

IN SILICO ANALYSIS FOR THE PREDICTION OF FUNCTIONAL NSSNPS FASN GENE IN WNT PATHWAY

Submitted in partial fulfillment of the requirements for the award of
Bachelor of Technology Degree in Bioinformatics

by

I.SHEBRIN DANI (Register No:42739023)



**DEPARTMENT OF BIOINFORMATICS
SCHOOL OF BIO AND CHEMICAL ENGINEERING**

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

**Accredited with Grade "A++" by NAAC
JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119**

APRIL 2024



SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited with Grade "A++" by NAAC
JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119
www.sathyabama.ac.in

DEPARTMENT OF BIOINFORMATICS

BONAFIDE CERTIFICATE

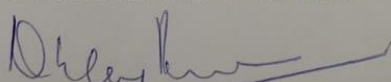
This is to certify that this project report is the bonafide work of **I.SHEBRIN DANI (42739023)** who carried out the project entitled **"IN SILICO ANALYSIS FOR THE PREDICTION OF FUNCTIONAL NSSNPS FASN GENE IN WNT PATHWAY"** under my supervision from January 2024 to April 2024.

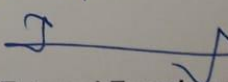

Internal Guide
(Dr. SwethaSunkar)


External Guide
(Dr. SwethaSunkar)


Head of the Department
(Dr. H.JEMMY CHRISTY)

Submitted for Viva voce Examination held on 20/04/2024


Internal Examiner


External Examiner

DECLARATION

I, **Shebrin Daniel (42739023)** hereby declare that the project report entitled "***IN SILICO* ANALYSIS FOR THE PREDICTION OF FUNCTIONAL NSSNPS FASN GENE IN WNT PATHWAY**" was done by me under the guidance of **Dr. Swetha Sunkar**, Head of the Department, Department of Bioinformatics, Sathyabama Institute of Science and Technology is submitted in partial fulfillment of the requirements for the award of Bachelor of Technology degree in Bioinformatics.

DATE : 20/4/2024

PLACE: Chennai

T. Shebrin Daniel

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to the board of management of **SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. REKHA CHAKRAVARTHY, Dean of the School of Science and Humanities** for providing me necessary support and details at the right time during the progressive reviews.

I sincerely thank **Dr.H.Jemmy Christy, Department of Bioinformatics, Sathyabama Institute of Science and Technology**, for his unwavering support and timely provision of necessary details during the progressive reviews. His support has been instrumental in helping me to overcome obstacles and achieve success in this project.

I would like to express my sincere and deep sense of gratitude to my project guide **Dr. Swetha Sunkar, Department of Bioinformatics, Sathyabama Institute of Science and Technology**, for her valuable guidance, and encouragement which paved way for the successful completion of my project work.

I wish to express my thanks to all faculty members and non-teaching staff members of the Department of Bioinformatics who were helpful in many ways for the completion of the project.

Heartfelt thanks to my lovable parents for their endless moral support, care, affection and blessings.

I.SHEBRIN DANI

ABSTRACT:

Cancer, a deadly disease in the current living, is caused by many factors, with common occurrences involving certain changes in genes that control cell growth and division. The Fatty Acid Synthase (FASN) gene, crucial for fatty acid biosynthesis, has been implicated in various cancers. In this study, we aimed to analyze potentially deleterious non-synonymous single nucleotide polymorphisms (nsSNPs) in the FASN gene to understand their role in cancer development and identify potential drug targets. From the dbSNP database, nsSNPs of the FASN gene were retrieved. Various computational tools including PolyPhen-2, SIFT, MUTPRED, Panther, and SNP & GO, DYNAMUT were employed to assess the pathogenicity of these nsSNPs and their potential impact on protein stability. Among the analyzed mutations, those predicted to be deleterious and destabilizing were prioritized. Homology modeling of both native and mutant FASN proteins was performed using SWISS MODEL, and the models were validated using PROCHECK. Structural comparisons between the native and mutant proteins were conducted using PDB Viewer, and VADAR server to analyze energy minimization, Root Mean Square Deviation (RMSD), and solvent accessibility (ASA). Additionally, the functional consequences of identified deleterious mutations on FASN enzymatic activity were explored. The analysis of pathogenicity identified a total of 32 mutations as deleterious. Subsequent stability assessment pinpointed 17 mutations that led to decreased stability within which the mutations L771M, R633C, R490H, and A641P were identified as potentially deleterious. Therefore, based on this computational analysis, the predicted non-synonymous single nucleotide polymorphisms (nsSNPs) might be implicated in disease mechanisms. Consequently, this protein emerges as a promising candidate for drug targeting, offering potential avenues for the treatment and enhanced comprehension of various human cancers.

