International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

An Integrated Next Generation Sequence Analysis Of Blood Cancer Along With Biopython

Kiranmai Venkatagiri[1], Uma Kumari[2]

[1]Trainee at Bioinformatics Project and Research Institute, Noida-201301, India
[2]Senior Bioinformatics Scientist, Bioinformatics Project and Research Institute, Noida -201301, India

Abstract

Blood cancers, including leukaemia, lymphoma, and myeloma, are among the most prevalent haematological malignancies. PDE6D, a key regulator of oncogenic RAS signalling, has been implicated in cancer progression. This study integrates Next-Generation Sequencing (NGS) with computational bioinformatics to analyse PDE6D mutations and their role in blood cancer pathogenesis. The 7PAC protein structure (PDE6D) was retrieved from the Protein Data Bank and analysed using PyMOL for structural visualization. Sequence alignment and homology studies were conducted using BLAST and Cluster Omega. Root Mean Square Deviation (RMSD) analysis assessed structural variations, while protein enrichment analysis and pathway analysis explored PDE6D's role in oncogenic signalling. B-factor analysis identified dynamic regions susceptible to mutation. BioPython was employed for genomic data analysis and functional annotation. STRING analysis revealed significant PDE6D interactions, particularly with RPGR, suggesting its involvement in cellular processes. KEGG pathway analysis confirmed its role in RAS-driven leukaemia. Root mean square analysis indicated high structural conservation, while B-factor analysis highlighted mutation-prone regions. Bioinformatics analysis suggested PDE6D mutations may contribute to dysregulated RAS signalling in blood cancers. This study highlights PDE6D as a potential biomarker and therapeutic target in blood cancer. By integrating NGS with bioinformatics tools, we identified critical genetic variations affecting its oncogenic role. Further in vitro and in vivo studies are required to validate PDE6D-targeted therapies, advancing precision medicine for blood cancer treatment.

Keywords: NGS; Structure Analysis; Blood cancer; BioPython; Proteomics; Mutations,Sequence Alignment

Introduction

Cancer is a disease where body cells grow uncontrollably in any part of body their by result in spreading to other areas through metastasis. In general terms Cancer can be defined as an event that occurs due to bypass of apoptosis, a phenomenon where normal cells undergo programmed cell death [1]. The known pathology for development of cancer is genetic mutations hence we can term cancer as a genetic disease that is caused due to genomic changes during cell division, DNA damage due to physical or chemical agents and inheritance from parents. Three main types of genes are involved in development of cancer. Over expression of Proto-oncogenes results in oncogenesis and mutations in tumor suppressor genes and DNA repair genes also results in development of cancer, hence these are sometimes termed as drivers of Cancer [2]. According to a study conducted by ICMR-Bengaluru, estimated cancer cases for the year

2022 was found to be 14, 61, 427 (100.4 per 1lakh population). Higher Cancer incidence was seen in women, breast cancer being the leading cause. In men lung cancer remains the most common followed by prostate cancer. India at present ranks third in Cancer prevalence after China and the United States of America. It is estimated that by 2025 there will be a 27.7% rise in cancer cases [3]. Blood cancers are the most common hematological malignancies that are mainly of three types. Leukemia, a type of blood cancer that affects blood components like white blood cells and bone marrow; Lymphoma, a blood cancer related to lymphatic system and myeloma, a blood cancer relating to plasma cells. According to 2019 global tumor burden report leukemia has highest incidence (8.22) and mortality rate (4.26) and people above 70 years of age have more risk of developing blood cancer (leukemia) [4]. The most common symptoms of blood cancer includes abdominal pain especially in the upper abdomen, bone or joint pain, easy bleeding or bruising, enlarged liver and glands such as spleen and lymph nodes, fatigue, frequent infections and unexplained weight loss. Acute lymphocytic lymphoma (ALL) affects the lymphocytes when exposed to toxins like benzene and radiation, chemotherapy and chromosomal abnormality can increase the risk of ALL. It is the most common cancer in children, with peak incidence between ages 2 and 9. Other leukaemia's are Acute mylogenous leukaemia (AML), Chronic lymphocytic leukaemia (CLL), Chronic mylogenous leukaemia (CML), Hairy cell leukaemia (HCL) [5]. According to Blood Cancer UK, annual morality is high for haematological malignancies when compared to prostate or breast cancer as diagnosis is difficult due to non-specific heterogeneous symptoms and one third of patients are diagnosed with blood cancer via emergency admission in the hospital and the survival rate of such admissions is only 40% for a period of three years [6]. A study conducted by The Cancer Genome Atlas (TCGA) on 200 AML patients identified 23 genes that were more frequently mutated than expected resulting in complication of analysis of genetic changes hence deep understanding of spectrum of driver genes mutations in large cohort populations is necessary to identify relevant treatment strategies apart of conventional approaches like surgery, chemotherapy and radiotherapy. Klaus H. Metzeler et al in their study analysed 664 adult AML patients for genetic alterations and concluded that presence of multiple mutations can influence the treatment outcomes in patients and hence their work emphasized on integrating genetic profiling into clinical practice for tailored therapies [7]. Over last few decades the Cancer research has been profoundly revolutionised that resulted in better classification of cancer types, diagnosis and treatment. When compared to solid tumours diagnosis of blood cancers is approachable easily due to intrinsic features of such neoplastic cells. Such cells do not have intercellular junctions thus making it possible to perform single cell analysis and moreover diagnosis involves minimal invasive procedures like collection of blood sample or iliac crest bone marrow aspiration. After the sampling we perform immune phenotyping by flow cytometry and conventional karyotype and fluorescence by in-situ hybridisation. In such cases Clinical Next Generation Sequencing (NGS) acts as adjuvant to diagnostic haematopathology [8]. NGS is a latest technology that sequences DNA and RNA by combining bioinformatics with sequencing chemistries and matrices allowing us for enormous parallel sequencing of various lengths of DNA and RNA. It is mostly used as a diagnostic aid in detection of variants and mutations in genetic diseases specifically in cancer, shedding its light on driver mutations. The DNA is usually fragmented into desired base-pair length which provides much more detailed information than Polymerase Chain Reaction (PCR) [9]. Unknown genetic variants can be identified effectively by sequencing methods and by employing NGS analysis a whole genome can be sequenced in a very short span with only $1000 which makes it more reliable than traditional sanger-sequencing. The applications of NGS in Cancer research include

