MOLECULAR INSIGHTS INTO KRAS ONCOGENIC PROTEIN VARIANTS AND IDENTIFICATION OF KRAS G12D NATURAL INHIBITORS: A MOLECULAR DOCKING AND SIMULATION APPROACH

BIKASH CHANDRA SAHU

Admission No. - 212121320

DEPARTMENT OF BIOINFORMATICS

CENTRE FOR POST GRADUATE STUDIES

ODISHA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY

BHUBANESWAR, 751003, ODISHA

2023

MOLECULAR INSIGHTS INTO KRAS ONCOGENIC PROTEIN VARIANTS AND

IDENTIFICATION OF KRAS G12D NATURAL INHIBITORS: A MOLECULAR DOCKING AND SIMULATION APPROACH

A Thesis Submitted to

Odisha University of Agriculture and Technology, Bhubaneswar for partial fulfilment of the requirement for the award of degree of

MASTER OF SCIENCE IN BIOINFORMATICS

BY

BIKASH CHANDRA SAHU

Admission No. – 212121320

DEPARTMENT OF BIOINFORMATICS
CENTRE FOR POST GRADUATE STUDIES
ODISHA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY BHUBANESWAR, 751003, ODISHA

2023

Advisor Dr. Surya Narayan Rath

Pondicherry University
A Central University under Ministry of Education, Govt. of India
Established by an act of Parliament No.53 of 1985
Kalapet, Pudhucherry-605014,
India.

Dr. K. M. Kumar
Associate Professor
Department of Bioinformatics
School of life Sciences
Email: kumarkm@bicpu.edu.in,
mobile: 9900765490

CERTIFICATE-I

This is to certify that **Mr. Bikash Chandra Sahu (Adm. No. 212121320)**, M.Sc. (Bioinformatics), **Department of Bioinformatics, OUAT, Bhubaneswar, Odisha** has carried out his project work entitled **"Molecular insights into KRAS protein variants and identification of KRAS G12D natural inhibitor- A molecular docking and simulation approach"** under my supervision and guidance for the period of June October, 2023. The thesis is required to be submitted in partial fulfilment for the award of a degree of Master of Science in Bioinformatics under **Odisha University of Agriculture & Technology, Bhubaneswar, Odisha.** The thesis embodies the bonafide and original work done by her and no part of the dissertation has been submitted anywhere for any degree, or diploma to any other University or Institution.

I wish him all success in his future endeavors.

Place: Pondicherry
Date: 15/11/2023

Dr. K. M. KUMAR
Associate Professor
Department of Bioinformatics
School of Life Sciences
Pondicherry University
Puducherry - 605014

ODISHA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY
DEPARTMENT OF BIOINFORMATICS
CENTRE FOR POST GRADUATE STUDIES

CERTIFICATE – II

This is to certify that the dissertation entitled "**Molecular insights into KRAS oncogenic protein variants and identification of KRAS G12D natural inhibitors: A molecular docking and simulation approach** " submitted by **Bikash chandra Sahu (Adm. No. 212121320)**, to the Odisha University of Agriculture & Technology, Bhubaneswar in the partial fulfillment of the requirements for the award of the degree of **Master of Science in Bioinformatics** has been approved by the students advisory committee after an oral examination of the same in collaboration with external examiner.

ADVISORY COMMITTEE

1. Dr. Surya Narayan Rath
   Assistant Professor
   Department of Bioinformatics
   CPGS, OUAT, Bhubaneswar
   Chairperson .....

2. Dr. K.M Kumar
   Associate Professor
   Department of Bioinformatics
   Pondicherry University, Puducherry
   Member
   Dr. K. M. KUMAR
   Associate Professor
   Department of Bioinformatics
   School of Life Sciences
   Pondicherry University
   Puducherry - 605014

3. Dr. Sukanta Kumar Pradhan
   Head of the Department
   Department of Bioinformatics
   CPGS, OUAT, Bhubaneswar
   Member .....

External Examiner

.....20/11/23...

ODISHA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY

Plagiarism Certificate
(For PG/PhD thesis)

This is to certify that the thesis entitled **"Molecular insights into KRAS oncogenic protein variants and identification of KRAS G12D natural inhibitors: A molecular docking and simulation approach"** Submitted by **Bikash chandra Sahu** bearing Admission No.**212121320** is plagiarism checked and has not crossed the limit as per the Anti-Plagiarism policy of OUAT.

Bikash chandra Sahu

Signature of the student

Dr. K. M. KUMAR
Associate Professor
Department of Bioinformatics
School of Life Sciences
Pondicherry University
Puducherry - 605014

Signature of the Major Advisor /Chairman

ACKNOWLEDGEMENT

I would like to acknowledge and express my gratitude to the following people for their magnificent support and contributions to my journey and to the creation of this work.

I take this great opportunity to express my gratitude and indebtedness towards my esteemed Advisor **Dr. K.M Kumar**, Associate Professor, Department of Bioinformatics, Pondicherry University for her invaluable guidance, support, encouragement, constant supervision, whole-hearted cooperation, and, timely advice at each difficult step of my project work.

With a feeling of profound gratitude, I owe a special word of thanks to my Major Guide **Dr. Suryanarayan Rath**, Assistant Professor, Dept. of Bioinformatics, CPGS, OUAT Puducherry for his vital encouragement, immense support, and timely advice during the entire period of my dissertation work.

On the completion of my project, I would like to extend my heartfelt thanks to **Dr. P.T.V. Lakshmi**, Head of The Department, Pondicherry University, Puducherry for granting me permission to join the prestigious University and carrying out the project work.

I am cordially thankful to my teachers **Dr. Sukanta Kumar Pradhan**, **Mrs. Sucharita Balabantray**, and **Dr. Sushma Rani Martha**, **Mr. Sujit Kumar Das**, **Mrs. Sudipta Mohanty** for their precious gift of knowledge. Their broad knowledge and logical way of thinking have been of great value to me.

I avail this opportunity to express my heartfelt thanks to my friends Shibangi Das, Kshirabdhi Dhal, for their cooperation and help, for which I could be able to complete my coursework smoothly.

I would like to place on record my sincere thanks to respected seniors for all their help and support throughout my coursework and research.

No appropriate words to express my full respect and indebtedness to my mother for her unlimited love, continuous support, affection, sacrifice, and moral support. I am grateful to have a best friend as my mother who guides me to choose the best path for me and has faith in me which always encourages me in life.

Above all, I thank my Almighty **Lord Jagannath** for giving me the opportunity and guidance to achieve my goal and to be successful in this part.

Finally, I would like to thank all those involved directly or indirectly with my work.

Bikash chandra Sahu

Place: Bhubaneswar
Date: 19/11/2023

Bikash chandra sahu
Adm No. 212121320

Name of the Student	: Bikash chandra Sahu
Admission no	: 212121320
Title of the Thesis	: Molecular insights into KRAS oncogenic protein variants and identification of KRAS G12D natural inhibitors: A molecular docking and simulation approach
Thesis submitted for	: Master of Science in Bioinformatics
Name of the Department	: Department of Bioinformatics
College and University	: Centre for Post Graduate Studies Odisha University of Agriculture and Technology, Bhubaneswar- 751003
Year of submission	: 2023
Name of the advisor	: Dr. Surya Narayan Rath Assistant professor Department of Bioinformatics CPGS, OUAT, Bhubaneswar

ABSTRACT

The KRAS protein belongs to the RAS GTPase protein family which are important in various cell signalling pathways and cell division. To find the mutational stability of KRAS protein molecular dynamics simulation is an efficient approach, there are no approved treatments for the KRAS G12D mutation, which is most prevalent in lung, colon, and pancreatic cancers. Due to their lack of binding pockets and smooth surface, switch II of the KRAS G12D mutation may be an effective target for treating a variety of malignancies, according to recent research. Therefore, in the present study, we targeted the switch II pocket (residues 57–76) of KRAS G12D with natural compounds in comparison with the reference KRAS Switch II inhibitor MRTX1133, which reported to be showing less GI tract absorption in animal models. Initially, we screened 17,752 natural compounds using XP docking in maestro and followed by ADMET analysis of the molecules based on drug-likeness properties, and ADME profiles best 5 molecules were chosen as leads with docking scores more than -8.5 kcal/mol for further studies. The molecular dynamics simulation revealed that the compound with PubChem ID 10207 named as Aloe emodin is more stable than the reference molecule. Further modification and structural optimization can lead to better binding affinity towards the KRAS G12D and can be considered as potential lead molecules for further studies.

Keywords: KRAS, Docking, Md simulation, Natural Compounds, Switch II Inhibitors

Advisor

Bikash chandra Sahu

Author

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATION

KRAS	**:**	Kirsten rat sarcoma
PDAC	**:**	Pancreatic ductal adenocarcinoma
GDP	**:**	Guanosine di-phosphate
kDa	**:**	Kilo Dalton
UTR	**:**	Untranslated region
GEFs	**:**	Guanosine Exchange Factors
BLAST	**:**	Basic Local Alignment Search Tool
GAPs	**:**	GTP activating proteins
IPTG	**:**	Isopropyl β- d-1-thiogalactopyranoside
SDS	**:**	Sodium Dodecyl Sulphate
NMR	**:**	Nuclear magnetic resonance
MD	**:**	Molecular dynamics
I-TASSER	**:**	Iterative Threading ASSEmbly Refinement
GROMACS	**:**	GROningen MAchine For Chemical Simulation
RMSD	**:**	Root Mean Square Deviation

CHAPTER I

1. INTRODUCTION

The human body is an incredibly complex biological system, with countless biochemical, biophysical, and physicochemical changes occurring simultaneously at any given moment. Enzymes play a pivotal role in catalyzing or being associated with all of this biotransformation that take place within the body. Several studies have found that the RAS genes, which are in charge of the pathophysiology of cancer, were first discovered as viral genes in the early 1980s when Harvey and Kirsten's murine sarcoma retroviruses were studied. Following the identification of the first human RAS genes and proteins in 1982, scientists saw that human cancer cell lines mutated to activate these genes.

The RAS protein in humans is a GTPase protein. Currently in humans the RAS protein family consists of more than 150 proteins Three genes, known as N RAS (Neuroblastoma RAS viral oncogene homolog), HRAS (Harvey rat sarcoma viral oncogene homolog), and KRAS (Kirsten Rat sarcoma viral oncogene homolog), encode all the proteins of the RAS family. As the RAS protein is a GTPase protein it comprises of small guanosine triphosphate chemicals which acts as a molecular switch for various cell s pathways. Thirty percent of human malignancies are caused by mutations in the RAS gene; of these, eighty-six percent are caused by the KRAS oncogene, which also causes ninety percent of pancreatic cancer cases. With a dismal prognosis and resistance to chemotherapy provided by normal care, new treatments targeting KRAS are required.

KRAS GENE

KRAS has been recognized as KRAS-1 and KRAS-2 . KRAS-1 is a pseudogene, while KRAS2 is located in chromosome 12. KRAS-2's coding region is made up of six exons that are more than 45 kB long. It comprises the invariant coding regions 2, 3, and 4. Two mRNA variants, 4A and 4B, are produced by alternative splicing of exon 4. Additionally, KRAS-2 protein isoforms KRAS4A and KRAS4B are generated. During alternative splicing, exon 5 can be omitted. The C-terminal portion of KRAS4B is encoded by Exon 6, which also makes up KRAS-1's 3'UTR (3' untranslated region). The predominant transcriptional product in human cells is KRAS isoform 4B, which has 189 amino acids,

while isoform 4A has 188 amino acids. Both isoforms are created through alternative splicing.

KRAS PROTEIN

KRAS belongs to the superfamily of small GTPases and it mainly binds to GTP(Guanosine triphosphate) . The KRAS protein contains two regions which are the Nterminal catalytic domain also known as G domain (guanine binding domain) and the Cterminal hyper variable region known as HVR. The G domain or the catalytic domain is a highly conserved region with showing greater degree of homology. The G domain consists of three regions P-loop, switch I and switch II regions. The three regions are involved in different functions.G domain serves as a GTP-GDP switch and promotes GTP-GDP exchange. While the switch region forms the binding region for effector proteins, in addition to stabilizing nucleotide phosphates, the P-loop acts as the phosphate binding region. The primary factors that regulate the G-domain switch are GTPase activating proteins (GAPs) and guanine exchange factors (GEFs), which facilitate the GDP to GTP switch and activation.

MEMBRANE LOCALIZATION

Since KRAS is found to be linked to the inner membrane of the plasma membrane, the C-domain, which exhibits a high degree of polymorphism, is responsible for anchoring RAS to the inner membrane of the plasma membrane. It contains the CAAX box (cysteine, two aliphatic amino acids, any amino acid) and is responsible for posttranslational modifications such as prenylation. The terminal cysteine of the CAAX box gains a farnesyl group or geranylgeranyl moiety through prenylation, which is followed by cleavage and methylation. For KRAS monomers to be active, they must be localized For KRAS monomers to be active, they must be localized.

