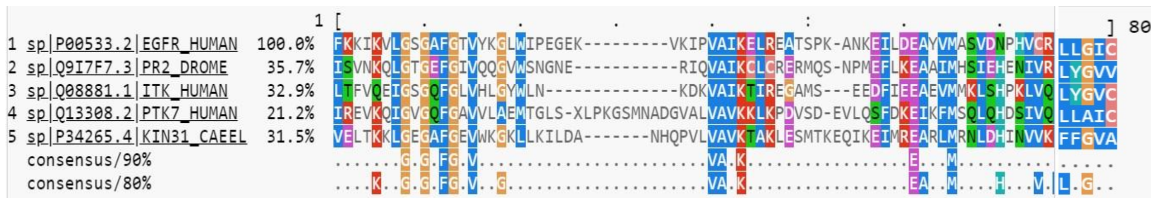


Figure 15: sequence similarity obtained using Phyre2

Pathway analysis has been performed to try to know the exact reason of the abnormailites causing by 6U9N protein.

ISSN: 0975-3583, 0976-2833 VOL14, ISSUE 12, 2023

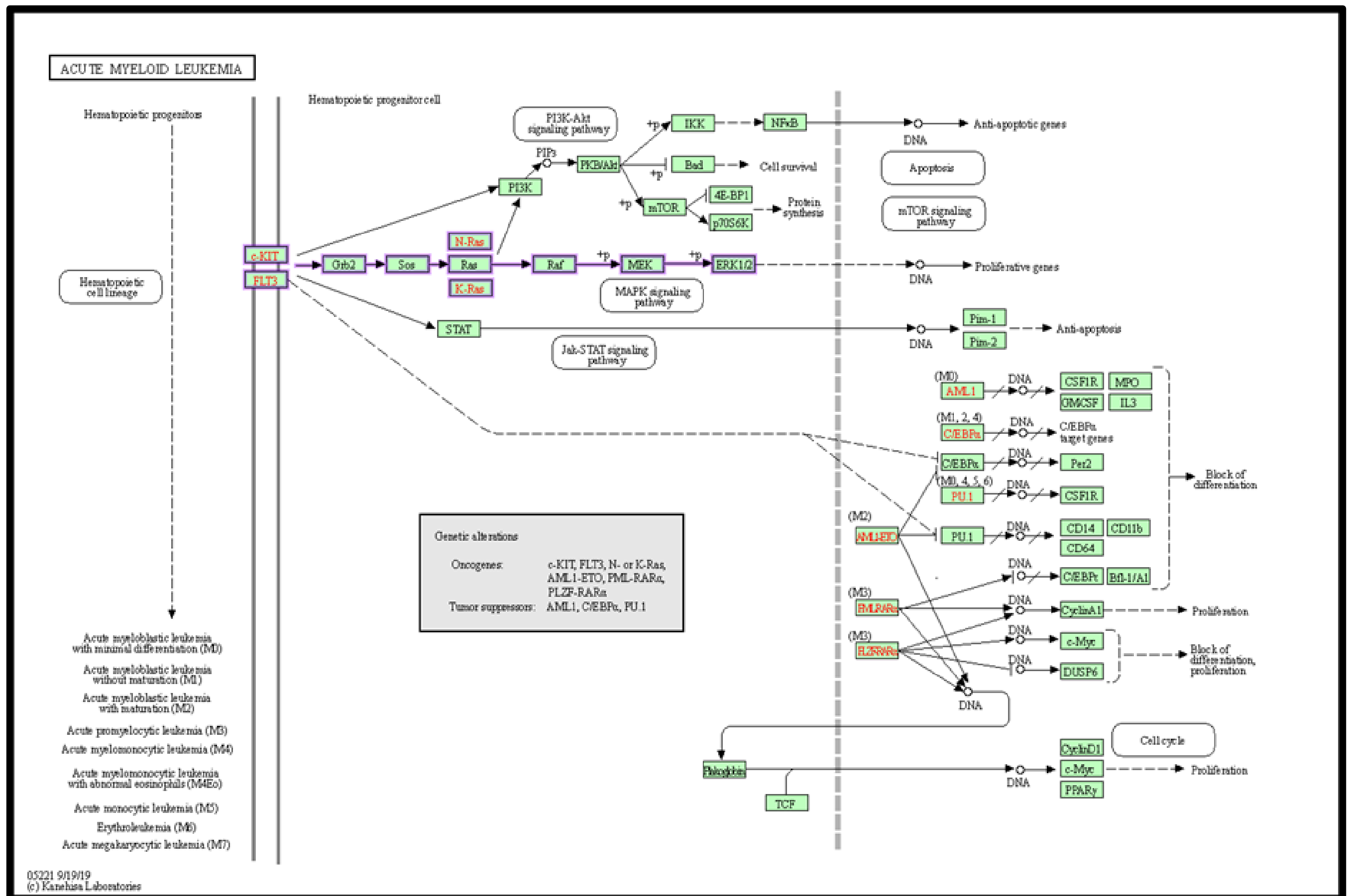


Figure 16: nt06210 ERK signaling pathway (cancer)

- N00003 Mutation-activated KIT to RAS-ERK signaling pathway
- N00004 Duplication or mutation-activated FLT3 to RAS-ERK signaling pathway
- N00012 Mutation-activated KRAS/NRAS to ERK signaling pathway

The another mutation pathway that has been observed from KEGG is –