identification of novel mutations in renal, lung, prostate and blood cancer. When NGS combined with Whole genome sequencing novel PML-RARA genetic recombinant was identified that was responsible for acute pro-myelocytic leukaemia which would not be possible by cytogenic techniques. Apart from diagnostic view NGS has wider applications in personalised medicine, that includes targeted genes sequencing, transcriptomic analysis, data validation and clinical interpretation followed by individualisation of Cancer therapy based on genetics of patient and tumour including a follow-up based on risk assessment and relapse of hereditary cancers. Overall early detection of diseases is achieved by employing NGS as a part of precision medicine [10, 11]. This study also employed the use of BioPython for data analysis that usually is an open source library for computational biology that can be applied in genomics, proteomics, transcriptomics and structural and functional bioinformatics. BioPython can perform multiple sequence alignment along with data visualisation making it easier for researchers in understandings large chunks of data in a very simplified form [12,29,31,32]. BioPython also implies sophisticated coding that helps in identification of drug target in view of blood cancer related genomic functionalities by means of multiple sequencing alignment analysis of nucleotide and protein sequences giving us the insights on mutations occurred in the sample. By using this information we can illustrate the 3D structure of proteins their by perform molecular docking techniques to find the drug target their by design the drug according to its actual binding site. This cascade results in effective implementation of personalised therapy to patients with blood cancer by which most of the adverse events caused due to use of trail and error of chemotherapeutic agents can be reduced their by ensuring patient safety [13]. This study is based on the RAS driven mutations and sheds its light on integration of NGS with BioPython in analysis of blood cancer. RAS, ”Rat Sarcoma” is a proto oncogene family composed of 36 human genes of which KRAS, HRAS, NRAS are most prominent in development of cancers in humans. These proteins are monomeric enzymes that require GTPase and bind to both GTP and GDP. RAS when bound to GTP looks like a coiled spring that in turn activates effector proteins like RAF, BRAF, P13K there by activating cascade of reactions resulting in regulation of cell growth in a normal pace as well as malignant proliferative phase [14]. Most mutations usually occur at codons 12, 13, and 61 and RAS isoforms are linked with HVR (Hypervariable region) in each isoform which gives us the idea of development of different types of cancers due to mutation of particular RAS isoform. According to COSMIC(Catalogue of somatic mutations in cancer) database 90% of tumours are due to KRAS mutations where 80% of mutations occur at codon 12 and mostly responsible for pancreatic cancer whereas NRAS mutations were strongly linked with hematopoietic tumours where 60% of mutations occur at codon 61 [15]. Among all KRAS oncogene is considered as main target for anti-cancer drug discovery. These mutations are usually identified in myeloid and T-cells and rarely in B-cells. Juvenile myelo monocytic leukaemia (JMML) is most common cancer associated with KRAS based abnormal RAS signalling. There were contrasting studies relating to the myelo proliferation due to KRAS mutations but study conducted by Braun et al and Chan et al used Mx1-cre transgene to induce somatic expression and tumourogenesis was observed in multiple tissues and the mice developed myelo proloferative disease. The anticipated pathogenesis is oncogenic mutations occur in the KRAS gene of haemopietic stem cells their by occur in common lymphoid progenitors (CLP) and common myeloid progenitors (CMP) there by resulting in lymphoma and JMML respectively [16]. PDE6D is essential for movement of normal membrane bound KRAS into the cells. PDE6D encodes the delta subunit of rod specific photoreceptor phosphodiesterase (PDE) which is a key enzyme in photo transduction cascade. It promotes the release of prenylated target proteins from cellular membranes and mutations in this may

lead to neurological disorders, retinal disorders and blood related disorders. The interaction between KRAS and PDE6D sustains spatial organisation of KRAS by facilitating its diffusion in the cytoplasm their by leading to onco-genesis. So if we block the binding of PDE6D to KRAS then it does not reach its normal membrane destination and remains inactive [17, 18]. Study conducted by Gunther Zimmermann et al, demonstrated that inhibition 0f PDE6D-KRAS interactions by means of small molecules provides novel approach to mitigate RA abnormal signalling, their team employed the use of deltarasin and injected the compound into mice and tumour growth was completely stopped in mice that were injected with 10mg/kg of deltarasin intra peritoneal route of administration [19]. Sara Canovas Nunes et al, from their study provided evidence relating to PDE6D dependent RAS signalling with downstream activation of P13K/AKT, they concluded that inhibition of PDE6D effectively triggered anti-leukemic response in AKT dependent cells [20]. This study focuses on integration of next generation sequencing along with BioPython in analysis of leukaemia and related mutations along with identification of potential drug targets that will help in tailoring treatment plan according to the patients.