KEYWORDS: FASN, PolyPhen-2, SIFT, MUTPRED, Panther, Homology modeling

TABLE OF CONTENTS

CHAPTER	TITLE	PAGE NO.
	ABSTRACT	i
	LIST OF FIGURES	v
	LIST OF TABLES	vi
	LIST OF ABBREVIATIONS	vii
1	INTRODUCTION	1
	1.1 Signs And Symptoms	2
	1.2 Classification	5
	1.2.1 Carcinoma	5
	1.2.2 Sarcoma	5
	1.2.3 Leukemia	5
	1.3 Common Types Of Cancer	5
	1.3.1 Breast Cancer	5
	1.3.2 Lung Cancer	6
	1.3.3 Cervical Cancer	7
	1.4 Wnt Pathway	8
	1.4.1 Wnt Signaling Mechanism	9
	1.4.2 Fasn Involved In Wnt Pathway	11
	1.4.2 Single Nucleotide Polymorphisms	12
2	LITERATURE SURVEY	14
	2.1 Wnt Pathway And Its Importance	14
	2.2 Wnt Pathway Mechanism	15
3	AIM AND SCOPE	16

3.1 Aim And Objective	16
3.1.1 Aim Of The Project	16
3.1.2 Objective Of The Project	16
3.2 Scope Of The Project	16
4 MATERIALS AND METHODS	17
4.1 Mining Of Data	17
4.2 Identification Of Deleterious Nssnps	17
4.2.1 Panther	17
4.2.2 Mutpred2	17
4.2.3 Polyphen	18
4.2.4 Sift	18
4.2.5 Snp & Go	18
4.3 Analyzing The Effects Of Nssnps On Stability Of Protein	19
4.4 Homology Modeling And Structural Analysis	19
4.5 Computation Of Accessible Surface Area (Asa)	19
4.6 Protein Ligand Docking	20
5 RESULTS AND DISCUSSIONS	22
5.1 Identification Of Functionally Significant Snps	22
5.2 Effects Of Nssnps On Stability Of Protein	24
5.3 Effects Of Nssnps On Stability Of Protein using DYNAMUT	25
5.4 Pairwise Structural Analysis Of Native And Mutant Models	26
5.5 Homology Modelling Of Native And Mutant Models	28
5.6 Validation Of Native And Mutant Model Structure Using Procheck	32
5.7 Analysis Of Accessible Surface Area (ASA) And Solvent Accessibility	34

	5.8 Dockthor Analysis	36
	5.9 Interactions	41
6	SUMMARY AND CONCLUSION	45
7	REFERENCES	46

LIST OF FIGURES

FIGURE	DESCRIPTION	PAGE NO
1	WNT PATHWAY MECHANISM	10
2	TYPES OF SNPS	13
3	HOMOLOGY MODELLED OF 3D STRUCTURE OF FASN	29-31
4	FASN NATIVE AND MUTANT 3D STRUCTURE USING DOCKTHOR	38-40
5	INTERACTIONS RESULTS	42-44

LIST OF TABLES

TABLE	DESCRIPTION	PAGE NO
1	IDENTIFICATION OF FUNCTIONALLY SIGNIFICANT SNPS	22-24
2	EFFECTS OF SELECTED NSSNPS OF FASN USING I-MUTANT 2.0	24--25
3	EFFECTS OF SELECTED NSSNPS OFFASN ON PROTEIN STABILITY USING DYNAMUT	26
4	PAIRWISE STRUCTURE ALIGNMENT VALUE RESULTS	27
5	VALIDATION OF THE STRUCTURES USING PROCHECK	33
6	ASA COMPUTATIONS OF PROTEIN	34-35
7	DOCKTHOR RESULTS	37
6	ASA COMPUTATIONS OF PROTEIN	34-35

LIST OF ABBREVIATIONS

ABBREVIATIONS	EXPANSION
AT	acetyl transacylase
MT	malonyl transacylase
DH	dehydratase
ER	enoyl reductase
TE	thioesterase
PKC	protein kinase C
TCF	t-cell factor
LEF	lymphoid enhancer facto
APC	adenomatous polyposis coli
PP2A	protein phosphatase 2A
GSK3	glycogen synthase kinase
CK1 α	casein kinase 1 α
PDZ	post synaptic density
FASN	fatty acid synthase
KS	β -ketoacyl synthase
AAS	amino acid substitution
FASTA	fast alignment search tool
SNP	single nucleotide polymorphism
SIFT	sorting intolerant from tolerant
GO	Gene ontology
PDB	Protein data bank
ASA	accessible surface area
RMSE	root mean square error