- N00116 Mutation-inactivated RUNX1 to transcription
- N00111 AML1-ETO fusion to CEBPA-mediated transcription
- N00112 AML1-ETO fusion to PU.1-mediated transcription
- N00108 AML1-ETO fusion to transcriptional activation
- N00109 PML-RARA fusion to transcriptional activation
- N00113 PML-RARA fusion to transcriptional repression
- N00110 PLZF-RARA fusion to transcriptional activation
- N00114 PLZF-RARA fusion to transcriptional repression

ISSN: 0975-3583, 0976-2833 VOL14, ISSUE 12, 2023

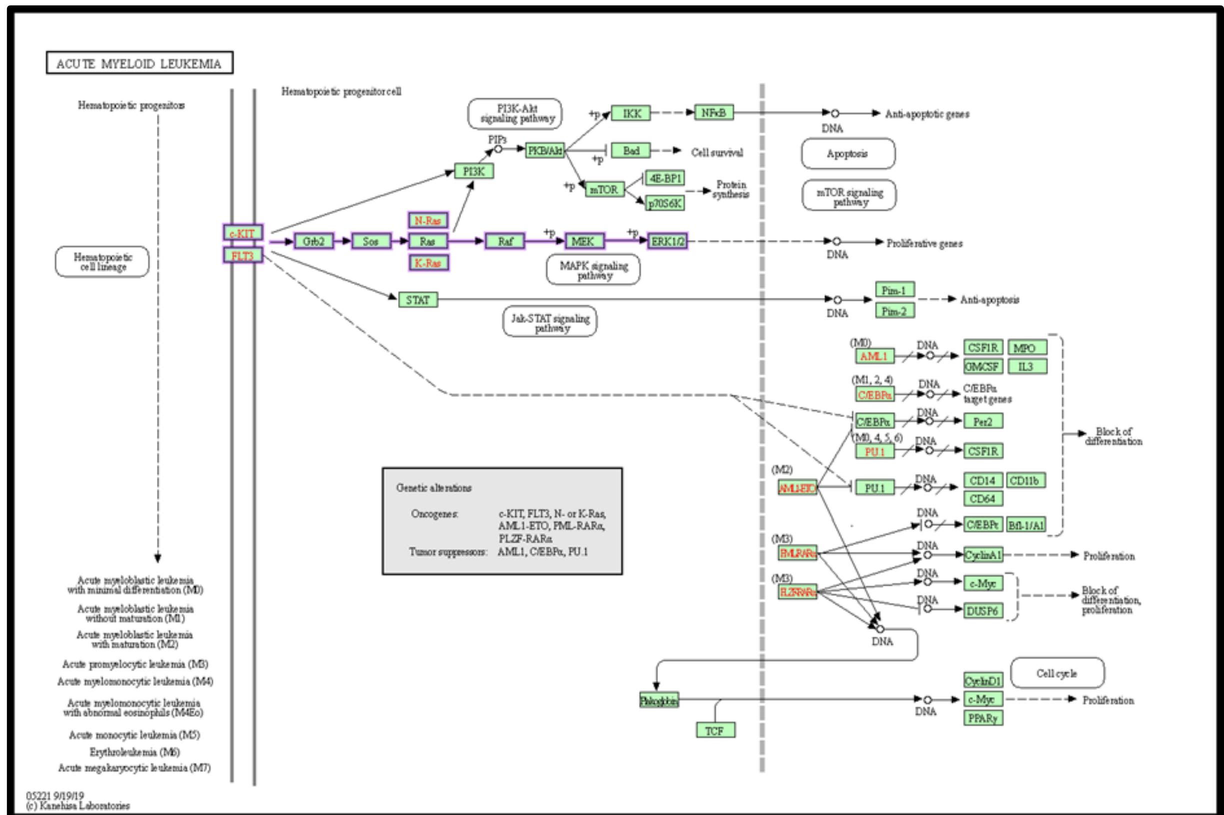
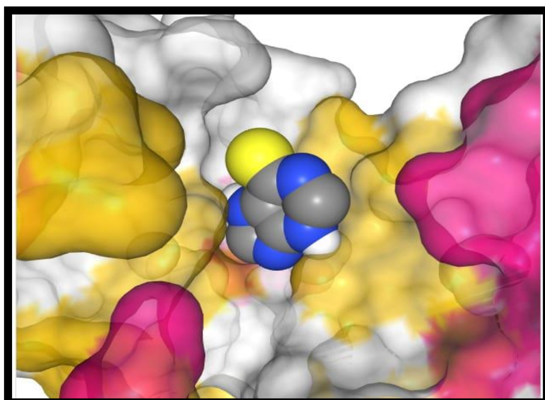