Materials and Methods:

Several NGS tools have been employed on the sample as the main part of research including Next Generation Sequencing along with BioPython application in analysis of our sample. The sample is retrieved from Protein Data Bank (PDB) and the sample we used is coded as 7PAC. PDB is a global repository that stores three dimensional structural data of biological molecules which helps researchers to access and analyse protein function and also helps us in drug discovery. It is most widely used database as it has relevant visualisations tools along with easy data retrieval [21]. The 7PAC represents the crystal structure of PDE6D in apostate that is a protein expressed in Homo *sapiens* with E.Coli-BL21 (DE3) expression system. After the Sample retrieval various bioinformatics tools were employed. For finding homologs, BLAST-P has been employed. Basic Local Alignment Search Tool is a program that compares sequences of DNA and RNA or proteins to find similar matches that helps in identification of homologs in other species and helps in annotating and classifying organisms based on matching score. It is maintained by NCBI (National Centre for Biotechnology Information that requires command line applications to perform search locally [22,29,30,31,32]. For finding protein-protein interaction, STRING has been employed. Network enrichment analysis refers to Search Tool for the Retrieval of interacting genes/proteins that is used to identify protein-protein interaction identifying proteins that are similar in work function that helps to predict functional associations between proteins that helps in pathway analysis and disease research [23]. To find the gene pathway of KEGG database has been utilized. Kyoto Encyclopedia of Genes and Genomes (KEGG) is a collection of databases that deals with genomes, biological pathways, diseases, drugs and other chemical substances. It has a collection of manually drawn pathways representing interaction networks pertaining to metabolism, genetic information processing, cellular processes, environmental information processing and so on [24]. B-factor analysis has been performed through PyMol, a visualization tool. PyMol software has been written in python language. Pymol helps in visualisation of proteins where we can study structure of protein molecules that are usually helpful in drug design and active site identification [25,26,27].

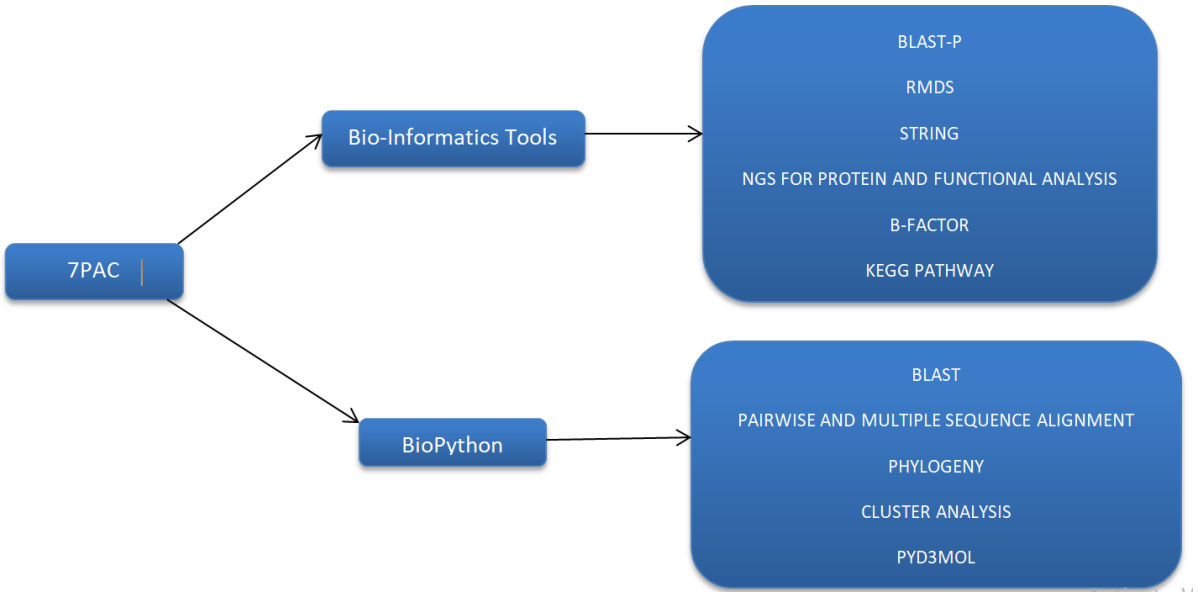


Figure 1: Flow chart representing the methodology employed in assessing the role of PDE6D in blood cancer

Results And Discussion:

1. Sequence Retrieval and Conversion

The 7PAC molecule, retrieved from PDB, is converted into FASTA format and is further used for analysis. Fast-ALL also known as FASTA is a sequence alignment tool that compares biological sequences to find similarities in DNA.

FASTA sequence of 7PAC has been presented below:

>pdb|7PAC|B Chain B, Retinal rod rhodopsin-sensitive cGMP 3',5'-cyclic phosphodiesterase subunit delta

MGSHHHHHHGSENLYFQSAKDERAREILRGFKLNWMNLRDAETGKILWQGTEDLSVPGVEHE
ARVPKKIL
KCKAVSRELNFSSTEQMEKFRLEQKVYFKGQCLEEWFFEFGFVIPNSTNTWQSLIEAAPESQMM
PASVLT
GNVIIETKFFDDDLLVSTSRVRLFYV

2. Homology Search using BLAST-P

After converting into FASTA, to find the homologs of the protein, similarity search has been performed using BLAST-P. BLAST-P helps in comparison of protein query to protein database.

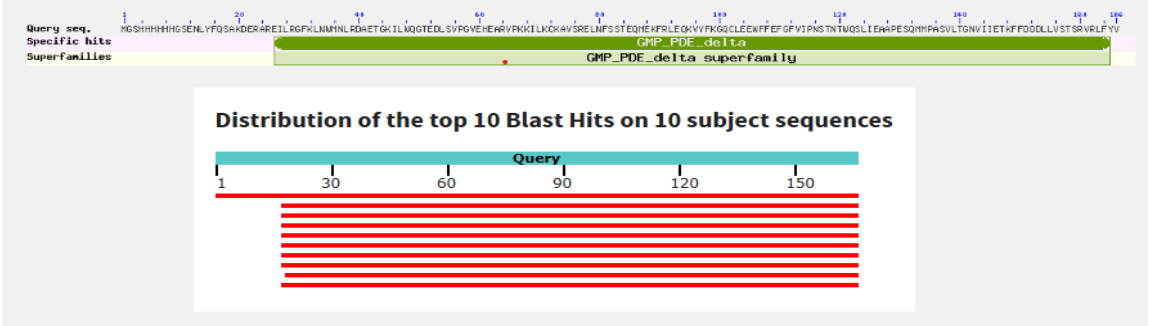


Figure 2: Graphical representation of Blast hits for top 10 sequences

By running BLAST-P the length of sequence was found to be 166 and here blast generated multiple sequence alignment of sample along with their graphical presentation. The BLAST analysis has detected putative conserved domains in the query sequence, specifically identifying the GMP_PDE_delta domain, which belongs to the GMP_PDE_delta superfamily. This suggests functional and structural relevance associated with this domain. The alignment scores are categorized into different ranges, with significant hits falling into the higher similarity regions. The distribution of the top 10 BLAST hits across 10 subject sequences shows strong alignment, as indicated by the continuous red bars, implying high conservation. The presence of a conserved domain further supports the potential biological role of this sequence in GMP_PDE_delta-related pathways.