ACTIVATION OF KRAS

Activation of receptor tyrosine kinase (RTK), growth factors like fibroblast growth factors (FGF), platelet-derived growth factors (PDGF), insulin-like growth factors (IGF), and epidermal growth factor (EGF), as well as cytokines, can all result in KRAS activation. With the help of RAS GEFs, GDP-bound KRAS is phosphorylated to a GTPbound KRAS active state, which is known as KRAS activation (fig 1). In drosophila, the similar son of sevenless gene was discovered., and CDC25 in S. cerevisiae was the first RAS GEF to be

discovered. Later, it is discovered that RAS guanine-nucleotide-releasingfactors (RASGrfs), son of sevenless proteins, and CDC25 have mammalian counterparts. EGFR is dimerized and phosphorylated by tyrosine kinase activation when EGF binds to its receptor, next the sequence homology 2 (SH2) domain of growth factor receptor bound protein 2 (GRB2) is bound by the phosphorylated EGFR-dimer. SOS is bound to the EGFR-GRB2 complex by its SH3 domain, which causes SOS to localize to the membrane. SOS1 and SOS2 are GEFs that facilitate GTP-RAS interaction and encourage the decoupling of GDP from RAS.A confirmational change in the switch I and switch II regions is brought about by GTP-RAS binding, which also encourages downstream signalling. KRAS is also activated by src homology phosphatase 2 (SHP2), another tyrosine phosphate enzyme, through processes that are still being worked out. Based on available data, SHP2 may have a scaffolding protein role, improving GRB2-SOS1 binding and encouraging KRAS activation. The autosomal dominant RAsopathy known as Noonan Syndrome, which is characterized by cardiac, endocrine, neurodevelopmental, and hematologic disorders, can be caused by mutations in the PTPN11 gene, which codes for SHP2.

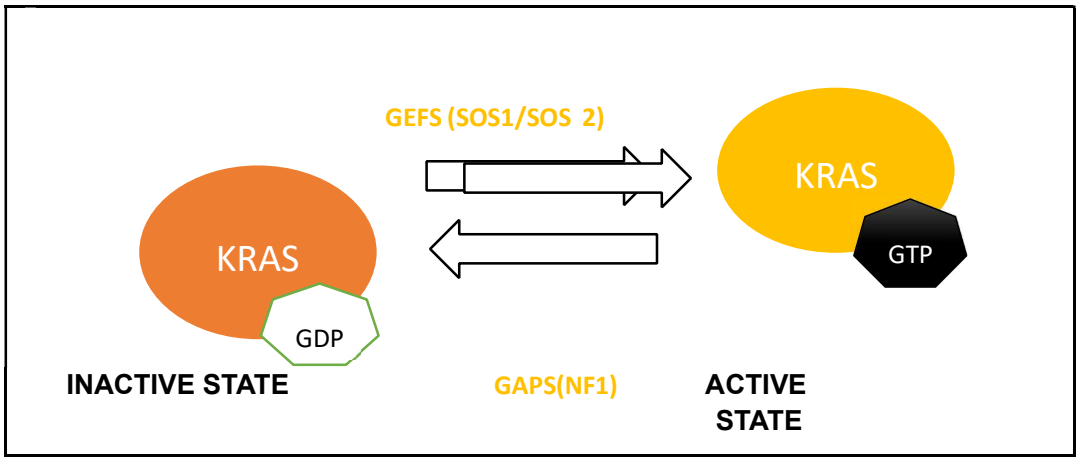


Figure 1: KRAS activation and inactivation

INACTIVATION OF THE PROTEIN

The inactivation of KRAS is a complex process that involves multiple molecules interacting with one another. After being activated, the RAS-GTP complex goes through intrinsic GTP hydrolysis, which leaves it in an inactive state called RAS-GDP. In order to speed up GTP hydrolysis and cause RAS deactivation, GAPs are crucial. The main process of RAS-GTP hydrolysis is GAP-mediated hydrolysis, with RAS GAPs such as neurofibromin 1 and p120GAP playing a significant role. RAS GAPs promote carcinogenesis by preventing GTP hydrolysis, whereas oncogenic KRAS mutations reduce GTPase activity and resistance.

KRAS MEDIATED PATHWAYS

Three main signaling pathways are triggered downstream by KRAS activation: the MAP kinase pathway, the P13K-AKT-mTOR pathway, and the pathways that causes tumor causing metastatic growth and invasion protein 1 (T1AM1-RAC) and RAS-related protein (RAL).

MAPK PATHWAY

Through phosphorylating RAS, RAF, MEK, and ERK, the MAPK pathway controls the cell cycle and proliferation. The dimer of mutant RAS increases the activation and binding of RAF, which initiates a significant downstream signaling pathway. The last enzyme, ERK, promotes cellular differentiation and proliferation by activating transcription factors.

PI3K PATHWAY

The PI3K Pathway is a multifaceted procedure. It starts with the activation of Phosphatidylinositol-4,5-phosphorylation of PIP2 into PIP3 (phosphatidylinositol 3,4,5triphosphate), which phosphorylates AKT and causes mTOR, FOXO, and NF-kB phosphorylation downstream. In the end, this procedure results in cell survival and apoptosis resistance. RAS mutations frequently cause an upregulation of the PI3K pathway. RAS's precise function in PI3K activation in normal cells is still unknown, though.

RAL PATHWAY

A protein called RAL guanine nucleotide stimulator (RALGDS), which is activated by RAS, promotes cell migration and RAS-dependent tumor growth. Additionally, RALGDS activates the JNK pathway, which promotes transcription and the advancement of the cell cycle. The phosphorylation of PAK serine/threonine kinases is caused by GEFs known as TIAM1-RAC. When KRAS is in its normal state, it functions as a crucial connector between a number of cellcycle pathways. By activating multiple downstream activation pathways, a KRAS mutation can cause oncogenesis.

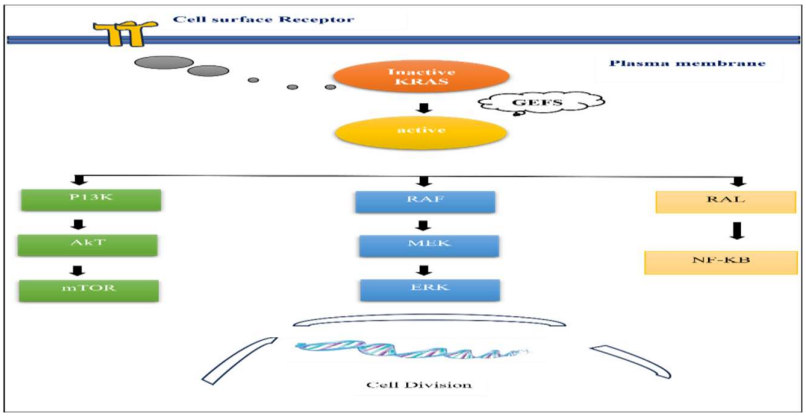


Figure 2: KRAS mediated cell signaling pathways

KRAS MUTATIONS IN HUMAN CANCER

Mutations of RAS genes accounts for about 30% of human cancers and KRAS is found almost in more than 85% of RAS mutated cancers. Over 85% of cases are pancreatic cancers, with colorectal cancer (~40%), non-small cell lung cancer (~30%), and cholangiocarcinoma (~20%) following in order of incidence. In lung and pancreatic cancers, KRAS mutations frequently occur early in the tumorigenesis process. On the other hand, driver mutations in the wnt signaling pathway typically precede KRAS mutations, which are a stage in the development of colon cancer

The majority of KRAS mutations are single-base missense mutations, which are located on exons 2 and 3 at codons 12 (83%), 13 (14%), or 61 (2%). These mutations enhance GTP binding, initiate KRAS, and convey signals via diverse pathways. A mutation in codon 12 inhibits GAP-mediated GTP hydrolysis and validates the GAP binding site, preventing KRAS deactivation. Six types of substitutions are reported for the glycine present at codon 12, which are aspartic acid(G12D), cysteine (G12C), valine(G12V), alanine (G12A), Arginine (G12R), serine (G12S). KRAS G12D is a mutant which occurs in higher numbers than other KRAS forms due to a single nucleotide change from glycine to aspartic acid. It downregulates thymine-DNA glycosylase, which in turn promotes proliferative signalling, invasions, invasions, metastasis, and apoptosis.

The development of pancreatic ductal adenocarcinoma (PDA) is a multistage process in which premalignant lesions accumulate mutations and eventually develop into malignant invasive carcinoma, which is categorized by degree of cellular dysplasia from PanIn 1 to PanIn 3. The recruitment of pro-tumorigenic cells and the control of the tumor microenvironment (TME), which lead to tumor invasion and metastatis, are associated with oncogenic KRAS. Mutant KRAS has recently been demonstrated to play a role in the initiation and maintenance of pancreatic cancer by activating paracrine hedgehog signaling, as well as in fostering reciprocal signaling between pancreatic tumor cells and the stroma.

Using a proteomic and phospho proteomic method, it was discovered that tumor cells with KRAS G12D mutations release sonic hedgehog, which stimulates pancreatic stellate cells, a subset of cancer-associated fibroblasts (CAFs) with the ability to control tumor growth and progression.

OBJECTIVES

I. Inspecting structural and conformational impact of mutations in KRAS oncoprotein.

II. Screening of effective natural compounds against potential KRAS mutant

CHAPTER II

2.REVIEW OF LITERATURE

KRAS, a small GTPase transductor protein, is involved in cell division control and resistance to chemotherapy and biological treatments. Mutations in KRAS make it difficult for the protein to transition between active and inactive states, increasing cell resistance. KRas4B, a member of a group of small GTPases, controls cell proliferation, differentiation, and survival. KRAS mutations are common in cancer and depend on its connection with the plasma membrane. The 188 amino acids that the KRAS gene codes for are members of the are mostly caused by mutations in the KRAS gene. Recently, KRAS G12C mutant allele-specific inhibitors were created that covalently alter Cys12's thiol, trapping KRAS in an inactive GDP bound form. Utilizing real-time NMR to concurrently observe GTP hydrolysis and inhibitor binding in order to investigate the mechanism of action of the covalent inhibitors in both in vitro and intracellular contexts. The rate-limiting step for ARS-853 modification is GTP hydrolysis, as demonstrated by in vitro NMR investigations, which also revealed that the rate constant of ARS-853 modification and GTP hydrolysis are the same. According to in-cell NMR studies, the ARS-853 reaction proceeds much more quickly than it does in vitro, which is indicative of endogenous GTPase proteins accelerating GTP hydrolysis.

There still aren't any approved medications to treat these mutations, even forty years after KRAS was discovered. Because of their smooth surface and lack of RAS family. RAS protein residues, which have a molecular mass of 21.6 kD.

Within the coding area, nonsynonymous single nucleotide polymorphisms (nsSNPs) have been found; some of these are linked to various disorders. But structural modifications aren't yet well defined. Additionally, genetic information was taken from many sources, such as ENSEMBLE, dbSNPs, and the 1000 Genome Project, and evaluated using in silico techniques. Depending on the tools being utilized Researchers have discovered that K-Ras4B can be dispersed in both hard and loosely packed membrane domains, and its membrane binding domain interacts with phospholipids.

Located at the plasma membrane, the small GTPase KRAS connects intracellular signalling networks to external growth factor activation by acting as a molecular switch. During this procedure, effectors like RAF kinase are brought to the plasma membrane by KRAS and

activated through a sequence of intricate molecular processes. Understanding the activation of RAF kinase and assessing new treatment options for the suppression of oncogenic KRAS mediated signalling depend on defining the membrane-bound state of KRAS. Our approach involved combining various biophysical measures with computational techniques to produce a consensus model for membrane-anchored, legitimately processed KRAS. Using a combination of neutron reflectivity, fast photochemical oxidation of proteins (FPOP), and other information, identified a third considerably populated state, in contrast to the two membrane-proximal conformations previously reported. KRAS mutant lung adenocarcinomas (LUAD) target downstream with tolerable toxicity. Recent developments in focusing KRAS mutations in various cancers, such as the KRAS G12C protein, have opened new avenues for effective treatment. KRAS is the most commonly mutated oncogene in human cancers. The main oncogene responsible for cellular survival and proliferation, Kirsten Rat Sarcoma (KRAS) is the most commonly mutated oncogene in all malignancies. The main oncogene responsible for cellular survival and proliferation, Kirsten Rat Sarcoma (KRAS) is the most commonly mutated oncogene in majority of human cancers.