Figure 17 :nt06240 Transcription (cancer)

The Transcription (Cancer) pathway (nt06240) plays a pivotal role in the dysregulation of transcriptional processes associated with tumorigenesis and cancer progression. In this study, KEGG pathway analysis revealed notable alterations in key regulatory elements and genes within this pathway in the context of [specify experimental conditions or tumor type]. Several transcription factors, such as [specific TFs], exhibited significant dysregulation, potentially contributing to aberrant gene expression profiles observed in cancer cells. Moreover, core components involved in RNA polymerase activity and transcriptional regulation, including [specific genes/proteins], displayed altered expression or activity levels. The dysregulated pathway underscores the intricate interplay of transcriptional machinery in driving oncogenic processes, shedding light on potential therapeutic targets or diagnostic markers in cancer biology. Understanding the perturbations in this pathway may offer crucial insights into unraveling the transcriptional landscape associated with tumorigenesis.

Now the molecular docking has been performed using CBDOCK2.



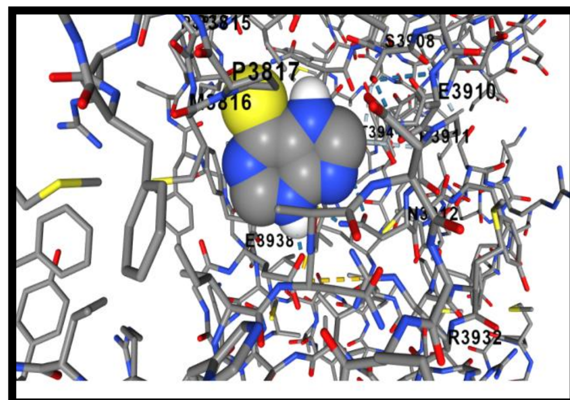


Figure 18 :Molecular docking result (Mercaptopurine)

TABLE 4: Docking score of protein-ligand of Mercaptopurine

CurPocket ID	Vina score	Cavity volume (Å^3)	Center (x, y, z)	Docking size (x, y, z)
C1	-5.3	1216	229, 175, 210	27, 16, 16
C3	-5.3	419	243, 173, 219	25, 16, 25
C2	-4.9	627	217, 189, 204	28, 16, 24
C4	-4.6	268	212, 197, 207	16, 16, 16
C5	-3.8	260	241, 189, 223	16, 16, 16