The alignment scores are categorized as given. **< 40** (not significant), **40 - 50** (low similarity), **50 - 80** (moderate similarity), **80 - 200** (high similarity) , **>= 200** (very high similarity, marked in red)

From the above sequences we have selected three samples that are close to 7PAC and they are 7QJK_AAA, 5TAR_B, 5X74_A , highest alignment score was measured with respect to 7PAC. The e-values obtained are 0.4 0.6 and 0.4 respectively. Out of these three samples we have selected 7QJK_AAA for RMSD calculation.

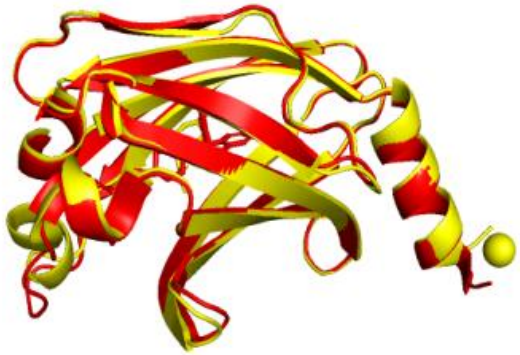
3. Root Mean Square Deviation Analysis

Root Mean Square Deviation is a mathematical measurement used to evaluate the average atomic displacement between two superimposed molecular structures. A lower RMSD value indicates higher structural similarity, while a higher RMSD value suggests greater deviation.

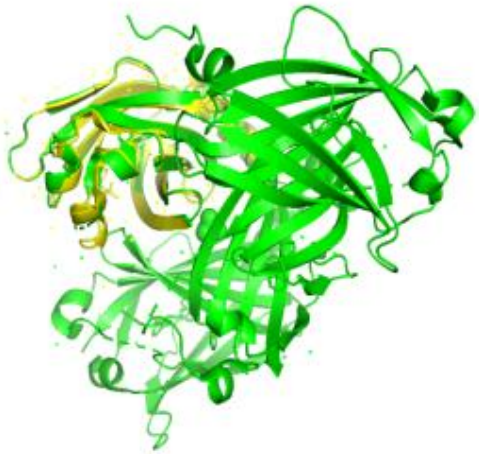
Table 1: RMSD Results

Protein Pair	Atoms Aligned	Cycle 1 RMSD	Cycle 2 RMSD	Cycle 3 RMSD	Cycle 4 RMSD	Cycle 5 RMSD	Final RMSD
5TAR_B **vs Reference**	1202	1.84 (60 rejected)	1.09 (86 rejected)	0.79 (52 rejected)	0.68 (39 rejected)	0.62 (17 rejected)	**0.603 (948 atoms)**
7QJKAAA **vs Reference**	1175	1.46 (70 rejected)	0.87 (86 rejected)	0.59 (44 rejected)	0.52 (31 rejected)	0.49 (8 rejected)	**0.479 (929 atoms)**
5X74A **vs 7PAC**	1214	1.60 (62 rejected)	0.90 (79 rejected)	0.65 (57 rejected)	0.56 (35 rejected)	0.52 (24 rejected)	**0.495 (956 atoms)**

The RMSD values obtained (0.4 - 0.6 Å) indicate a high degree of structural similarity between the analyzed proteins, suggesting that their atomic positions are well-aligned.



Figures 3: representing aligned structures of 5TAR-B



Figures 4: representing aligned structures of 7QJKA

Figures 5: representing aligned structures of 5X74A

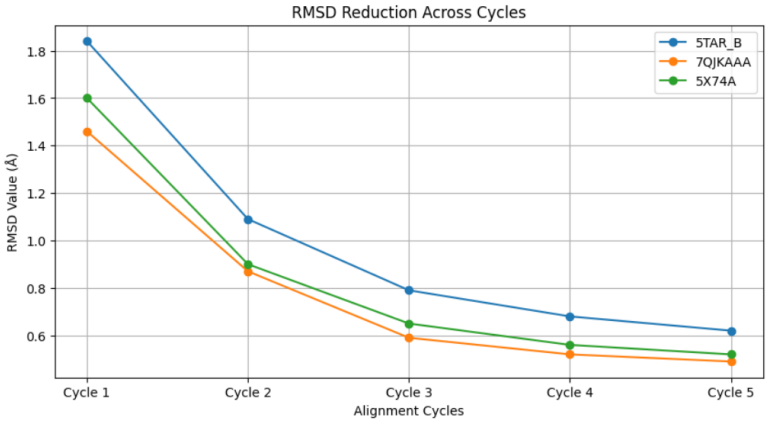


Figure 6: Line plot showing the RMSD values across alignment cycles

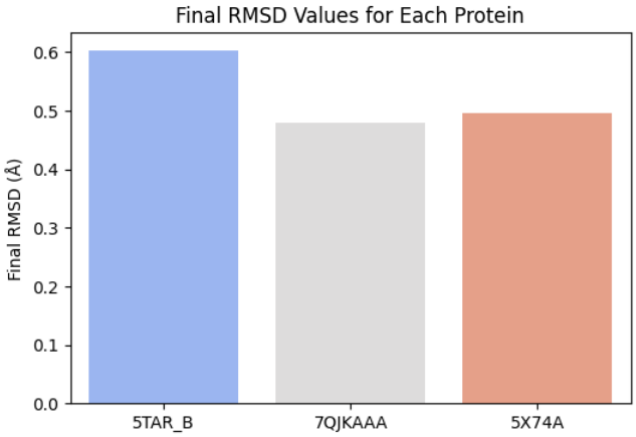


Figure 7: Bar plot comparing the final RMSD values of different protein alignments

line plot is showing the RMSD values across alignment cycles and **bar plot** is comparing the final RMSD values of different protein alignments. An RMSD value of 0.4-0.6 A^0 indicates a high degree of structural similarity implying that the two compared structures are almost identical in terms of their atomic positions. This level of similarity is often desirable in structural biology and computational modelling as it suggests accuracy and reliability in the structural comparison or alignment.