Activating KRAS mutations are linked to the initiation and progression of pancreatic ductal adenocarcinoma (PDAC) and are seen in more than 90% of cases. While KRAS is a crucial oncogene and hence a valuable target for therapy, it has proven extremely difficult to suppress therapeutically, and specific mutant inhibitors of KRAS have just recently been identified. The activation of KRAS signalling and the function of mutant KRAS in the development of PDAC are covered in this review. It has long been believed that KRAS cannot be manipulated, and numerous attempts at drug development centered on indirect targeting have shown negative results

The most commonly mutant oncogene in all cancers is Kirsten Rat Sarcoma (KRAS), the primary oncogene responsible for cellular survival and proliferation. More than 90% of patients had activating KRAS mutations, which are associated with the development and course of pancreatic ductal adenocarcinoma (PDAC). Although KRAS is an important oncogene and hence a useful target for treatment, it has been very challenging to suppress therapeutically, and only lately have particular mutant inhibitors of KRAS been found. This review discusses the activation of KRAS signalling and the role of mutant KRAS in the development of PDAC. It's been well accepted for a long time that KRAS cannot be altered, and multiple attempts to produce drugs with indirect targeting have failed. Belonging to the GTPase Ras family, the human K-Ras protein hydrolyzes GTP to GDP and simultaneously

transitions from its active to its inactive form. It is a crucial oncoprotein because a number of mutations, especially those at residue position 12, are highly frequent in a variety of human malignancies. Thus, the K-Ras protein is a crucial target for the development of medicinal anti-cancer drugs. In the GTP complexed active forms, we describe here the nearly full sequence-specific resonance assignments of the wild-type and the oncogenic G12C and G12D mutants, including the functionally significant Switch I and Switch II regions. A thorough functional dynamics analysis of wild-type K-Ras and its G12 mutant is based on these assignments.

More than 25% of human cancers contain oncogenic mutations involving the KRAS GTPases. The lack of a good small-molecule binding site makes KRAS essentially undruggable. In this instance, we provide the unusual crystal structure of His-tagged KRASG12D, which displays an amazing conformational shift. In an adjacent unit cell, the Switch I loop of KRASG12D structure extends into the Switch I/II pocket of another HisKRASG12D, forming an intricate interface akin to high-affinity protein-protein complexes. Using alanine-scanning investigations and alchemical free energy perturbation estimates based on explicit-solvent molecular dynamics simulations, we investigate the contributions of amino acids at this interface. Given that they increased the protein-protein interaction by more than 1.5 kcal/mol, a number of interface amino acids were discovered to be hot spots. Many malignancies an appropriate site for the inhibitor to bind to, the KRAS mutants have appeared to be unbreakable for the last forty years. Numerous investigations aimed at targeting KRAS led to the effective development of sotorasib (AMG510), a KRAS G12C inhibitor that received FDA approval in 2021. The cysteine residue presents in the 12^{th} position, which contains the switch II pocket of KRAS, forms a covalent bond with sotorasib.

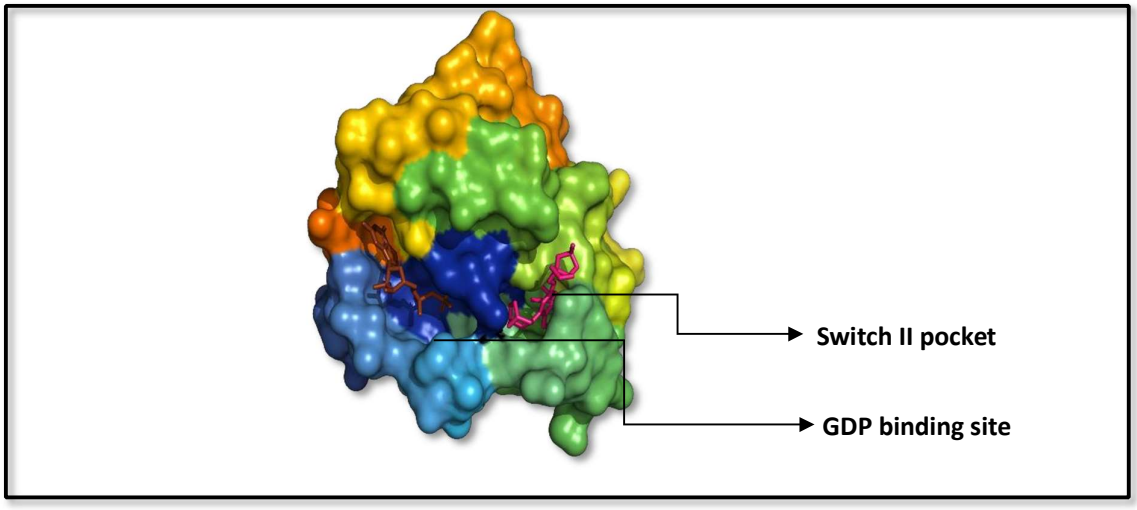


Figure 3: Showing the switch II pocket and GDP binding site of KRAS G12D

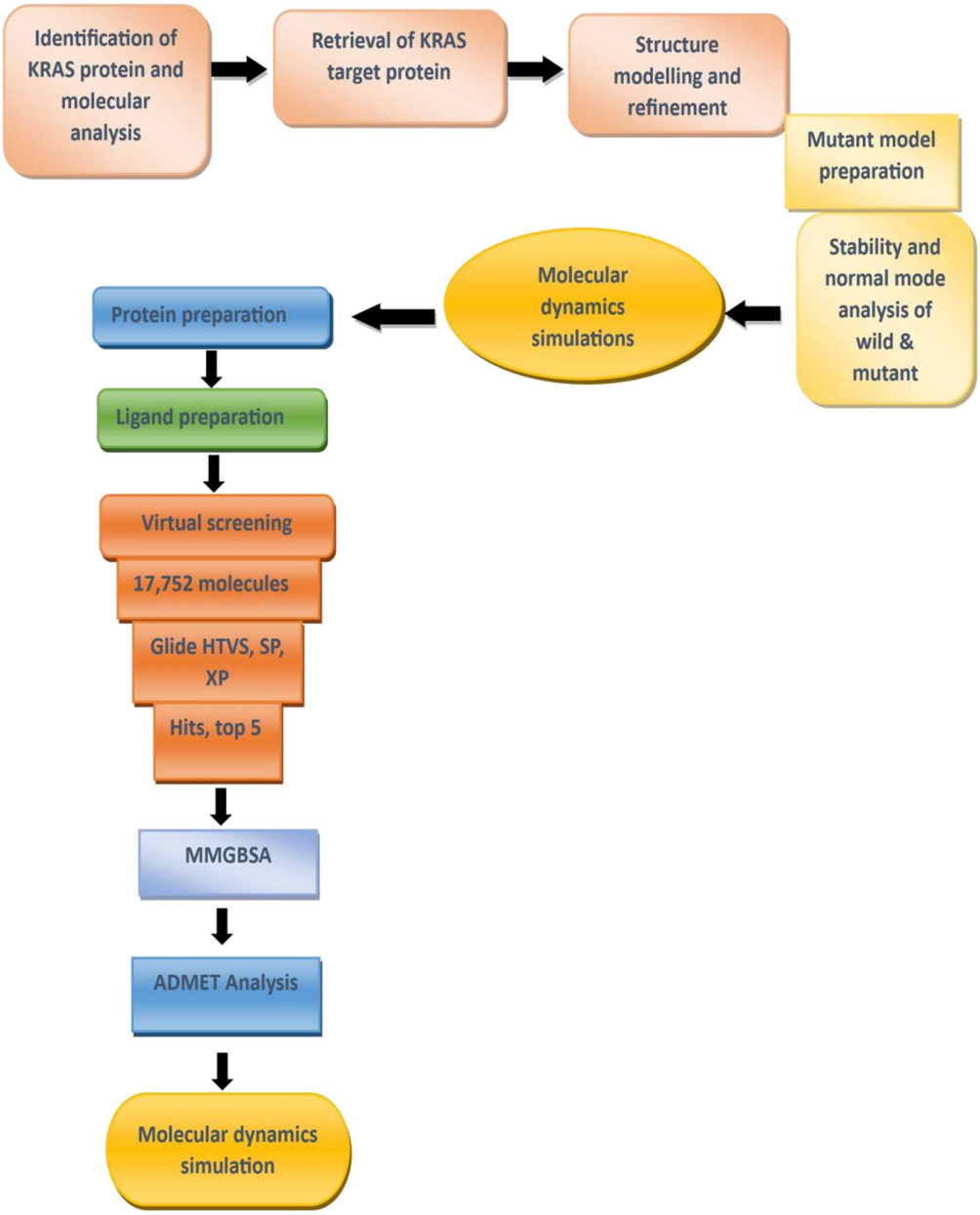
Similarly direct KRAS G12D inhibitors have been developed and studied. Few KRAS switch II pocket inhibitors are MRTX1133, BI-2852, and compound 3144 which still lacks effective inhibition. The MRTX1133 is one of the potent inhibitors of KRAS G12D, it targets the switch II pocket of KRAS G12D. From a recent study it has been reported that MRTX1133 is a potent inhibitor but still as a oral drug it is impractical. Because of its high molecular weight (600.6 g/mol) and other physio-chemical and ADME properties. MRTX1133 exhibits a high efflux ratio, or a high B-A rate in caco-2 permeability, which is detrimental to absorption through the gastrointestinal tract. In mice, MRTX1133 is only 0.5% at a dose of 30 mg per kilogram. Poor GI absorption is most likely the cause of mice's low oral bioavailability. Thus, MRTX1133's oral molecule will have limitations in many treatment aspects, including increased dosage requirements, high excipient loading, and pill burden.

To better understand the molecular mechanisms underlying the effects of mutations on functions of KRAS, we performed the molecular dynamics simulation of the wild type and the six KRAS codon 12 mutants. Molecular dynamics simulation is a powerful computational technique that can provide insights into the structural dynamics and stability. Through MD simulations we aimed to investigate how the mutations affecting the stability and dynamics of KRAS protein.

Further to investigate the effects of these mutations on the binding of natural compounds, we performed virtual screening for the wild type and the KRAS G12D mutant and also aimed to find some natural compounds as potential switch II inhibitor for KRAS G12D mutation as it is prevalence in human cancers and lacks standard drugs.

CHAPTER III

3.MATERIALS AND METHOD



3.1 Identification of KRAS protein and molecular analysis

The human KRAS protein belongs to the RAS GTPase family of proteins. In human cancers mutation in RAS protein is found in about 30% which is more than a quarter of total human cancers. KRAS is present in almost 80% of RAS mutated cancers. This GTPase protein is a small protein with 188 amino acids with a molecular weight of 21 kda. Three mutation hotspot regions are there in the protein, codon12, codon13 and the codon 61. Mutation is mostly found in the 12^{th} codon with six different type of amino acids substitution in place of glycine.

3.2 Retrieval of KRAS target protein

The full-length amino acid sequence of human KRAS protein was obtained from the Uniprot database with Uniprot ID P0116 in FASTA format, which can be accessed from here(https://www.uniprot.org/uniprotkb/P01116/). The Human KRAS 4B isoform is generally corresponds to as the KRAS protein. The full length of KRAS 4B is 188 amino acids.

3.3 Protein structure modelling, refinement and Validation

To get the modelled structures for KRAS protein, the 188 amino acid length sequence were submitted to Robetta (Kim et al.,2004)(https://robetta.bakerlab.org/), I-TASSER and Swiss model for predicting the protein structure Robetta uses ab initio method while the ITASSER uses threading algorithm and Swiss model predicts structures based on homology modelling.

The final stage used to enhance the quality of a predicted protein model is protein model refining. TO refine the generated model SPDV tool is used. For minimization of protein steepest descent and conjugate gradient method, each with 1000 steps were used. The structure validation has been done using online tools. The SAVES server was used for Ramachandran plot analysis for structure validation. The ProSA (Protein Structure Analysis) tool, which can be accessed at (https://prosa.services.came.sbg.ac.at/prosa.php), is widely utilized for the enhancement and verification of experimental protein structures, as well as for protein prediction and modelling purposes.

3.4 Mutant model preparation

For the preparation of the mutant models, the SPDV (Swiss Pdb viewer) was used. Which can be accessed from(https://spdbv.unil.ch/disclaim.html), version 4.1 was used. There are six mutations reported for the 12^{th} codon in the KRAS protein. Six mutant models have been generated using the MUTANT panel of the Swiss Pdb viewer, followed by energy minimization of all six mutant models.

3.5 Stability and normal mode analysis

The stability and NMA analysis of the mutant and the wild type structures have been performed using the DynaMut (Rodrigues et al. 2018). DynaMut combines graphbased signatures and normal mode dynamics to generate a consensus prediction regarding the impact of a mutation in protein stability.

3.6 Molecular dynamics simulation

Molecular Dynamics is a computer-based simulation method for analyzing the physical movements of atoms and molecules. The atoms and molecules are allowed to interact for a fixed period of time giving a view of the dynamic evolution of the system. It shows how atomic coordinates evolve under given conditions over a period of time. The simulation was performed using the GROMACS 2022.4 molecular dynamics package to analyze the molecular stability of KRAS wild-type and mutants. The OPLS-AA/L allatom force field was utilized. The periodic boundary condition was implemented in all dimensions, while the solvated systems were neutralized by adding proper NA and CL ions. Energy minimization was carried out using the steepest descent algorithm, followed by the conjugate gradient method. Following the same protocol for all the six mutants and the wild type a 100nanosecond molecular dynamics simulation run was performed.