Chain A: PRO3837 ILE3838 SER3908 GLU3910 PRO3911 ASN3912 ARG3932 GLU3938 GLU3939 THR3941 CYS3959

Chain B: ILE3838 HIS3839 ASN3906 HIS3907 SER3908 TYR3944 PRO3956 CYS3957 ASN3958 CYS3959

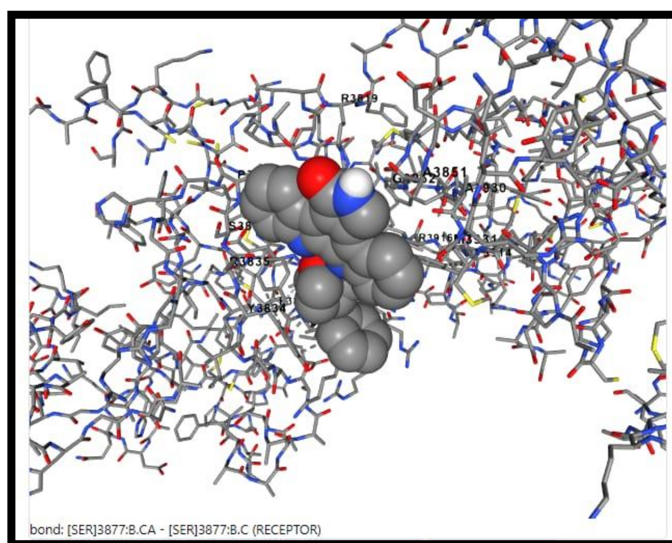


Figure 19 :Docking of protein-ligand of Midostaurin

Figure showing the protein-ligand docking

Chain A: LEU3823 LYS3824 SER3827 LYS3828 VAL3831 GLY3832 VAL3833 LEU3843 LYS3846 ILE3856 GLU3857 ALA3859 ASN3900 ALA3901 ARG3903

Chain B: HIS3839 GLY3840 ARG3841 GLY3842 TYR3858 THR3868 GLU3872 CYS3882 TYR3883 MET3884 PHE3885 ARG3886 VAL3892 ARG3903 PHE3904 ILE3905 ASN3906 TYR3942 TYR3944

ISSN: 0975-3583, 0976-2833 VOL14, ISSUE 12, 2023

TABLE 5: Docking score of protein-ligand of Midostaurin

CurPocket ID	Vina score	Cavity volume (Å^3)	Center (x, y, z)	Docking size (x, y, z)
C4	-9.2	268	212, 197, 207	24, 24, 24
C2	-9.1	627	217, 189, 204	24, 24, 24
C3	-9.0	419	243, 173, 219	24, 24, 24
C5	-7.8	260	241, 189, 223	24, 24, 24
C1	-7.7	1216	229, 175, 210	24, 24, 24

Studying acute myeloid leukemia is important for understanding its molecular mechanisms, improving early diagnosis, and developing targeted therapies. This can ultimately lead to more effective treatment and potential cures for this aggressive form of leukemia, resulting in better patient outcomes and quality of life. In this project, our goal was to conduct a comprehensive computational analysis of acute myeloid leukemia using a human blood sample. We started by obtaining the protein sample from the protein database (PDB) and downloading the molecular structure in a common file format. We then visualized the structure using RasMol software, a molecular visualization tool commonly used for analyzing and understanding the three-dimensional structures of biological macromolecules such as proteins and nucleic acids. RasMol allowed us to explore the graphical representation of the protein's three-dimensional structure using various visualization styles, including wireframe, ball and stick, space-filling, ribbon, and cartoon. The software also allowed us to color, select, zoom, and highlight specific atoms of interest, making it a valuable tool in our analysis. In addition to RasMol, we also utilized PyMOL software for visualizing the structure and examining the protein-ligand binding. PyMOL is another powerful molecular visualization software widely used in the field of structural biology. With its advanced rendering and analysis capabilities, PyMOL enabled us to analyze and manipulate three-dimensional structures of biomolecules. Its user-friendly interface and scripting options made it easy to communicate complex molecular structures. One of the features in PyMOL that we used is called "roving density," which allows the calculation and visualization of local electron density maps around selected atoms or regions. This feature helped us analyze protein-ligand interactions and structural features within molecular complexes by studying the distribution of electron density. PyMOL also supports volume rendering, a technique used to visualize and analyze three-dimensional density maps, such as electron density maps obtained from X-ray crystallography or cryo-electron microscopy. This feature allowed us to represent molecular surfaces and visualize spatial distributions of electron density within biological macromolecules, providing valuable insights into their structural features and functional mechanisms. For further analysis, we performed homology modeling of the protein sequence using the BLAST tool. Furthermore, we utilized the KEGG pathway database to analyze various signaling pathways, such as the mutation-activated pathway and transcription pathway, among others. We also conducted multiple sequence alignments and analyses using the COBALT server, which provided us with a result indicating the membrane