4. Network and Pathway Analysis of PDE6D

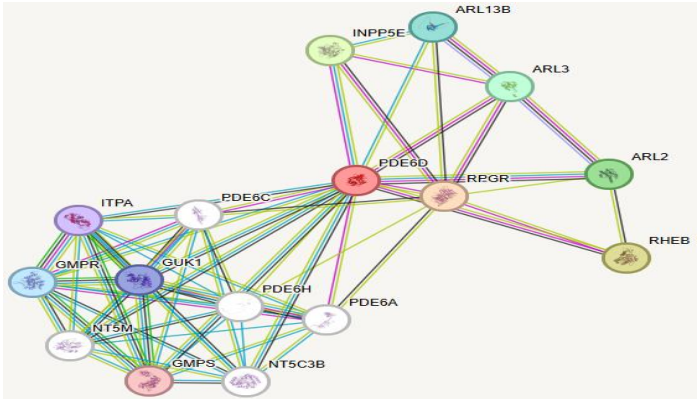


Figure 8: Network analysis of PDE6D with other proteins

This network summarises the predicted associations of the protein PDE6D and the nodes are proteins and the edges represent the predicted functional associations. Redline indicates presence of fusion

evidence, green line indicates neighbourhood, blue line indicate co-occurrence, purple line indicates experimental, yellow line indicates text-mining, light blue indicates database evidence. The presence of more number of lines between two proteins indicates the degree of confidence prediction of the interaction. When string analysis was performed for our sample we found that the protein RPGR has more number of interactions with a combined score of 0.994 and this protein is responsible for ciliogenesis and plays an important role in photoreceptor integrity. Overall average node degree was 7.88 and local clustering coefficient was found to be 0.866 with a total of 16 nodes and 63 edges.

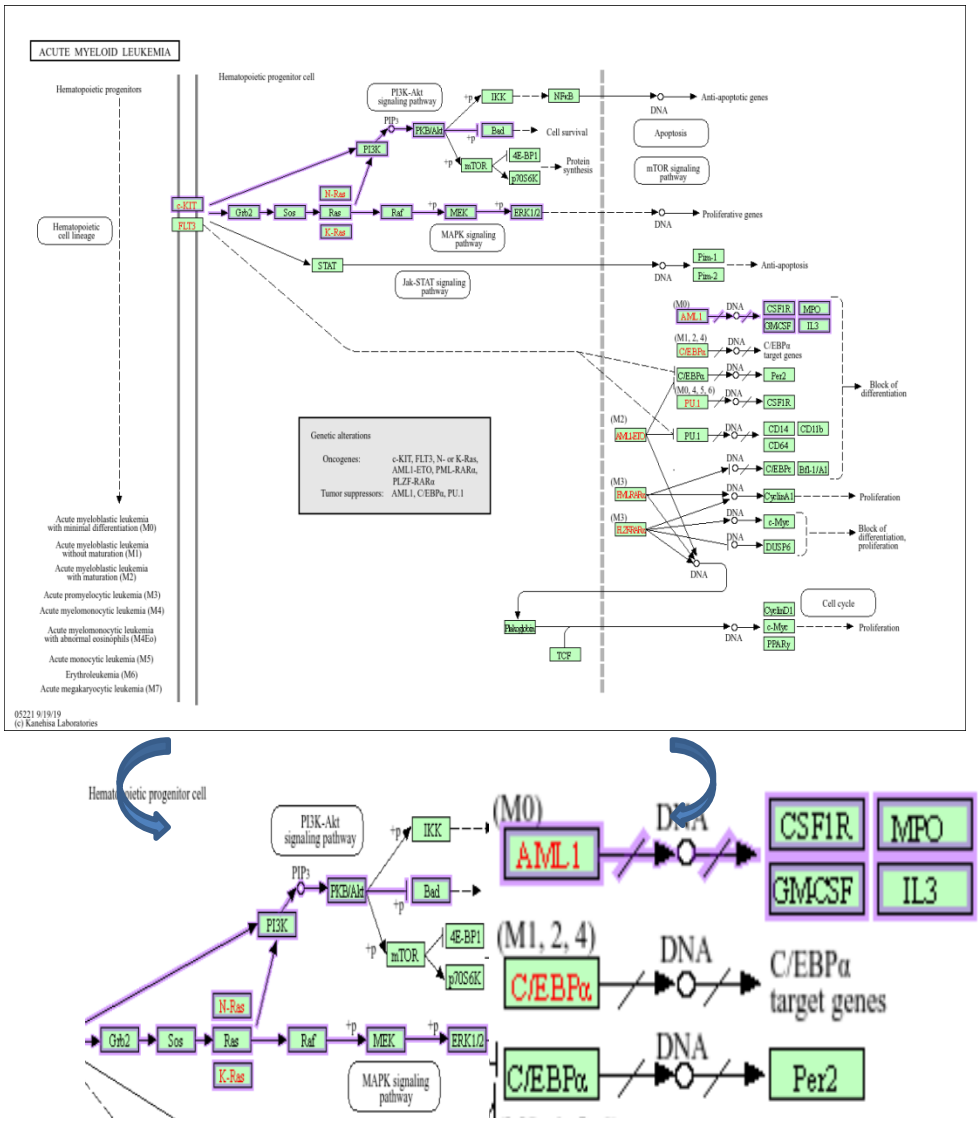


Figure 9: KEGG Pathway representing role of RAS mutation in development of acute myeloid leukaemia

Research shows that PDE6D is primarily involved in RAS Signalling pathway and is investigated as potential therapeutic target, this statement can be elucidated with the help of KEGG pathway analysis, when we searched for Acute myeloid leukaemia(05221 N) related pathway we obtained KRAS and NRAS(hsa04014) pathway where it usually is responsible for direct or indirect cell survival, cell growth, cell migration and gene expression and hence any imbalance between PDE6D and RAS signalling may manifest abnormal changes in above mentioned cell mediated activities that might have a chance of

development of myeloid cancer to an individual. Thus KEGG pathway indirectly gives us strategies that helps us to target drugs that can actually inhibit RAS signalling or PDE6D protein their by attaining anti-leukemic effect.