3.7 Protein preparation

An x-ray crystal structure of KRAS G12D mutation with GDP and switch two inhibitor MRTX1133 with a sequence length of 169 amino acids was retrieved from the RCSB PDB with PDB code 7RPZ with a resolution of 1.95 Å and an observed R-value of which is in a standard range. A few criteria must be met in order to choose a specific protein from Uniprot

(www.uniprot.org/UniProt/P01116), including an acceptable resolution of the protein structure, the presence of a co-crystal, and the identification of the protein's most crucial organism of the protein must be *Homo sapiens*. A vital phase in the preparation of proteins is energy minimization, where net interatomic forces are reduced to almost nothing. Therefore, in the OPLS2005 force field in Maestro, the protein's best structural confirmation comes from the protein's minimization in molecular modeling.

After bringing the protein into the protein preparation wizard, the strategy for minimizing the protein's energy usage is initiated. The pH was 7.5 ± 0 and the high module was used to fill in the gaps in the chains and loops. The protein was optimized in PROPKA pH 7.5, and the OPLS2005 forcefield was used to minimize the protein and enforce the elimination of water beyond 3Å.

3.8 Receptor grid generation and validation of docking protocol

In order to generate the grid, the energy minimized structure underwent additional processing. The receptor grid generation panel, which illustrates the protein's binding pocket. The switch II binding ligand of the KRAS protein was chosen in order to carry out the receptor's grid generation, thereby excluding the ligand from both the receptor-ligand docking calculation and grid generation itself. Lastly, using the site's default settings, the receptor grid generation was put into practice, The RMSD of the docked co-crystal ligand and the energy-minimized co-crystal ligand was computed to validate the docking protocol using the co-crystal ligand and protein complex following the generation of the receptor grid.

3.9 Ligand preparation

1. In this study, natural product-based compounds are used as ligand molecules. The natural product-based library retrieved from the ChemDiV database (accessed on date August 25, 2023) was downloaded along with 95 anticancer phytochemicals from the IMPPAT (access date August 25, 2023) database. A total of 17,752 structures are imported into Maestro for energy minimization and 3D structures were generated using the LigPrep application. The elimination of ligand errors and the production of optimized structures that can be utilized for Glide screening and molecular dynamics simulation are the two main functions of LigPrep. Further the energy minimization of the 17,752 structures were carried out using OPLS2005 forcefield.

The 3D structure of KRAS G12D inhibitor MRTX1133 was imported and prepared in a similar manner using the OPLS2005 force field.

3.10 Structure-based virtual screening by molecular docking

A very effective technique for determining ligand hits, Glide which stands for grid-based ligand docking with energetics in the Schrodinger suite's maestro interface which also aids in lead optimization for structure based virtual screening through molecular docking. High throughput virtual screening (HTVS) was used to flexibly dock the compounds to the receptor for the first screening at a rate of two seconds per compound. The top 1000 molecules from the 17,752 molecules were then chosen for standard precision (SP) docking, which was done at a rate of ten seconds per compound. And the best 30 molecules were

docked using Glide XP (extra precision). The top 5 molecules were then chosen for additional examination.

The same Glide XP docking was also carried out for the KRAS G12D inhibitor MRTX1133.

3.11 MMGBSA

MMGBSA which stands for the mechanics generalized born surface area is used to calculate the relative free binding energy of the ligand and determines the ligand binding affinity toward the receptor .The Prime module of Schrödinger suite was used for the MMGBSA calculation of the 5 best molecules utilizing the OPLS2005 force field and VSGB 2.0 solvation model. In order to determine the binding free energy, the equation is as follows;

$$\Delta G(binding\ affinity) = \Delta G(solvation\ \ energy) + \Delta E(minimized\ energy) + \Delta G(surface\ area\ energies)$$

$\Delta G\ (solvation\ energy)$is the difference between the total solvation energies of the unligated

KRAS G12D and the corresponding inhibitors and the solvation energy of the KRAS G12D-ligand complex. The term "minimized energy," or ΔE, refers to the difference between the energy of the ligand-KRAS G12D complex and the total energy of the ligand and unligated KRAS G12D. $\Delta G\ (surface\ area\ energies)$ is the difference between the surface area energy of the KRAS G12D-inhibitor complex and the sum of energies for unligated KRAS G12D and respective inhibitors. Since the affinity for binding is free, it incorporates both enthalpy and entropy. The MMGBSA prime module was used to determine the binding energies and affinities of the protein-ligand complex, optimized free receptors, and free ligands. The ligand's strain energy was calculated and visualized using Prime's energy visualizer by immersing it in a solution that the VGSB suite had automatically generated.

3.12 Drug-likeness and ADMET predictions

The QikProp module in Maestro was used to analyze the ADMET properties. The drug-likeness properties of the top five molecules were compared to the standard KRAS G12D inhibitor MRTX1133, allowing drug-likeness to be evaluated using Lipinski's rule of five. QikProp predicts polar surface area, oral absorption, and the octanol/water partition coefficient in addition to analyzing drug-likeness and ADMET properties. For the top five

compounds and MRTX1133, it also forecasts their apparent caco-2 cell permeability, IC50 values, and aqueous solubility.

3.12 Molecular dynamics (MD) simulation

To evaluate the stability of the most successfully identified leads through comparison with the reference KRAS G12D inhibitor, MRTX1133. GROMACS 2023.2 is used for performing out the complexes' molecular dynamics simulation. With respect to the protein-ligand complexes, the MD simulation ran for 100 ns. Before the dynamics were finished, the system was submerged in a TIP3P solvent model. The system was prepared using the Charmm force field by placing it inside a cubic box and then adding NA and CL ions to neutralize it. Other parameters of the system such as temperature and pressure were maintained constantly at 300K and 1.01325 bar during the whole period of simulation.

CHAPTER - 4

4. Results and discussion

4.1 protein model validation

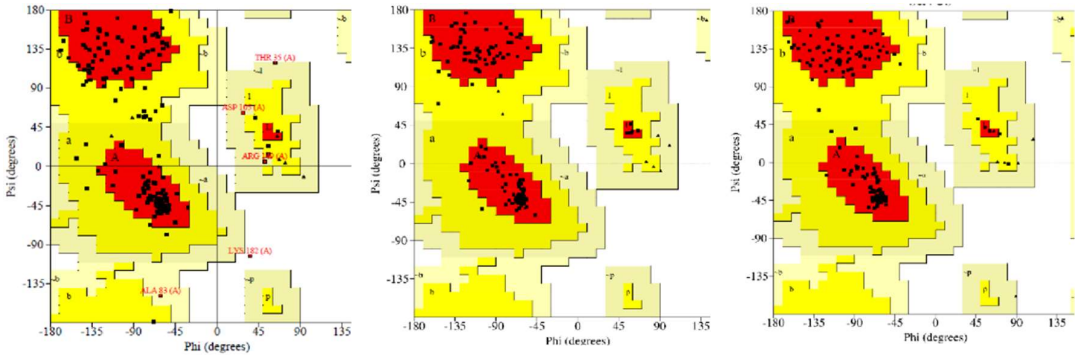


Figure 4: Ramachandran plot analysis of modelled structures (a)I-TASSER (b) Robetta (c) swiss model

4.2 mutant models generated using SPDV

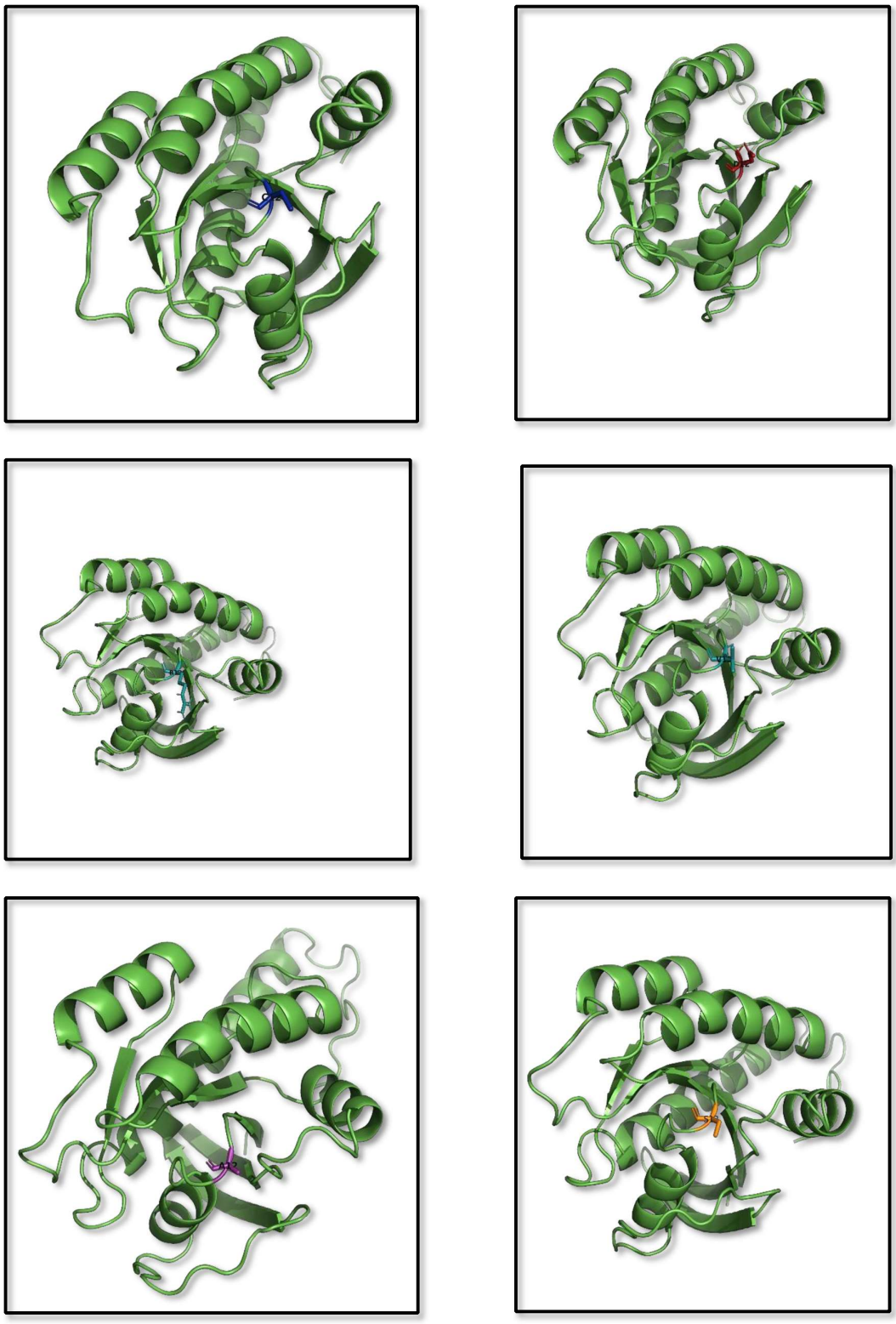


Figure 5: all six mutant models G12C, G12D, G12S, G12V, G12A, G12V

Table 1: Comparative summary of Ramachandran plot and PRosA and PRochecK Statistics

Models	I-TASSER	Robetta	Swiss model
Residue in the most favoured region	80%	95%	92.1%
Residue in disallowed region	0.6%	0.00%	0.00%
ProsA Z score	-5.6	-6.2	-5.1
Total residue	188	1888	168

From the Ramachandran plot statistics (Table3), it can be observed that the Robetta model is best among the three models with 95% residues in the most favorable region.

4.3 Stability and NMA analysis using DynaMut

Table 2: DynaMut results for stability analysis of wild and mutants in Kcal/mol

Mutant type	ΔΔG	Stability	ΔΔENcoM
G12D	0.402	stabilizing	0.798
G12C	1.024	stabilizing	0.927
G12V	1.425	stabilizing	0.956
G12S	0.227	stabilizing	-0.483
G12R	0.250	stabilizing	0.809
G12A	0.220	stabilizing	0.825
Wild KRAS	0.802	stabilizing	0.925

The six mutant and the wild type models were submitted in DynaMut for stability and NMA analysis. From the table in can be observed that the mutant G12S is the most unstable one with ΔΔENcoM score -0.483 kcal/mol, which represents the value of normal mode analysis of a protein structure. The most stable one is the mutant G12V, which can be depicted from its ΔΔG score and ΔΔENcoM score. Higher value represents higher stability. The G12V and G12C mutants were also showing higher score than the wild type making them more stable than the wild type. Overall, all the structures were in range of stability except the G12S mutant as it is having a negative ΔΔENcoM value.