ISSN: 0975-3583, 0976-2833 VOL14, ISSUE 12, 2023

preference. Finally, we employed CB-DOCK (ClusproBiomolecular Docking Server) for molecular docking to study the interaction between human acute myeloid leukemia and two associated ligands: mercaptopurine and midostaurin as potential future treatment options. CB-DOCK is a web-based platform that enables protein-protein and protein-ligand docking simulations, allowing researchers to predict and analyze the interactions between biomolecules. Through the use of a clustering algorithm, CB-DOCK provided us with refined and representative models of potential molecular complexes. This process allowed us to successfully perform structure-based molecular docking.

CONCLUSION

In summary, this study focuses on the analysis of a 3D model and prediction of the active site in acute myeloid leukemia. The molecular analysis and docking studies performed in this research provide a deeper understanding of myeloid leukemia at a molecular level and offer potential ideas for the development of targeted treatments. These findings highlight the significance of early detection, genetic analysis, and computational approaches in advancing research on acute myeloid leukemia and enhancing patient outcomes. Further investigations and clinical trials are necessary to confirm the effectiveness of the identified ligands and investigate their potential therapeutic use in treating acute myeloid leukemia cancer.

REFERENCES

1. Stone, R. M., O'Donnell, M. R., &Sekeres, M. A. (2004). Acute myeloid leukemia. *Hematology. American Society of Hematology. Education Program*, 98–117. https://doi.org/10.1182/asheducation-2004.1.98
2. Rubnitz, J. E., Gibson, B., & Smith, F. O. (2010). Acute myeloid leukemia. *Hematology/oncology clinics of North America*, *24*(1), 35–63. https://doi.org/10.1016/j.hoc.2009.11.008
3. Rubnitz, J. E., Inaba, H., Leung, W. H., Pounds, S., Cao, X., Campana, D., Ribeiro, R. C., &Pui, C. H. (2014). Definition of cure in childhood acute myeloid leukemia. *Cancer*, *120*(16), 2490–2496. https://doi.org/10.1002/cncr.28742
4. Vakiti A, Mewawalla P. Acute Myeloid Leukemia. [Updated 2023 Aug 8]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan-. https://www.ncbi.nlm.nih.gov/books/NBK507875/
5. Döhner, H., Weisdorf, D. J., & Bloomfield, C. D. (2015). Acute Myeloid Leukemia. *The New England journal of medicine*, *373*(12), 1136–1152. https://doi.org/10.1056/NEJMra1406184
6. Khwaja, A., Bjorkholm, M., Gale, R. E., Levine, R. L., Jordan, C. T., Ehninger, G., Bloomfield, C. D., Estey, E., Burnett, A., Cornelissen, J. J., Scheinberg, D. A., Bouscary, D., &Linch, D. C. (2016). Acute myeloid leukaemia. *Nature reviews. Disease primers*, *2*, 16010. https://doi.org/10.1038/nrdp.2016.10
7. Bennett, J. M., Catovsky, D., Daniel, M. T., Flandrin, G., Galton, D. A., Gralnick, H. R., & Sultan, C. (1985). Proposed revised criteria for the classification of acute myeloid leukemia. A report of the French-American-British Cooperative Group. *Annals of internal medicine*, *103*(4), 620–625. https://doi.org/10.7326/0003-4819-103-4-620
8. Vardiman, James & Thiele, Jüergen& Arber, Daniel &Brunning, Richard &Borowitz, Michael &Porwit, Anna & Harris, Nancy & Beau, Michelle &Hellström-Lindberg, Eva &Tefferi, Ayalew& Bloomfield, Clara. (2009). Vardiman JW, Thiele J, Arber DA, Brunning RD, Borowitz MJ, Porwit A, Harris NL, Le Beau MM, Hellstrom-Lindberg E, Tefferi A, Bloomfield CD. The 2008 revision of the World Health

Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. Blood. 114: 937-951. Blood. 114. 937-51. 10.1182/blood-2009-03-209262.