B-factor is also known as temperature factor or atomic displacement factor and usually represents the degree of atomic displacement around their mean position in a crystal structure and higher B-factor generally represents greater flexibility. It usually finds protein structure regions with high mobility or high activity that helps in comparing active sites for ligand related conformational analysis. This usually helps in drug designing by targeting flexible regions that usually creates drug optimisation effect. We performed B-Factor analysis using PyMol and the results are represented in the below diagram.

5. B-Factor Analysis

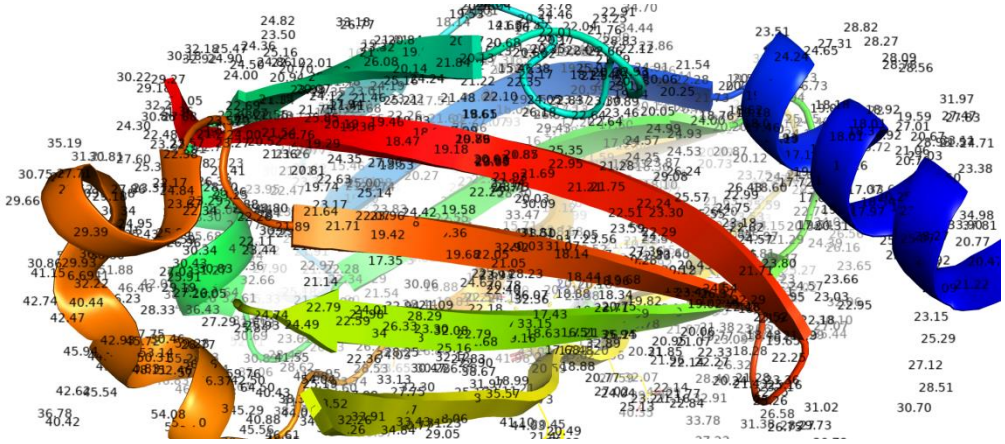


Figure 10: B-Factor Analysis of PDE6D molecule showing varied regions of molecule stability

sample has been represented by different gradient of colours usually signifying a range of B-Factor values. Here some protein sheets are represented by colour red indicating regions with high B-factor value and blue and its related shades usually represent lower B-Factor values and shades of yellow represent median range of B-Factor values. From the above image we can infer that blue labelled regions fall under more stable region with a value between 0-30A^0 and usually we found ligand binding site at this region. Yellow shaded region is usually moderately flexible with average B-values of 30 to 50A^0 and we usually see alpha and beta helical sheets in this area. The most important area essential for our protein analysis is the red labelled region that usually holds any greater than 50A^0 and is more dynamic, mobilised and disordered. This highly unstable region usually shows high rate of mutations and our sample has shown red region areas with B-Factor values greater than 50A^0 hence B-Factor analysis helps us in mutation analysis which will further assess us in development of drug molecules that will target the mutated regions and this can further be applied in personalised vaccine development in terms of cancer.

6. Proteomic Analysis using BioPython

Second part of our research employs use of BioPython for proteomic analysis and we tried to cover all the components that are essential for doing a proteomic analysis.

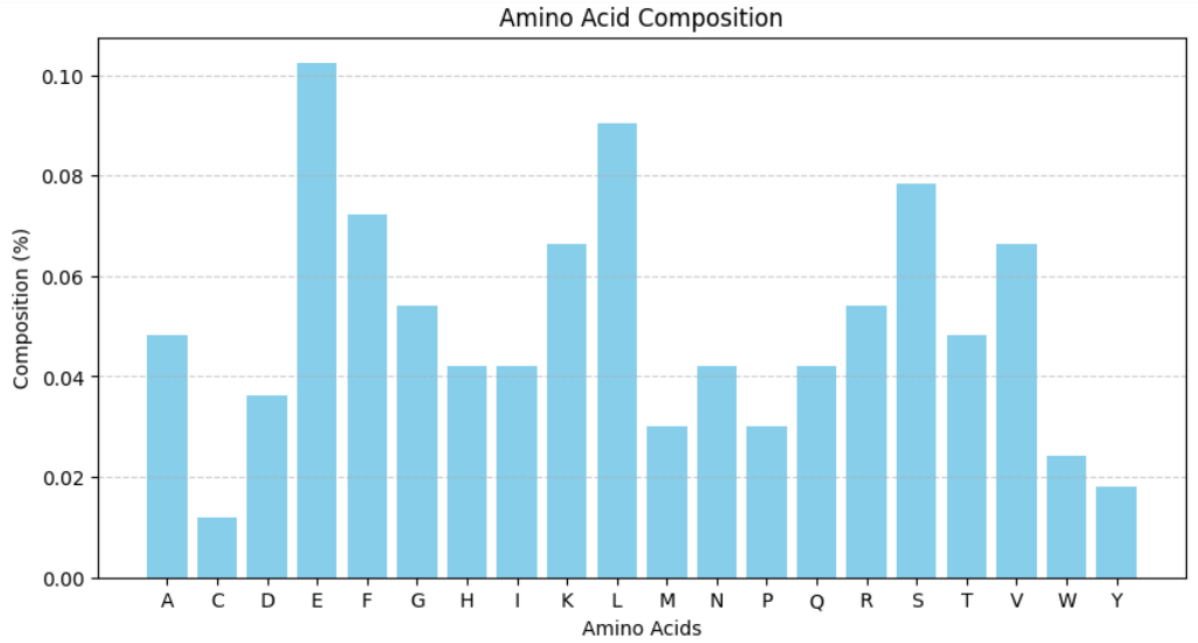


Figure 11: Amino Acid Composition of Protein

The bar chart represents the amino acid composition of a protein sequence. The x-axis lists the amino acid residues, while the y-axis indicates their relative composition in percentage. The highest composition is observed for glutamic acid (E) at approximately 10.24%, followed by leucine (L) and serine (S). Cysteine (C) has the lowest composition among the residues. The distribution highlights variations in amino acid frequency, which may influence the structural and functional properties of the protein.