4.4 Molecular dynamics simulation

I. Root Mean Square Deviation (RMSD) **WILD**

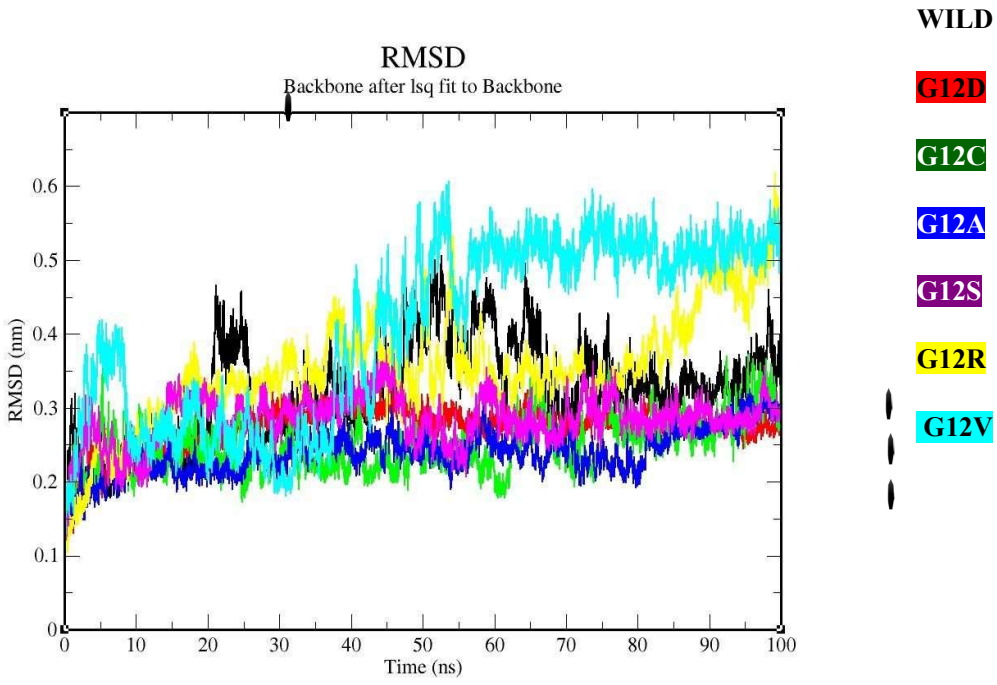


Figure 6: RMSD from MD simulation of wild and mutant proteins

The RMSD plot shows that in the wild type a sudden increase in RMSD value after 20ns from 2.1nm to 4.6nm with a continuous fluctuation for the next five nanoseconds. The highest RMSD value for the wild type is recorded at 50ns which is 5.1nm. It is observed that after 72ns the graph is gradually converged within a range of 0.25nm to 0.3nm. Of the six mutants, four (G12D, G12C, G12S, G12A) mutants show less RMSD value compared to the wild type indicating more stability. The G12A and G12C show RMSD values 0.2nm to 0.25nm whereas it is observed that the G12D and G12S show RMSD value in a similar range that is from 0.25nnm to 0.3nm.

On the other hand, two mutants, G12V and G12R initially demonstrated promising RMSD values. However, it was observed that the G12V mutant experienced a sudden increase in RMSD value after 40ns, while the G12R mutant displayed a similar trend after 80ns. These mutants exhibit higher RMSD values in contrast to the wild type. The graph shows that the G12V mutant reaches an RMSD value of 0.6nm at 50ns, which is almost twice that of the wild type and other mutants, while the G12R mutant reaches its highest RMSD value of 6.2nm at 50ns, which is the highest RMSD value among all indicating its significant instability.

II. Root Mean Square Fluctuation (RMSF)

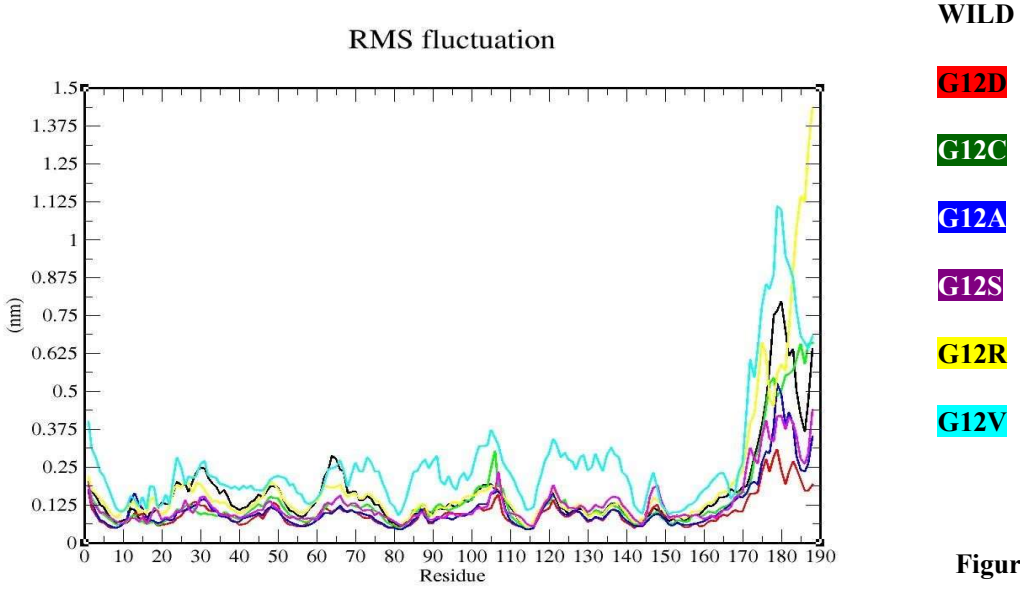


Figure 7: RMSF plot from MD simulation of wild and mutant proteins

Based on the RMSF plot, it seems that there is major fluctuation in the terminal residues after the 170th position across all the models. However, the four mutant models (G12D, G12C, G12A, G12S) seem to show lesser fluctuation when compared to the wild type. Additionally, up to the 170th position of all these four mutants, the RMSF value ranges from 0 to 0.125nm. On the other hand, the G12V mutant exhibits significant fluctuation throughout its entire structure, while the G12R mutant shows the highest fluctuation of 1.375nm at the last amino acid residue. It's worth noting that no significant fluctuation is observed at the 12th position for all of them, with G12A showing the highest value of 0.168nm

III. Radius of gyration (Rg)

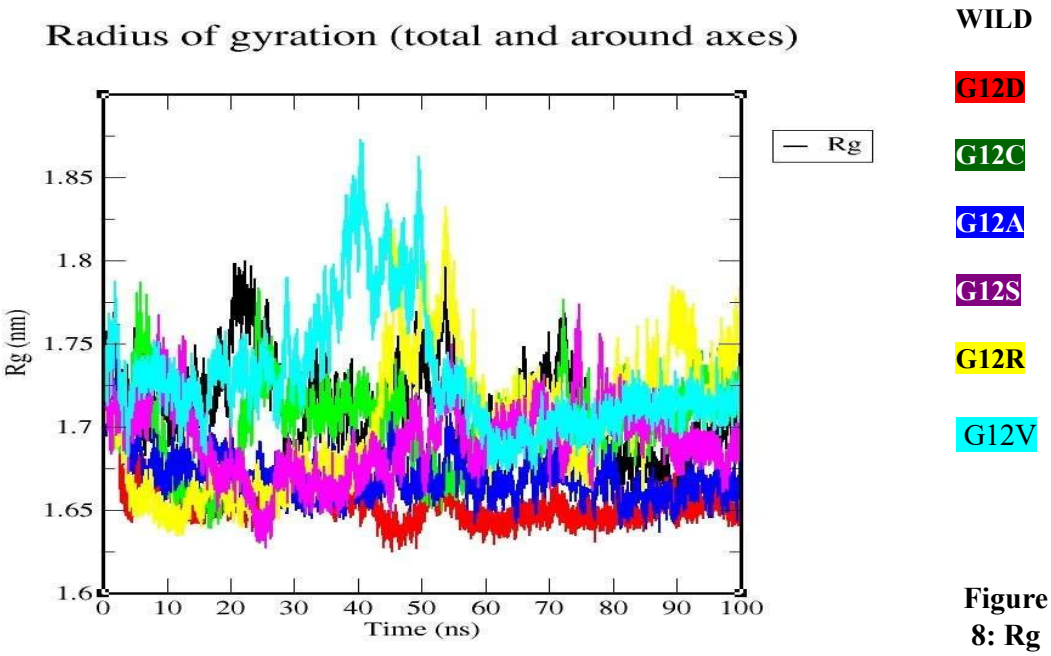


Figure 8: Rg from MD simulation of wild and mutant proteins

Based on the Rg plot, it appears that the structure being analyzed is fairly rigid. The majority of Rg values fall within a range of 1.65 to 1.7 nm, indicating consistency in the structure's shape. Interestingly, the wild type and G12C mutant show similar graphs, suggesting little variation between the two. However, the G12D mutant displayed the lowest Rg value of 1.62nm between 45ns to 50ns, indicating a potential stability in the structure during that time period. The G12V mutant showed the highest Rg value of 1.9nm and experienced significant fluctuations in its Rg values, indicating a lack of stability in the structure compared to the other mutants. After 70ns, it can be observed that both G12V and G12R mutants showed higher Rg values than the wild type. This suggests that these mutants may have a less stable structure compared to the wild type at this later time point.

4.4 Protein preparation, receptor grid generation and validation of docking protocol

Utilizing the protein preparation wizard of the maestro interface, the KRAS G12D protein was prepared at pH 7.5± 0 by applying the applications of Epik, ProtAssign, and Impref. For researchers, the protein preparation wizard streamlines multiple tools and the intricate process of energy minimization into an easy-to-use workflow. Further, the workflow for receptor grid generation described in the methodology was applied to the binding pocket of the co-crystal ligand for analysis in Glide based ligand docking. Prior to the structure-based virtual screening, the docking protocol underwent additional validation. This involved using the Glide module to dock the co-crystal ligands and check

if the experimental binding mode of the structure and the lowest energy state of the ligand predicted by glide were similar. The super imposition between the docked pose of the molecule is presented in fig 1. The RMSD between the two poses is less than 2Å.
Hence the docking protocol was validated.

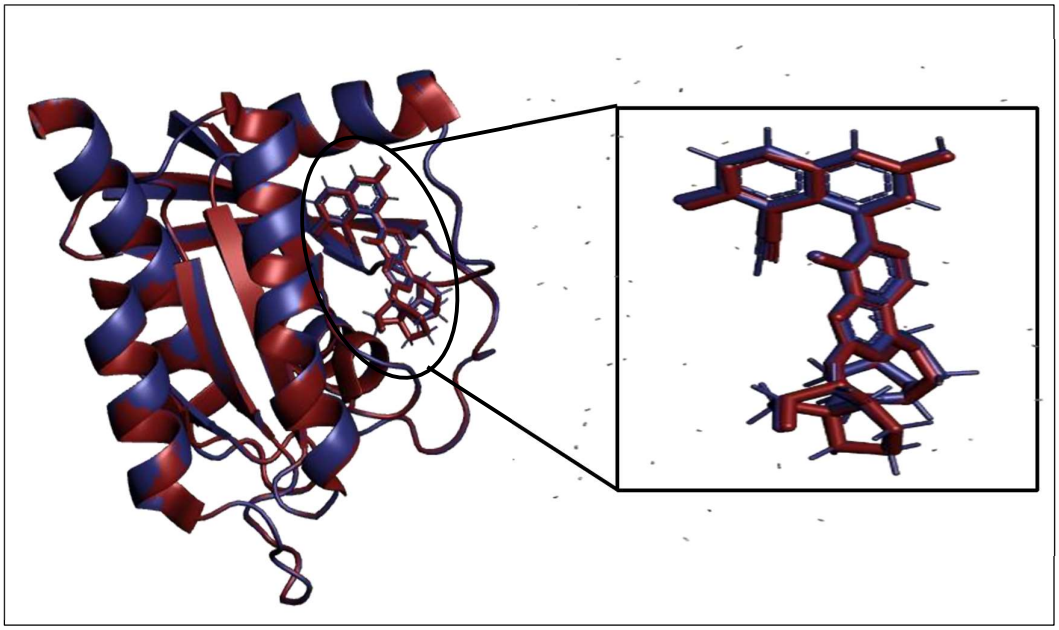


Figure 9: Validation of docking protocol

4.4 Structure based virtual screening by molecular docking

The screening of a large library of chemical compounds is time taking as well as less effective considering the false positive and negative rates. So, virtual screening using molecular docking is a preferred computational technique. The main advantage of this methodology is that it is rapid and time saving while requiring minimum investment. Furthermore, ligand preparation of the 18,752 molecules were carried out using Ligprep in the same pH of 7.5, same like the protein preparation.

The 18,752 molecules underwent structure-based virtual screening with Glide HTVS. Docking score was used to select the top 1000 molecules and further for SP docking and the top 30 molecules for XP docking, which allowed for the accurate study of ligand-receptor binding interactions in the same grid created with PDB ID 7RPZ. A quick understanding of energy minimization and ligand-protein affinity may be acknowledged by the ligand docking study.. We have selected XP docking in order to obtain more accurate results, as HTVS and SP may generate false positive results.