9. Rio-Machin, A., Vulliamy, T., Hug, N., Walne, A., Tawana, K., Cardoso, S., Ellison, A., Pontikos, N., Wang, J., Tummala, H., Al Seraihi, A. F. H., Alnajar, J., Bewicke-Copley, F., Armes, H., Barnett, M., Bloor, A., Bödör, C., Bowen, D., Fenaux, P., Green, A., … Dokal, I. (2020). The complex genetic landscape of familial MDS and AML reveals pathogenic germline variants. *Nature communications*, *11*(1), 1044. https://doi.org/10.1038/s41467-020-14829-5
10. Caldwell, J. T., Ge, Y., &Taub, J. W. (2014). Prognosis and management of acute myeloid leukemia in patients with Down syndrome. *Expert review of hematology*, *7*(6), 831–840. https://doi.org/10.1586/17474086.2014.959923
11. Rosenauer, A., Raelson, J. V., Nervi, C., Eydoux, P., DeBlasio, A., & Miller, W. H., Jr (1996). Alterations in expression, binding to ligand and DNA, and transcriptional activity of rearranged and wild-type retinoid receptors in retinoid-resistant acute promyelocyticleukemia cell lines. *Blood*, *88*(7), 2671–2682.
12. Shen, Z. X., Chen, G. Q., Ni, J. H., Li, X. S., Xiong, S. M., Qiu, Q. Y., Zhu, J., Tang, W., Sun, G. L., Yang, K. Q., Chen, Y., Zhou, L., Fang, Z. W., Wang, Y. T., Ma, J., Zhang, P., Zhang, T. D., Chen, S. J., Chen, Z., & Wang, Z. Y. (1997). Use of arsenic trioxide (As2O3) in the treatment of acute promyelocyticleukemia (APL): II. Clinical efficacy and pharmacokinetics in relapsed patients. *Blood*, *89*(9), 3354–3360.
13. Kadia, T. M., Faderl, S., Ravandi, F., Jabbour, E., Garcia-Manero, G., Borthakur, G., Ferrajoli, A., Konopleva, M., Burger, J., Huang, X., Wang, X., Pierce, S., Brandt, M., Feliu, J., Cortes, J., &Kantarjian, H. (2015). Final results of a phase 2 trial of clofarabine and low-dose cytarabine alternating with decitabine in older patients with newly diagnosed acute myeloid leukemia. *Cancer*, *121*(14), 2375–2382. https://doi.org/10.1002/cncr.29367
14. Grimwade, D., Hills, R. K., Moorman, A. V., Walker, H., Chatters, S., Goldstone, A. H., Wheatley, K., Harrison, C. J., Burnett, A. K., & National Cancer Research Institute Adult Leukaemia Working Group (2010). Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. *Blood*, *116*(3), 354–365. https://doi.org/10.1182/blood-2009-11-254441
15. Döhner, H., Estey, E. H., Amadori, S., Appelbaum, F. R., Büchner, T., Burnett, A. K., Dombret, H., Fenaux, P., Grimwade, D., Larson, R. A., Lo-Coco, F., Naoe, T., Niederwieser, D., Ossenkoppele, G. J., Sanz, M. A., Sierra, J., Tallman, M. S., Löwenberg, B., Bloomfield, C. D., & European LeukemiaNet (2010). Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood*, *115*(3), 453–474. https://doi.org/10.1182/blood-2009-07-235358
16. Daver, N., Schlenk, R. F., Russell, N. H., & Levis, M. J. (2019). Targeting FLT3 mutations in AML: review of current knowledge and evidence. *Leukemia*, *33*(2), 299–312. https://doi.org/10.1038/s41375-018-0357-9
17. Cui, W., Aouidate, A., Wang, S., Yu, Q., Li, Y., & Yuan, S. (2020). Discovering Anti-Cancer Drugs *via* Computational Methods. *Frontiers in pharmacology*, *11*, 733. https://doi.org/10.3389/fphar.2020.00733
18. Carter, J. L., Hege, K., Yang, J., Kalpage, H. A., Su, Y., Edwards, H., Hüttemann, M., Taub, J. W., &Ge, Y. (2020). Targeting multiple signaling pathways: the new

approach to acute myeloid leukemia therapy. *Signal transduction and targeted therapy*, *5*(1), 288. https://doi.org/10.1038/s41392-020-00361-x