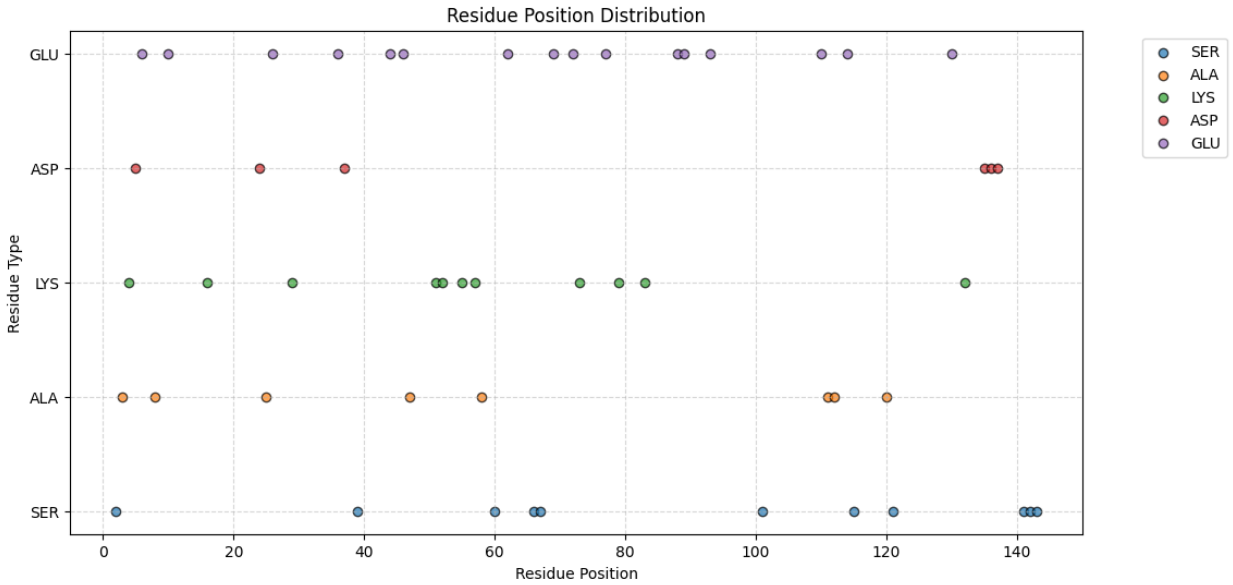


Figure 12: Distribution of specific amino acid residues across different positions

The scatter plot illustrates the distribution of specific amino acid residues across different positions in a protein sequence. The x-axis represents residue positions, while the y-axis denotes different residue types. Five residues are highlighted: serine (SER), alanine (ALA), lysine (LYS), aspartic acid (ASP),

and glutamic acid (GLU). Each residue is color-coded, and its position in the sequence is marked by a dot. The distribution pattern suggests that certain residues are more concentrated in specific regions, potentially influencing the protein's structural and functional properties.

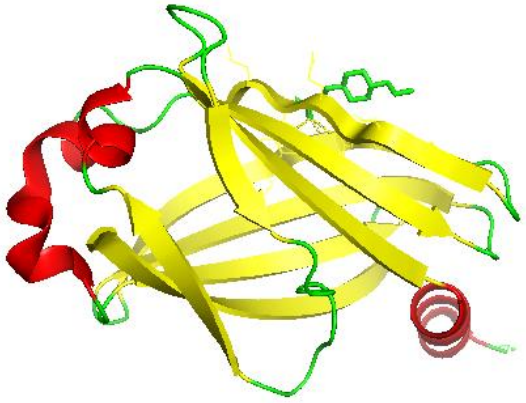


Figure 13: Computational Analysis of Proteomics sample representing Helix (red) Sheet (Yellow) loop (green)

The analysis reveals a protein structure characterized by a significant presence of beta-sheets, forming a prominent core of the molecule. Several alpha-helices are also observed, contributing to the overall architecture. The loops, represented in green, connect these secondary structural elements, providing flexibility and defining the surface topology of the protein. This visual representation aids in understanding the protein's folding pattern and its potential functional domains. The predominance of beta-sheets suggests a stable, possibly globular structure, while the arrangement of helices and loops likely plays a role in interactions with other molecules

Conclusion

This study integrated NGS and bioinformatics to investigate PDE6D's role in blood cancer. Our analysis revealed significant PDE6D interactions, particularly with RPGR, and highlighted its involvement in RAS-driven leukaemia. RMSD analysis indicated high structural conservation, while B-factor analysis identified mutation-prone regions. These findings suggest that PDE6D mutations may contribute to dysregulated RAS signalling in blood cancers, underscoring PDE6D as a potential biomarker and therapeutic target. Further research, including in vitro and in vivo studies, is crucial to validate PDE6D-targeted therapies and advance precision medicine for blood cancer treatment