Compound IDs 5317471, 2851716, 10207, 46074039, and 5315830 are the top five molecules with docking scores greater than -8.5 kcal/mol; these molecules are designated as leads and are listed in Table1, representing their compound ID, molecular formulas, docking scores, and binding free energies. MMGBSA ΔG score with the protein.

Top five molecules exhibit scores for docking within the range of -8.593 kcal/mol to -9.553 kcal/mol, although the KRAS switch II inhibitor shows a docking score of -10.307.

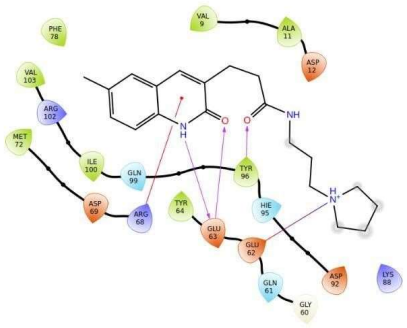
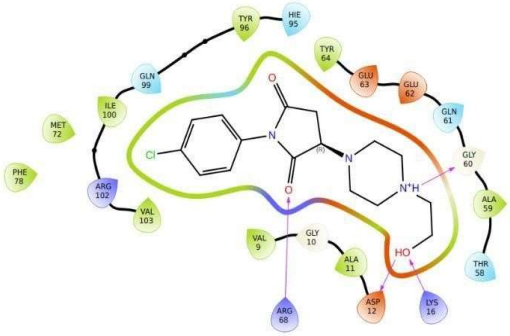
Table 3: docking scores and binding free energy of the top five natural compounds.

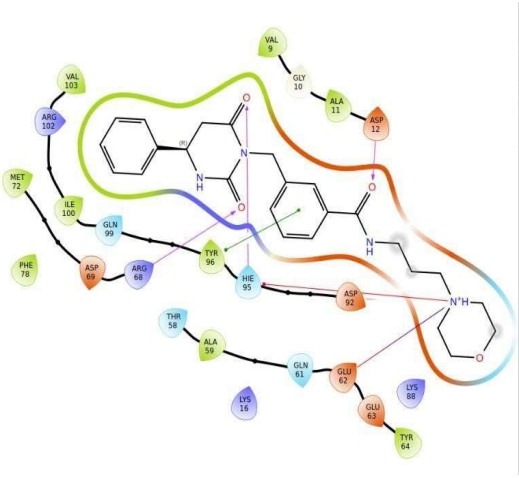
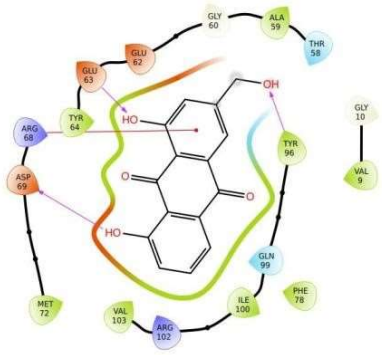
Compound (PubChem CID)	Molecular formula	Docking score (Kcal/mol)	MMGBSA ΔG (Kcal/mol)
531 74712	$C_{20}H_{27}N_3O_2$	-9.553	-85.26
10207	$C_{15}H_{10}0_5$	-9.094	-78.94
2851716	$C_{16}H_{20}CIN_3O_3$	-8.961	-76.83
46074039	$C_{25}H_{30}N_4O_4$	-8.649	-81.08
53158305	$C_{15}H_{15}FN_4O_2S$	-8.593	-87.54
MRTX1133	$C_{33}H_{31}F_3N_6O_2$	-10.307	-134.37

MMGBSA

From the table the MMGBSA values of the reference and the top five lead compounds can be observed. The molecular mechanics burn generalised surface area represents the binding free energy of protein ligand complexes. The binding free energy is represented in Kcal/mol. The reference molecule MRTX1133 showed the highest binding free energy as among all the ligand molecules. The MMGBSA for the lead compounds with PubChem ID 53174712, 10207, 2851716, 46074039, 53158305 are, 85.26, -78.94, -76.83, -81.08, -87.54 respectively. More negative the MMGBSA value greater the binding free energy. The MMGBSA scores of the lead compounds are also in standard range representing their ability for considering as potential lead compounds

2D interactions of the top five leads

Compound name	2D interactions
Lead 1		H bond: (03) Glu62(2) Tyr96 Hydrophobic: Val9, Ala11, Tyr96, Tyr64, Phe78, Ile100, Val103 Charged(-ve):Asp12, Glu62, Glu63, Asp69 Charged(+ve):Arg68,Lys88, Arg102 Pi-cation: Arg68 Salt-bridge: Glu62 Polar: Gln61,
Lead 2		H bond:(04) Asp12, Gly60, Lys16, Arg68, Hydrophobic: Val9, Ala11, Ala59, Tyr64, Tyr96, Met72, Phe78, Ile100, Val103 Charged(-ve): Asp12, Glu62, Glu63, Charged(+ve):Lys16, Arg68, Arg102 Polar: Thr58, Gln61, Hie95, Gln99

Lead 4		H bond: (03) Asp12, Arg68 Hie95 Hydrophobic: Val9, Ala11, Ala59, Tyr64 Met72, Phe78, Tyr96, Ile100, Val103 Charged(-ve): Asp12, Glu62, Glu64, Asp69 Charged(+ve): Lys16, Lys88, Arg68, Arg102 Salt bridge: Glu62 Pication: Hie95 Pi-pi stacking: Tyr96 Polar: Thr58, Gln61, Hie95, Gln99
Lead 3		H bond: (02) Glu63, Asp69 Pi- cation: Arg68, Hydrophobic: Val9, Ala59, Tyr64, Met72, Ile100, Va103, Charged(-ve): Asp69, Glu62, Glu63, Charged (+ve): Arg68, Arg102 Polar: Thr58, Gln99

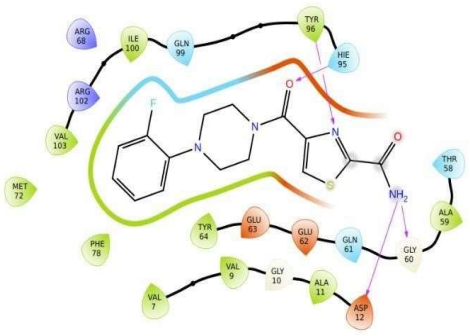
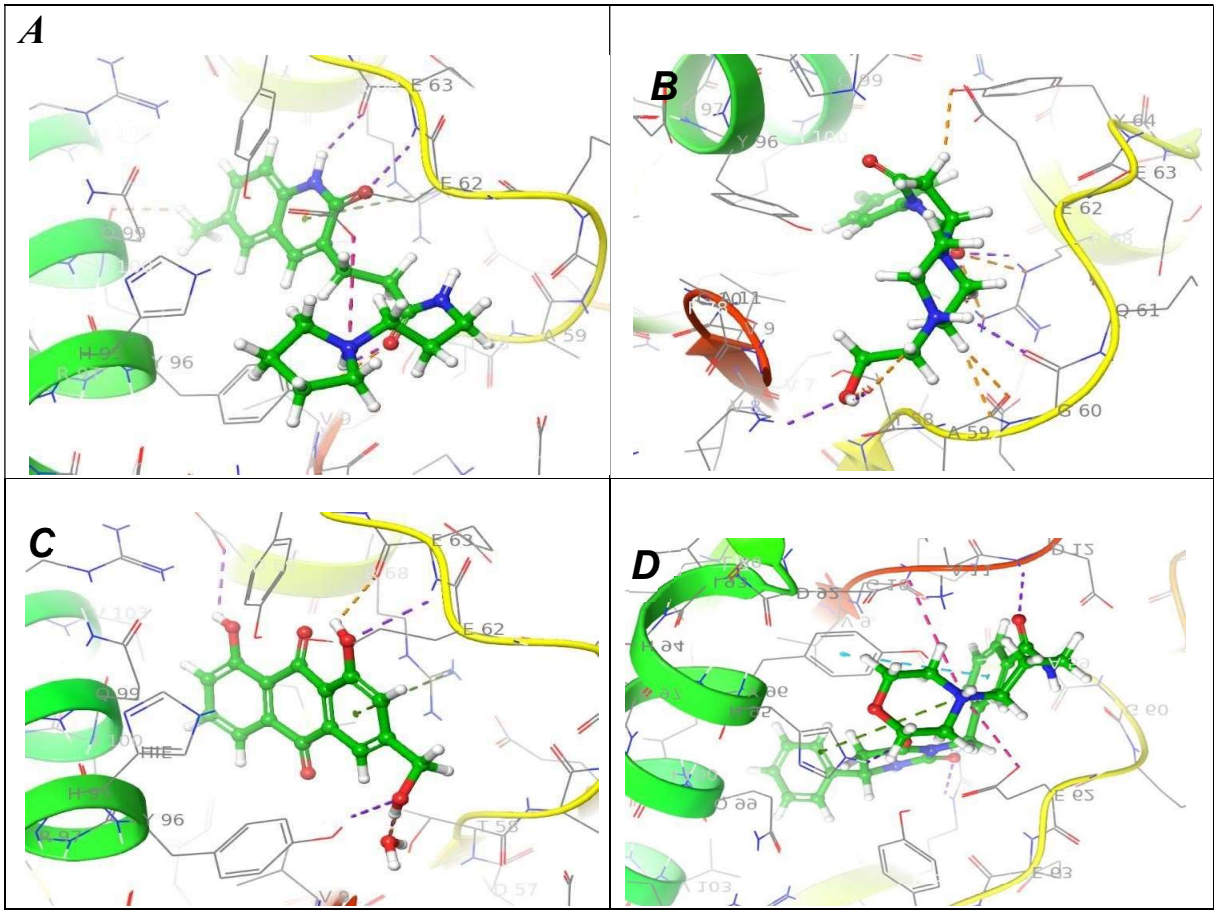
Lead 5		H bond: Asp12, Gly60, Hie95, Tyr96 Hydrophobic: Val7, Val9, Ala11, Ala59, Tyr64, Met72, Phe78, Tyr96, Ile100, Val103 Charged(+ve): Arg68, Arg102 Charged(-ve): Asp12, Glu62, Glu63,

figure 10: Showing protein ligand interactions in 2D for all the five lead molecules

The top five molecules of the natural product based library with PubChem ID 53174712, 2851716,10207, 460740839, 53158305 demonstrated substantial interactions when compared to every other molecule in the compound library, in the switch II binding pocket of KRAS G12D (PDB ID:7RPZ) based on the analysis of the top five compounds grid based extra precision docking, it can be concluded that Asp12 amino acid residue was most actively found in hydrogen bonding interactions, followed by Glu60, Arg68, Tyr95, and Hie 96.

Hydrophobic interactions are mostly shown by the Val9, Ala11, Ala59, Tyr64, Met72, Phe78, Tyr96 and the Val103 amino acid residues. The lead1(ID :53174712) compound also formed a salt bridge interaction with the Glu62 amino acid residue apart from the three hydrogen bond interactions. Only one ligand the lead4 (53158305) compound is showing a pi-pi stacking and pi-cation interaction with Tyr96. The +ve charge and -ve charge interactions of the lead compounds with Lys16, Arg68 and Asp12, Glu62 also showed various interactions.

The KRAS switch II inhibitor MRTX1133 also showed significance interaction in the binding. It can be observed that the top five lead compounds were showing a similar level of interaction in the Switch II pocket of KRAS G12D. Thus, it can be said that the top five lead molecules have the potential to effectively inhibit the protein.


A
B
C
D

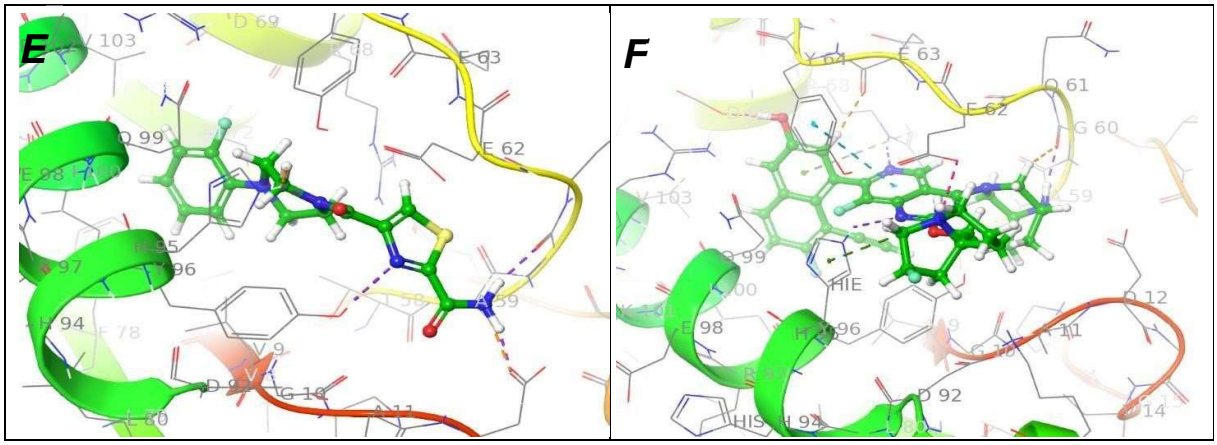


Figure 11: showing 3D interaction of the five lead molecules and the standard MRTX1133 in the switch II pocket region of KRAS G12D. A, B, C, D, E and F are the five lead molecules and the standard inhibitor respectively.