19. Farrar, J. E., Schuback, H. L., Ries, R. E., Wai, D., Hampton, O. A., Trevino, L. R., Alonzo, T. A., Guidry Auvil, J. M., Davidsen, T. M., Gesuwan, P., Hermida, L., Muzny, D. M., Dewal, N., Rustagi, N., Lewis, L. R., Gamis, A. S., Wheeler, D. A., Smith, M. A., Gerhard, D. S., &Meshinchi, S. (2016). Genomic Profiling of Pediatric Acute Myeloid Leukemia Reveals a Changing Mutational Landscape from Disease Diagnosis to Relapse. *Cancer research*, *76*(8), 2197–2205. https://doi.org/10.1158/0008-5472.CAN-15-1015
20. Christine M. Segeren, Mars B. van 't Veer, The FAB classification for acute myeloid leukaemia is it outdated, The Netherlands Journal of Medicine, Volume 49, Issue 3,1996,Pages 126-131,ISSN 0300-2977,https://doi.org/10.1016/0300-2977(96)00024-1.
21. Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B. A., Wang, J., Yu, B., Zhang, J., & Bryant, S. H. (2016). PubChem Substance and Compound databases. *Nucleic acids research*, *44*(D1), D1202–D1213. https://doi.org/10.1093/nar/gkv951
22. Wang, Y., Bolton, E., Dracheva, S., Karapetyan, K., Shoemaker, B. A., Suzek, T. O., Wang, J., Xiao, J., Zhang, J., & Bryant, S. H. (2010). An overview of the PubChemBioAssay resource. *Nucleic acids research*, *38*(Database issue), D255–D266. https://doi.org/10.1093/nar/gkp965
23. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic acids research*, *28*(1), 235–242. https://doi.org/10.1093/nar/28.1.235
24. Westbrook, J., Feng, Z., Jain, S., Bhat, T. N., Thanki, N., Ravichandran, V., Gilliland, G. L., Bluhm, W., Weissig, H., Greer, D. S., Bourne, P. E., & Berman, H. M. (2002). The Protein Data Bank: unifying the archive. *Nucleic acids research*, *30*(1), 245–248. https://doi.org/10.1093/nar/30.1.245
25. Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., Feng, Z., Gilliland, G. L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J. D., &Zardecki, C. (2002). The Protein Data Bank. *Actacrystallographica. Section D, Biological crystallography*, *58*(Pt 6 No 1), 899–907. https://doi.org/10.1107/s0907444902003451
26. Roger A. Sayle, E.James Milner-White, RASMOL: biomolecular graphics for all,Trends in Biochemical Sciences,Volume 20, Issue 9, 1995,Pages 374-376, ISSN 0968-0004, https://doi.org/10.1016/S0968-0004(00)89080-5.
27. J.Tony Pembroke, Bio-molecular modelling utilising RasMol and PDB resources: a tutorial with HEW lysozyme, Biochemistry and Molecular Biology Education, Volume 28, Issue 6, 2000, Pages 297-300, ISSN 1470-8175, https://doi.org/10.1016/S1470-8175(00)00050-3.
28. Seeliger, D., & de Groot, B. L. (2010). Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *Journal of computer-aided molecular design*, *24*(5), 417–422. https://doi.org/10.1007/s10822-010-9352-6
29. Mura, C., McCrimmon, C. M., Vertrees, J., &Sawaya, M. R. (2010). An introduction to biomolecular graphics. *PLoS computational biology*, *6*(8), e1000918. https://doi.org/10.1371/journal.pcbi.1000918
30. McGinnis, S., & Madden, T. L. (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic acids research*, *32*(Web Server issue), W20–W25. https://doi.org/10.1093/nar/gkh435