References

1. Understanding Cancer [Internet]. Available from: https://www.cancer.gov/about-cancer/understanding/what-is-cancer.
2. E NK, Koutras A, Syllaios A, Schizas D, Kalagasidou S, Pagkalos A, et al. Basic principles of molecular biology of cancer cell-Molecular cancer indicators. J buon. 2021;26(5):1723-34.
3. Sathishkumar K, Chaturvedi M, Das P, Stephen S, Mathur P. Cancer incidence estimates for 2022 & projection for 2025: Result from National Cancer Registry Programme, India. Indian J Med Res. 2022;156(4&5):598-607.
4. Zhang N, Wu J, Wang Q, Liang Y, Li X, Chen G, et al. Global burden of hematologic malignancies

and evolution patterns over the past 30 years. Blood Cancer Journal. 2023;13(1):82.
5. Mathers CD, Boschi-Pinto C, Lopez AD, Murray CJL, World Health Organization. Global Programme on Evidence for Health P. Cancer incidence, mortality and survival by site for 14 regions of the world / Colin D. Mathers ... [et al.]. Geneva: World Health Organization; 2001.
6. Black GB, Boswell L, Harris J, Whitaker KL. What causes delays in diagnosing blood cancers? A rapid review of the evidence. Primary Health Care Research & Development. 2023;24:e26.
7. Metzeler KH, Herold T, Rothenberg-Thurley M, Amler S, Sauerland MC, Görlich D, et al. Spectrum and prognostic relevance of driver gene mutations in acute myeloid leukemia. Blood. 2016;128(5):686-98.
8. Ramkissoon LA, Montgomery ND. Applications of next-generation sequencing in hematologic malignancies. Human Immunology. 2021;82(11):859-70.
9. Qin D. Next-generation sequencing and its clinical application. Cancer Biol Med. 2019;16(1):4-10.
10. Eren K, Taktakoğlu N, Pirim I. DNA Sequencing Methods: From Past to Present. Eurasian J Med. 2022;54(Suppl1):47-56.
11. Guan YF, Li GR, Wang RJ, Yi YT, Yang L, Jiang D, et al. Application of next-generation sequencing in clinical oncology to advance personalized treatment of cancer. Chin J Cancer. 2012;31(10):463-70.
12. S. KUaG. Ngs and sequence analysis with biopython for prospective brain cancer therapeutic studies. . International Journal for Research in Applied Science and Engineering Technology 2023;;11(4):3318-3329.
13. Tyagi E, Kumari U. Next Generation Sequencing Analysis of NSP5 NSP6 and NSP7 SARS CoV-2 Nonstructural Protein with Bio-Python2024.
14. Keeton AB, Salter EA, Piazza GA. The RAS–Effector Interaction as a Drug Target. Cancer Research. 2017;77(2):221-6.
15. Prior IA, Lewis PD, Mattos C. A comprehensive survey of Ras mutations in cancer. Cancer Res. 2012;72(10):2457-67.
16. Zhang J, Wang J, Liu Y, Sidik H, Young KH, Lodish HF, et al. Oncogenic Kras-induced leukemogeneis: hematopoietic stem cells as the initial target and lineage-specific progenitors as the potential targets for final leukemic transformation. Blood. 2009;113(6):1304-14.
17. PDE6D phosphodiesterase 6D [Homo sapiens (human)] [Internet]. Available from: https://www.ncbi.nlm.nih.gov/gene/5147.
18. Abankwa D. PDE6D/KRAS inhibitor development for cancer treatment (inhibitPDE-RASv2) 2021 [Available from: https://www.uni.lu/fstm-en/research-projects/inhibitpde-rasv2/.
19. Zimmermann G, Papke B, Ismail S, Vartak N, Chandra A, Hoffmann M, et al. Small molecule inhibition of the KRAS-PDEδ interaction impairs oncogenic KRAS signalling. Nature. 2013;497(7451):638-42.
20. Canovas Nunes S, De Vita S, Anighoro A, Autelitano F, Beaumont E, Klingbeil P, et al. Validation of a small molecule inhibitor of PDE6D-RAS interaction with favorable anti-leukemic effects. Blood Cancer J. 2022;12(4):64.
21. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 2000;28(1):235-42.
22. Krauthammer M, Rzhetsky A, Morozov P, Friedman C. Using BLAST for identifying gene and protein names in journal articles. Gene. 2000;259(1):245-52.

23. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47(D1):D607-d13.
24. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27-30.
25. Barthels F, Schirmeister T, Kersten C. BANΔIT: B'-Factor Analysis for Drug Design and Structural Biology. Mol Inform. 2021;40(1):e2000144.
26. Uma Kumari, Shruti Gupta, NGS and Sequence Analysis with Biopython for Prospective Brain Cancer Therapeutic Studies, International Journal for Research in Applied Science & Engineering Technology,Volume 11, Issue IV ,2023
27. Kukreja, Vinita &Kumari, Uma. (2023). Data Analysis of Brain Cancer with Biopython. 8. 2146-2154.
28. Uma kumari, Devanshi Gupta, Computational analysis of Glioma Transcriptome and proteome with Bio python ;2023Volume-13, Issue-1, Pp 1-14,IJBTR
29. Kumari Uma & Tripathi, Kartik. (2023). Computational Analysis and Molecular Docking Approach for Liver Cancer. International Journal for Research in Applied Science and Engineering Technology. 11. 1828 to 1633. 10.22214/ijraset.2023.54927.
30. Uma kumari,Meenakshi Pradhan,Saptarshi Mukherjee,Sreyashi Chakrabarti;,Ngs Analysis approach for neurodegenerative disease with Biopython",2023 ,volume 10,issue 9 , http://doi.one/10.1729/Journal.36043
31. Vineeta Johri , Ekta Tyagi , Uma Kumari. (2024). Next Generation Sequencing Analysis of NSP5 NSP6 and NSP7 SARS CoV-2 Nonstructural Protein with Bio-Python. *South Eastern European Journal of Public Health*, 1653–1666. https://doi.org/10.70135/seejph.vi.2936 Dec 2024
32. Uma Kumari,Gurpreet Kaur *et al*,"Biopython/Network Of Protein Identification And NGS Analysis Of Glioma Cancer ATP Competitive Type III C-MET Inhibitor : Volume 11, Issue 2 : 27-Jun-2024 :pp 41-51 : http://doi.one/10.1729/Journal.40229