4.3 Drug likeness prediction and ADMET analysis

Based on analysis from docking scores, ligand receptor interactions, and MMGBSA calculations binding affinity for drug like proporties and ADMET analysis, the top five molecules were chosen. The ADMET analysis was carried out with ADMET Lab 2.0 and the Maestro Qikpro application. The standard molecule and the top five lead compounds were shown in Table 4 along with other properties such as drug likeness. The physicochemical and lipophilicity properties were listed in Table 3.

Table 4 Physicochemical properties, Lipophilicity and Water solubility

Compound	*Lead1*	*Lead2*	*Lead3*	*Lead4*	*Lead5*	*Standard*
Mol weight (g/mol)	341.45	337.8	270.24	450.53	334.37	600.63
H bond acceptor	3	5	5	5	4	10
H bond donor	2	1	3	2	1	2
Heavy atoms	25	23	20	33	23	44

Molar refractivity	105.46	98.01	69.92	134.95	91.34	171.82
TPSA	65.2	64.09	94.83	90.98	107.77	86.64
Consensus Log Po/w	2.79	1.03	1.50	1.79	1.63	4.76
Log S (ESOL)	-3.05	-2.26	-3.04	-3.17	-3.04	-8.96
Aqueous solubility	Soluble	soluble	soluble	soluble	soluble	Poorly soluble

From Table 3 and Table 4 it was observed that the MRTX1133 is having a higher molecular weight which may make it unfit for used as a oral drug thus violating the Lipinski's rule of five for standard drugs. It has also three violations in Ghose rule. All the compounds along with the standard inhibitor were not permeable in the blood brain barrier but the lead1 compound is having blood brain permeability. The TPSA for all the molecules observed to be in optimal range for all the molecules

Table 5: Pharmacokinetics, Drug likeness and toxicity predictions

Properties	**Lead1**	**Lead2**	**Lead3**	**Lead4**	**Lead5**	**Standard**
GI absorption	*High*	*High*	*High*	*High*	*High*	*High*
BBB Permeant	*Yes*	*No*	*No*	*No*	*No*	*No*
P-gp Substrate	*Yes*	*Yes*	*No*	*Yes*	*No*	*No*
Log Kp	*-6.89*	*-7.91*	*-6.66*	*-8.08*	*-7.17*	*-6.31*
Lipinski's violation	*No*	*No*	*No*	*No*	*No*	*2 violations*
LC_{50}FM	*3.084*	*2.300*	*4.517*	*2.525*	*2.667*	*4.588*
LC_{50}DM	*5.082*	*3.330*	*6.716*	*3.691*	*4.825*	*6.875*

The dissolution of the active ingredient in the drug is the next step in the absorption process, which starts with the tablet or capsule disintegrating. Compounds with proper water solubility are good for complete oral absorption where low solubility is detrimental

for drug absorption. Solubility of all the compounds are listed in Table 3, it can be observed that the lead compounds are having better solubility as compared to the standard molecule.

The Log Po/w hold a prominent position with significant influence over hydrophobic binding to macromolecules, such as target receptors, and membrane permeability., or metabolizing enzymes. Log Po/w in the range 0 to 3 for lead like molecules are considered as proper. The Log Po/w value of the five lead compounds ranges from 1 to 2.8, which is in the optimal range whereas the standard molecule is having a higher Log Po/w value of 4.76 indicating its lower membrane permeability which is one of the essential properties of drug molecules. Considering the Drug likeness and ADME parameters the lead compounds could be potent switch II inhibitors for KRAS G12D.

Molecular dynamics simulation:

To predict the stability of the compounds and protein ligand complexes molecular dynamics simulation is the preferred computational method. It also mimics the stability of the complexes inside human body.

1.Root mean square deviation (RMSD)

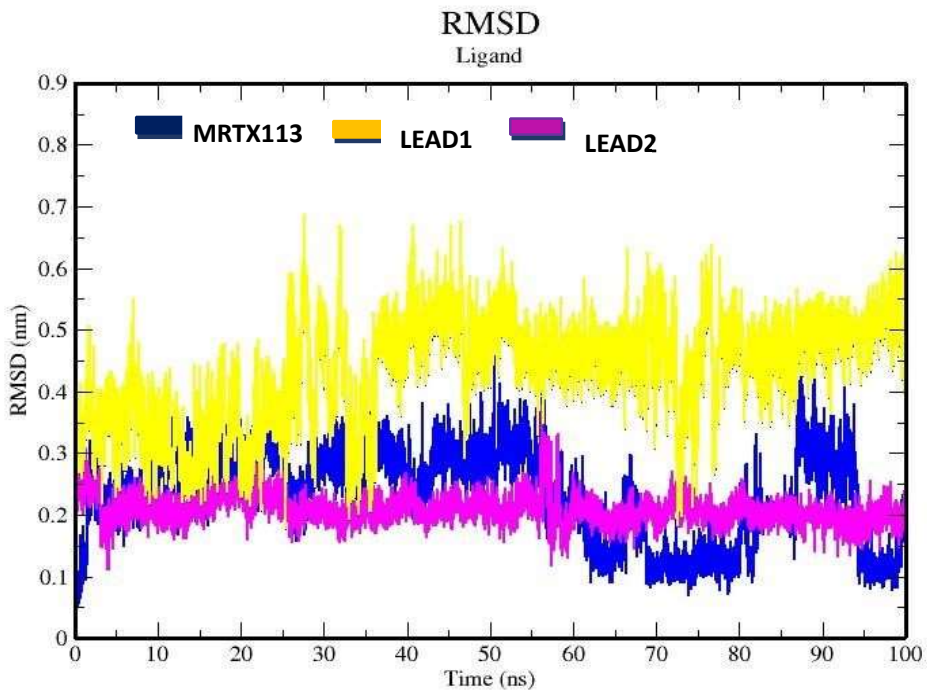


Figure 12: RMSD plot of top two lead and standard molecule

The RMSD plot generated from the protein ligand simulation of the stand compound and the top two lead compounds were represented in figure. From the RMSD plot for the three ligands it can be observed that the lead1compound is showing RMSD value in a range of 0.2nm to 0.6nm for the first 30 seconds followed by fluctuations for the next couple of nano seconds. After 35ns the graph was gradually showed a constant RMSD value in a range of 0.4nm to 0.55nm till 70nano seconds then again it showed fluctuations between the time period of 70ns to 80ns and converged in a range of 0.35 to 5ns.

The RMSD value of standard compound MRTX1133 is showing variation in a range of 0.25nm to 0.35 nm. It can be observed that the highest fluctuation observed was at 50ns with a RMSD value of 0.45nm. Maximum stability for the system was observed after 50ns till 80ns with RMSD value in a range of 0.07nm to 0.29nm. After 80ns the graph showed a slight increase in RMSD till 90ns and gradually showing a decline in RMSD value.

From the plot it can be clearly noticed that the RMSD for the lead1 compound was not depicting much more variations throughout the 100 ns time period. The RMSD value is in a constant range for the first 55 nano seconds with a slight fluctuation in RMSD at the beginning. Highest RMSD for the compound was observed between 52ns to 58ns with a sudden increase of RMSD to 0.37nm and after 60ns again it showed RMSD in a constant range, which indicates its stability in the system.

For protein ligand complexes the ideal range for RMSD fluctuations is between 0.2nm to 0.3nm mostly it may also vary in different complexes. The lead1 compound showed RMSD variation in a greater range which depicts that the ligand binding pose has changed in a greater range throughout the course of simulation. Therefore, this ligand could be subjected to further studies with lead optimization to acquire potent KRAS G12D inhibitor. For the standard molecule also, it was observed that it showed less fluctuations in RMSD as compared to the lead1 compound but still for half of the time it showed RMSD value out of the ideal range, if we notice the RMSD plot for the lead3 compound it showed a constant RMSD and lower than the standard compound indicating the ligand binding pose has not changed much during the simulation and making it more efficient than the standard compound in terms of change in binding pose and stability. The simulation also revealed that the lead2 compound (ID 10207) could be potential lead compounds towards KRAS G12D inhibitors.

CHAPTER V

Conclusion:

KRAS protein humans plays a crucial role in cell signalling pathways, which regulates cell growth, differentiation and survival. It is encoded by the KRAS gene and belongs to the RAS family of proteins. KRAS is encoded by the KRAS gene, Mutation in KRAS gene is associated with several types of cancers including lungs cancer, pancreatic cancer, and colorectal cancer which make this protein an important target for cancer therapy. The major issue of concern towards KRAS mediated cancers is that there are no standard drugs available to treat this mutation except the KRAS G12C mutation. In 2021 FDA approved two drugs Sotorasib and Adagrasib for the treatment of KRAS cancer in with G12C mutations. Following decades of research and development of KRAS drugs, the FDA approved the first mutant selective KRAS G12C inhibitors, but there are no standard drugs for other KRAS mutated cancers.

In this study we aimed to discover the mutational landscape of KRAS oncogenic protein and to find how the different mutations of KRAS impacting changes towards its conformation and stability by a computational approach using molecular dynamics simulations to get a better understanding about this protein. There are six (G12C, G12D, G12V, G12S, G12R, G12A) different types mutations are reported for 12^{th} codon of human KRAS protein. To inspect the difference among the mutants and the wild KRAS protein molecular dynamics simulation was carried out. During the 100ns simulation it is observed that some mutants are more stable than the wild type throughout the period of simulation. The G12D, G12C and G12V mutants has shown more stability as compared to the wild type. Particularly the G12D mutant was more stable compared to all other structures which indicates its more chances of occurring in human cancers as compared to the others.

So, we have focused to find some potential inhibitors for KRAS G12D mutant from natural compounds as there are no approved drugs are available for this KRAS G12D mutated cancers, MRTX1133 is a known inhibitor for KRAS G12D but still it is in investigational studies and also reported that it shows less GI tract absorption in animal

models. Virtual screening was carried out with a library of 17,752 natural compounds and five compounds were taken as lead. The lead compounds were better in terms of ADME properties as compared to MRTX1133. Further from MD simulation we got one lead compound with compound ID 10207 to be more stable than MRTX1133 towards the binding pocket of KRAS G12D. From the study it can be concluded that further modification and lead optimization of these natural compounds may result in more binding affinity towards KRAS G12D mutations and can be used as potential lead compounds for further studies.

REFERENCES

Karnoub, A.E.; Weinberg, R.A. Ras oncogenes: Split personalities. Nat. Rev. Mol. Cell Biol. **2008**, 9, 517–531.

Cox, A.D.; Der, C.J. Ras history: The saga continues. Small GTPases **2010**, 1, 2–27. [

Jemal, A.; Siegel, R.;Ward, E.; Hao, Y.; Xu, J.; Thun, M.J. Cancer statistics. CA Cancer J. Clin. **2009**, 59, 225–249

Herdeis, L. *et al.* Stopping the beating heart of cancer: KRAS reviewed. *Curr. Opin. Struct. Biol.* **71**, 136–147.

Simanshu, D.K.; Nissley, D.V.; McCormick, F. RAS Proteins and Their Regulators in Human Disease. Cell **2017**, 170.

Capon DJ, Seeburg PH, McGrath JP, et al. Activation of Ki-ras2 gene in human colon and lung carcinomas by two different point mutations.Nature. 1983;304(5926):507–13 Sweet RW, Yokoyama S, Kamata T, Feramisco JR, Rosenberg M, Gross

M. The product of ras is a GTPase and the T24 oncogenic mutant isdeficient in this activity. Nature. 1984;311(5983):273–5.

Bourne HR, Sanders DA, McCormick F. The GTPase superfamily:conserved and molecular mechanism. Nature.1991;349(6305):117–27.

Kato K, Cox AD, Hisaka MM, Graham SM, Buss JE, Der CJ. Isoprenoid addition to Ras protein is the critical modification for its membrane association and transforming activity. Proc Natl Acad Sci U S A.1992;89(14):6403–7.

12 Clarke S. Protein isoprenylation and methylation at carboxyl-terminal cysteine residues. Annu Rev Biochem. 1992;61:355–86.

Nan X, Tamguney TM, Collisson EA, et al. Ras-GTP dimers activate the mitogenactivated protein kinase (MAPK) pathway. Proc Natl Acad Sci US A. 2015;112(26):7996–8001

Broek D, Toda T, Michaeli T, et al. The S. cerevisiae CDC25 gene product regulates the RAS/adenylate cyclase pathway. Cell. 1987;48(5):789–99.