ISSN: 0975-3583, 0976-2833 VOL14, ISSUE 12, 2023

31. Boratyn, G. M., Camacho, C., Cooper, P. S., Coulouris, G., Fong, A., Ma, N., Madden, T. L., Matten, W. T., McGinnis, S. D., Merezhuk, Y., Raytselis, Y., Sayers, E. W., Tao, T., Ye, J., &Zaretskaya, I. (2013). BLAST: a more efficient report with usability improvements. *Nucleic acids research*, *41*(Web Server issue), W29–W33. https://doi.org/10.1093/nar/gkt282
32. Mark Johnson, Irena Zaretskaya, Yan Raytselis, Yuri Merezhuk, Scott McGinnis, Thomas L. Madden, NCBI BLAST: a better web interface, *Nucleic Acids Research*, Volume 36, Issue suppl_2, 1 July 2008, Pages W5–W9, https://doi.org/10.1093/nar/gkn201
33. Jason S. Papadopoulos, RichaAgarwala, COBALT: constraint-based alignment tool for multiple protein sequences, *Bioinformatics*, Volume 23, Issue 9, May 2007, Pages 1073–1079, https://doi.org/10.1093/bioinformatics/btm076
34. Kanehisa, M., &Goto, S. (2000). KEGG: kyotoencyclopedia of genes and genomes. *Nucleic acids research*, *28*(1), 27–30. https://doi.org/10.1093/nar/28.1.27
35. Minoru Kanehisa, Miho Furumichi, Mao Tanabe, Yoko Sato, KanaeMorishima, KEGG: new perspectives on genomes, pathways, diseases and drugs, *Nucleic Acids Research*, Volume 45, Issue D1, January 2017, Pages D353–D361, https://doi.org/10.1093/nar/gkw1092
36. Liu, Y., Yang, X., Gan, J., Chen, S., Xiao, Z. X., & Cao, Y. (2022). CB-Dock2: improved protein-ligand blind docking by integrating cavity detection, docking and homologous template fitting. *Nucleic acids research*, *50*(W1), W159–W164. https://doi.org/10.1093/nar/gkac394
37. Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., &Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic acids research*, *31*(13), 3784–3788. https://doi.org/10.1093/nar/gkg563
38. https://www.ebi.ac.uk/pdbe/pisa/pistart.html
39. Kelley, L., Mezulis, S., Yates, C. *et al.* The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**, 845–858 (2015). https://doi.org/10.1038/nprot.2015.053
40. Damian Szklarczyk, Andrea Franceschini, Stefan Wyder, KristofferForslund, Davide Heller, Jaime Huerta-Cepas, Milan Simonovic, Alexander Roth, Alberto Santos, Kalliopi P. Tsafou, Michael Kuhn, Peer Bork, Lars J. Jensen, Christian von Mering, STRING v10: protein–protein interaction networks, integrated over the tree of life, *Nucleic Acids Research*, Volume 43, Issue D1, 28 January 2015, Pages D447–D452, https://doi.org/10.1093/nar/gku1003
41. Al-Khayyat, M. Z., & Al-Dabbagh, A. G. (2016). In silico Prediction and Docking of Tertiary Structure of LuxI, an Inducer Synthase of Vibrio fischeri. *Reports of biochemistry & molecular biology*, *4*(2), 66–75.
42. Messaoudi, A., Belguith, H., & Ben Hamida, J. (2013). Homology modeling and virtual screening approaches to identify potent inhibitors of VEB-1 β-lactamase. *Theoretical biology & medical modelling*, *10*, 22. https://doi.org/10.1186/1742-4682-10-22
43. Dr Uma kumari, Devanshi Gupta, In silico RNA aptamer drug design and modelling,2022/4, Journal-JETIR,Volume-9, Issue-4, Pages 718-725
44. Uma kumari, NavjotKaurVirk, Identification of new potential drug for lung adenocarcinoma causing protein RMB10 using computer aided drug design approach, Publication date- 2022/6/11

ISSN: 0975-3583, 0976-2833 VOL14, ISSUE 12, 2023

45. Uma kumari, Shruti Gupta, NGS and Computer aided drug designing approach for prospective grade iv glioblastoma , JETIR 2023, Journal-IJBTR, Volume-10, Issue-5, Pages 280-290
46. Uma kumari, Devanshi Gupta,Computational analysis of Glioma Transcriptome and proteome with Biopython ,June 2023/4, Journal-IJBTR,Volume-13, Issue-1, Pages 1-14
47. Uma kumari,Saptarshi Mukherjee *etal;*Computationalanalysis for dementia Drug target identification and optimization,Corrosion and protection,Vol.51 Issue.2(2023)
48. Vinita Kukreja, Uma Kumari, "Genome Annotation of Brain Cancer and Structure Analysis by applying Drug Designing Technique", International Journal of Emerging Technologies and Innovative Research, 9(5);473-k479, May, 2022.