Bonfini L, Karlovich CA, Dasgupta C, Banerjee U. The Son of sevenless gene product: a putative activator of Ras. Science.1992;255(5044):603–6.

Wei W, Mosteller RD, Sanyal P, et al. Identification of a mammalian gene structurally and functionally related to the CDC25 gene of Saccharomyces cerevisiae. Proc Natl Acad Sci U S A. 1992;89(15):7100–4..

Buday L, Downward J. Epidermal growth factor regulates p21ras through the formation of a complex of receptor, Grb2 adapter protein,and Sos nucleotide exchange factor. Cell. 1993;73(3):611–20.

Chardin P, Camonis JH, Gale NW, et al. Human Sos1: a guanine nucleotide exchange factor for Ras that binds to GRB2. Science.1993;260(5112):1338–43. Gale NW, Kaplan S, Lowenstein EJ, Schlessinger J, Bar-Sagi D. Grb2

mediates the EGF-dependent activation of guanine nucleotide exchange on Ras. Nature. 1993;363(6424):88–92.

Egan SE, Giddings BW, Brooks MW, Buday L, Sizeland AM, Weinberg

RA. Association of Sos Ras exchange protein with Grb2 is implicated in tyrosine kinase signal transduction

Boriack-Sjodin PA, Margarit SM, Bar-Sagi D, Kuriyan J. The structural basis of the activation of Ras by Sos. Nature. 1998;394(6691):337–43.

Boriack-Sjodin PA, Margarit SM, Bar-Sagi D, Kuriyan J. The structural basis of the activation of Ras by Sos. Nature. 1998;394(6691):337–43.

Shi ZQ, Yu DH, Park M, Marshall M, Feng GS. Molecular mechanism for the Shp-2 tyrosine phosphatase function in promoting growth factor stimulation of Erk activity. Mol Cell Biol. 2000;20(5):1526–36.

Tartaglia M, Mehler EL, Goldberg R, et al. Mutations in PTPN11, encoding the protein tyrosine phosphatase SHP-2, cause Noonan syndrome. Nat Genet. 2001;29(4):465–8.

Trahey M, McCormick F. A cytoplasmic protein stimulates normal N-ras p21 GTPase, but does not affect oncogenic mutants. Science.1987;238(4826):542–5.

Takacs T, Kudlik G, Kurilla A, Szeder B, Buday L, Vas V. The effects of mutant Ras proteins on the cell signalome. Cancer Metastasis Rev.2020;39(4):1051–65.

Eblen ST. Extracellular-regulated kinases: signaling from Ras to ERK substrates to control biological outcomes. Adv Cancer Res.2018;138:99–142.

Fruman DA, Chiu H, Hopkins BD, Bagrodia S, Cantley LC, Abraham RT.

The PI3K pathway in human disease. Cell. 2017;170(4):605–35.

Fresno Vara JA, Casado E, de Castro J, Cejas P, Belda-Iniesta C, Gonzalez-Baron M. PI3K/Akt signalling pathway and cancer. Cancer Treat Rev.2004;30(2):193–204.

Hofer F, Fields S, Schneider C, Martin GS. Activated Ras interacts with the Ral guanine nucleotide dissociation stimulator. Proc Natl Acad Sci US A. 1994;91(23):11089–93.

Lambert JM, Lambert QT, Reuther GW, et al. Tiam1 mediates Ras activation of Rac by a PI(3)K-independent mechanism. Nat Cell Biol. 2002;4(8):621–5

Yokota J, Kohno T. Molecular footprints of human lung cancer progression.Cancer Sci. 2004;95(3):197–204.

Lohr M, Kloppel G, Maisonneuve P, Lowenfels AB, Luttges J. Frequency of K-ras mutations in pancreatic intraductal neoplasias associated withpancreatic ductal adenocarcinoma and chronic pancreatitis: a metaanalysis.Neoplasia. 2005;7(1):17–23.

Markowitz SD, Bertagnolli MM. Molecular origins of cancer: molecular basis of colorectal cancer. N Engl J Med. 2009;361(25):2449–60. S. Reddy Eda, Bioinformation, 2019, 15, 709–715.

K. Roos, C. Wu, W. Damm, M. Reboul, J. M. Stevenson, C. Lu,M. K. Dahlgren, S. Mondal, W. Chen, L. Wang, R. Abel,R. A. Friesner and E. D. Harder, J. Chem. Theory Comput.,2019, 15, 1863–1874.

Maestro User Manual, Schr¨odinger Press, 2015, pp. 1–337.

G. Madhavi Sastry, M. Adzhigirey, T. Day, R. Annabhimoju and W. Sherman, J. Comput.-Aided Mol. Des., 2013, 27, 221-234.

Prime User Manual, Schr¨odinger Press, 2015, pp. 1–133.

MOLECULAR INSIGHTS INTO KRAS ONCOGENIC PROTEIN VARIANTS AND IDENTIFICATION OF KRAS G12D NATURAL INHIBITORS: A MOLECULAR DOCKING AND SIMULATION APPROACH

BIKASH CHANDRA SAHU

Admission No. - 212121320

Submission date: 18-Nov-2023 04:12PM (UTC+0530)
Submission ID: 2232120106
File name: bikash_thesis_3.pdf (2.47M)
Word count: 9183
Character count: 48642

Plag-bikash thesis 3

ORIGINALITY REPORT

15%	11%	12%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.ncbi.nlm.nih.gov Internet Source	5%
2	krishikosh.egranth.ac.in Internet Source	2%
3	Prasanna Srinivasan Ramalingam, Purushothaman Balakrishnan, Senthilnathan Rajendran, Arunachalam Jothi et al. "Identification of Dietary Bioflavonoids as Potential Inhibitors against KRAS G12D Mutant—Novel Insights from Computer-Aided Drug Discovery", Current Issues in Molecular Biology, 2023 Publication	1%
4	pubs.rsc.org Internet Source	1%
5	Qingci Zhao, Ryoka Haga, Satoko Tamura, Ichio Shimada, Noritaka Nishida. "Real-time monitoring of the reaction of KRAS G12C mutant specific covalent inhibitor by in vitro and in-cell NMR spectroscopy", Scientific Reports, 2023	1%

	Publication	
6	Chunhua Yuan, Alexandar L. Hansen, Lei Bruschweiler-Li, Rafael Brüschweiler. "NMR 1H, 13C, 15N backbone resonance assignments of wild-type human K-Ras and its oncogenic mutants G12D and G12C bound to GTP", Biomolecular NMR Assignments, 2023 Publication	1 %
7	ici2016.org Internet Source	<1 %
8	www.frontiersin.org Internet Source	<1 %
9	journals.plos.org Internet Source	<1 %
10	www.osti.gov Internet Source	<1 %
11	Karnati Konda Reddy, Poonam Singh, Sanjeev Kumar Singh. "Blocking the interaction between HIV-1 integrase and human LEDGF/p75: mutational studies, virtual screening and molecular dynamics simulations", Molecular BioSystems, 2014 Publication	<1 %
12	Debojyoti Halder, Subham Das, R. S. Jeyaprakash. "Identification of natural product as selective PI3Kα inhibitor against NSCLC: multi-ligand pharmacophore modeling,	<1 %

	molecular docking, ADME, DFT, and MD simulations", Molecular Diversity, 2023 Publication Submitted to University of Otago Student Paper	<1 %
14	www.mdpi.com Internet Source	<1 %
15	E.J. Corey, Yong-Jin Wu. "Molecules Engineered Against Oncogenic Proteins and Cancer", Wiley, 2023 Publication	<1 %
16	Submitted to Napier University Student Paper	<1 %
17	edoc.ub.uni-muenchen.de Internet Source	<1 %
18	"Computational Modeling of Drugs Against Alzheimer's Disease", Springer Science and Business Media LLC, 2018 Publication	<1 %
19	jjbs.hu.edu.jo Internet Source	<1 %
20	Ferah Comert Onder, Serdar Durdagi, Kader Sahin, Bulent Ozpolat, Mehmet Ay. "Design, Synthesis, and Molecular Modeling Studies of Novel Coumarin Carboxamide Derivatives as	<1 %

	eEF-2K Inhibitors", Journal of Chemical Information and Modeling, 2020 Publication Submitted to Franklin & Marshall College Student Paper	<1 %
22	Submitted to La Trobe University Student Paper	<1 %
23	Debojyoti Halder, Subham Das, Aiswarya R., Jeyaprakash R. S.. "Molecular docking and dynamics based approach for the identification of kinase inhibitors targeting PI3Kα against non-small cell lung cancer: a computational study", RSC Advances, 2022 Publication	<1 %
24	Xile Zhou, Yang Ji, Jinming Zhou. "Multiple Strategies to Develop Small Molecular KRAS Directly Bound Inhibitors", Molecules, 2023 Publication	<1 %
25	mdpi-res.com Internet Source	<1 %
26	Kathryn Graham. "The Ras Signalling Pathway as a Target in Cancer Therapy", Resent Results in Cancer Research, 2007 Publication	<1 %
27	Abhijeet Kapoor, Gerard Martinez-Rosell,	<1

	Opioid Receptor Functional Selectivity", Scientific Reports, 2017 Publication pubmed.ncbi.nlm.nih.gov Internet Source	<1 %
29	Hoa B. Nguyen, Lawrence A. Quilliam. "Chapter 274 Rap GEF Family", Springer Science and Business Media LLC, 2012 Publication	<1 %
30	Jill Hallin, Vickie Bowcut, Andrew Calinisan, David M. Briere et al. "Anti-tumor efficacy of a potent and selective non-covalent KRASG12D inhibitor", Nature Medicine, 2022 Publication	<1 %
31	Xuewei Liu, Danfeng Shi, Shuangyan Zhou, Hongli Liu, Huanxiang Liu, Xiaojun Yao. "Molecular dynamics simulations and novel drug discovery", Expert Opinion on Drug Discovery, 2017 Publication	<1 %
32	ebin.pub Internet Source	<1 %
33	rcastoragev2.blob.core.windows.net Internet Source	<1 %
	Anupam Banerjee, Pralay Mitra. "Ebola Virus	<1

<1%

with Human PKR", Journal of Proteome Research, 2020
Publication

Parameswar Sahu, Rosaleen Sahoo, Abhishek Kumar Sahu, Sundeep Singh Saluja, Banshidhar Behera. " Repurposing phytochemicals of against maltase-glucoamylase using molecular docking, MMGBSA, MD simulation and linear regression to identify potential anti-diabetic compounds ", Journal of Biomolecular Structure and Dynamics, 2023
Publication

36 Priscilla Cascetta, Arianna Marinello, Chiara Lazzari, Vanesa Gregorc et al. "KRAS in NSCLC: State of the Art and Future Perspectives", Cancers, 2022 <1%
Publication

37 Demare, S.. "Accurate automated logP"o"/"w measurement by gradient-flow liquid-liquid partition chromatography", Journal of Chromatography A, 20071214 <1%
Publication

38 Kaushal Parikh, Giuseppe Banna, Stephen V. Liu, Alex Friedlaender, Aakash Desai, Vivek Subbiah, Alfredo Addeo. "Drugging KRAS: <1%

BIKASH CHANDRA SAHU | M.Sc. Bioinformatics

Orissa University of Agriculture and Technology,

Bhubaneswar

bikashchandrasahu1252000@gmail.com

+91(7608886119)

Objective

To be employed in a creative exciting and dynamic environment where I can contribute by applying the disciplined knowledge gained.Highly organized and hardworking individual looking for a responsible position to gain practical experience.

Area of interest

System Biology

Structural Bioinformatics

Programming language

Web Design

Education

2021 - 2023

Orissa University of Agriculture and Technology, Bhubaneswar

M Sc. Bioinformatics

2018- 2021

Utkal University, Bhubaneswar

Bachelor of science in Zoology

2016- 2018

Central Board Of Secondary Education, New Delhi

Intermediate Science/ +2

2016

Central Board Of Secondary Education, New Delhi

Matriculation/ 10th

Skills and techniques

Operating system - Windows , Linux(basics)

Database Programs - SQL

Programming language - C, R, Python (basics)

Tools - AutoDock, MEGA, Modeller

Software and techniques - MS office (Word,PowerPoint,Excel)

Workshop Attended

Computational Biology and Biomedical Informatics at National Institute of Technology, Assam

Hands-on training in Molecular Diagnosis Of Infectious Disease - Infections

Personal Details

Address- At/Po- Talagard,Dist-Angul, Odisha
D.O.B- 11th july 2001
Nationality-Indian
Gender-Male
Marital Status-Unmarried
Languages known - English,Odia,Hindi

EXPERIENCES

Completed M.Sc Dissertain Project under the guidance of Dr. K.M. Kumar at Department of Bioinformatics, Pondicherry University.

Declaration

I hereby declare that all the statements made above are true, complete and correct to the best of my knowledge and belief.

Date: 20.11.2023

Place: Bhubaneswar

Bikash chandra Sahu

Signature