CHAPTER 1

INTRODUCTION

Cancer is a prevalent cause of death in adults, ranking as the second most common cause of mortality after heart disease. It is no longer viewed as a single disease but rather as a diverse group of over 100 diseases characterized by the uncontrolled proliferation and dissemination of abnormal cells. This classification recognizes cancer as a family of diseases where cells exhibit unregulated growth, possess the ability to invade nearby tissues, spread to other parts of the body, and manifest in various forms throughout the body (Ruggero, 2013). Cancer can originate in any part of the body, disrupting the normal process of cell growth and division. Under healthy conditions, cells grow, multiply, and die in an orderly manner to meet the body's needs. However, in cancer, this process is disrupted, leading to the survival of old or damaged cells and the formation of new cells when unnecessary. These abnormal cells continue to divide uncontrollably, giving rise to tumors. These aberrant cells collectively form a neoplasm, also known as a tumor, characterized by unregulated growth that can manifest as a mass or diffuse distribution. The hallmark of cancer lies in the development of malignant tumors from these neoplastic cells, marking a significant deviation from the normal cellular processes (Cooper, 2000). These hallmarks include continuous growth and division given even contrary signals, avoidance of programmed cell death, uncontrolled count of cell divisions, tissue invasion and formation of metastases. Cancer is a complex and devastating disease that affects millions of people worldwide (Gao & Ge, et., al.,2023). Every year, millions of people are diagnosed with cancer and many lose their lives to this deadly illness. The impact of cancer extends beyond individual health, affecting families, communities, and healthcare systems. Cancer research has become a crucial area of focus in the field of biomedical research due to its profound impact on society and the economy. The understanding of cancer has significantly evolved over the years, leading to advancements in diagnosis, treatment, and prevention. These advancements have brought us closer to the era of personalized medicine, where targeted therapies can be tailored to individual patients based on their specific type and variant of cancer. This review

presented an update on recent advances and breakthroughs in cancer therapies. These advances offer hope for improved outcomes and an increased life expectancy for cancer patients. The science of extremely small materials is poised to revolutionize cancer diagnostics, imaging, and treatment and could finally usher in the long-awaited era of personalized medicine

1.1 SIGNS AND SYMPTOMS

Signs and symptoms play a crucial role in indicating underlying health issues, injuries, or diseases, highlighting that something is amiss within the body. While signs and symptoms are often used interchangeably, they have distinct differences. A sign is observable by others, such as fast breathing, fever, or abnormal lung sounds detected with a stethoscope, providing external evidence of a disease. On the other hand, a symptom is experienced by the individual, like weakness, aches, or shortness of breath, which may not be apparent to others, reflecting internal sensations related to a disease.

In medical contexts, a single sign or symptom may not suffice to diagnose a condition accurately. For instance, a rash could indicate various issues like allergies, skin infections, or measles. However, when combined with other symptoms like high fever, sore throat, or chills, healthcare providers can better understand the underlying illness. In cases where symptoms are ambiguous, medical tests such as blood tests, x-rays, or biopsies may be necessary for a definitive diagnosis.

Cancer often presents with nonspecific symptoms, underscoring the importance of appropriate cancer screening to mitigate risk factors. Screening protocols are tailored to different age groups, and individuals with risk factors like smoking, alcohol consumption, sun exposure, or genetic predispositions should be vigilant about potential cancer symptoms. Regular screenings and heightened awareness of symptoms are vital in early cancer detection and improving treatment outcomes (Alwan et al., 2001). The best way to fight cancers is only by early detection and by eliminating and decreasing risk factors.

Accordingly, the individuals need to know which symptoms might lead to cancer. They should not ignore any warning symptom that might lead to early diagnosis and possibly to a cure. Sometimes cancer doesn't exhibit any signs or

symptoms that exclusively indicate the disease and also unfortunately every symptom or complaint of cancer can be explained by a harmless condition as well. Some of the common symptoms that may occur with cancer are as follows:

1. Persistent cough or blood-tinged saliva which usually represent simple infections such as sinusitis and bronchitis. They could even be symptoms of lung, neck and head.
2. A change in bowel habits which are dependent on the diet and fluid intake. Cancer may also exhibit diarrhea occasionally. An evaluation needs to be done for the significant changes in bowel habits which cannot be explained easily through dietary changes.
3. Blood in the stool which is frequently caused by hemorrhoids.
4. Anemia which should always be investigated is a condition where people have fewer red blood cells than expected in their blood. Sometimes this condition cannot be explained.
5. Unexpected weight loss where most people with cancer will lose weight at some point. This is supposed to happen most often with cancers of the pancreas, stomach, esophagus (swallowing tube), or lung.
6. Fever, very common with cancer, but more often happens after cancer has spread from where it started. If the cancer or its treatment affects the immune system almost all people with cancer will have fever at some time which can make it harder for the body to fight infection. In blood cancers like leukemia or lymphoma, less often, fever may be an early sign of cancer.
7. Fatigue, an extreme tiredness that doesn't even get better with rest. It appears to be an important symptom as cancer grows. But this may happen early in some cancers, like leukemia.
8. Pain might be an early symptom with some cancers like bone cancers or testicular cancer. This symptom might be able to persist a tumor or different cancers depending upon the occurrence of pain. For example, a headache that does not go away or get better with treatment may be a symptom of a brain tumor. Also back pain can be a symptom of cancer of the colon, rectum, or ovary.

9. Along with skin cancers, there are some other cancers that cause skin changes that can be seen. These signs and symptoms include:

- Darker looking skin (hyper pigmentation)
- Yellowish skin and eyes (jaundice)
- Reddened skin (erythema)
- Itching (pruritus)
- Excessive hair growth

10. Skin cancers sometimes bleed and look like sores that do not heal. A long-lasting sore in the mouth could possibly be an oral cancer that should be dealt with right away, especially in people who smoke, chew tobacco, or often drink alcohol. 11. Sometimes white patches and spots can be seen inside the mouth on the tongue that may be diagnosed as leukoplakia which is a precancerous area that's caused by frequent irritation. Often it is caused by smoking or other tobacco use. People smoking pipes or using oral or spitting tobacco are at a high risk for leukoplakia.

12. dyspepsia or swallowing issues that don't escape is also signs of cancer of the muscle system (the swallowing tube that goes to the stomach), stomach, or pharynx (throat).

13. Any changes in the skin or even any wart, mole or a freckle that may change a color, size and shape contribute to the list of symptoms. Melanoma can be taken as a skin change that can be treated if found early.

14. Hoarseness or Nagging cough could be a possible symptom. Any cough that does not go away may be a sign of lung cancer. Hoarseness could be a sign of cancer of the larynx (voice box) or thyroid gland.

The above enlisted signs and symptoms are the common ones seen with respect to cancer. But there are many others that are not listed here. Any changes that majorly affect the way your body works or the way you feel especially if it lasts for a long time or gets worse let a doctor know. If it has nothing to do with cancer, the doctors could possibly find out more about what's going on and, if needed, treat it. And if diagnosed as cancer, you will give yourself the chance to have it treated early, when treatment works best.

1.2 CLASSIFICATION

Cancers are classified depending on the site at which cancer started or the type of tissues where the cancer has originated. A person for example can have breast cancer, which is a type of carcinoma, or a cancer that arises from epithelial tissue, a type of tissue that forms a particular layer of the skin. The classifications of cancer by tissue type includes:

1.2.1 CARCINOMA:

A type of cancer that develops in epithelial tissues, such as those in the gastrointestinal tract and mucous membranes. 80 to 90 percent of cancer cases are of carcinomas estimated by National Cancer Institute. The epithelial cells are of different types and these cancers can develop into different types of carcinoma. They include squamous cell carcinoma, adenocarcinoma, transitional cell carcinoma, basal cell carcinoma.

1.2.2 SARCOMA:

A type of cancer that starts in connective tissues which are the supporting tissues of the body. These tissues include bone, cartilage, tendons and fibrous tissue that support organs. Sarcomas can be bone sarcomas or soft tissues sarcomas. Sarcomas are diagnosed much less than carcinomas. They make up less than 1 in every 100 cancers diagnosed every year

1.2.3 LEUKEMIAS:

These are the cancers of blood cells. It is a condition in which too many white blood cells are formed in bone marrow. They are not completely developed and so don't work properly. This type of cancer is very uncommon and makes up only 3 out of every 100 cancers diagnosed. They may be lymphomas and myeloma. There may be brain and spinal cord cancers also.

1.3. COMMON TYPES OF CANCER:

1.3.1 BREAST CANCER:

Breast cancer is a type of cancer that starts in the breast tissue, with the majority of cases occurring in women over 50 years old, but it can also affect men and younger women. The breast is made up of glands, ducts, fatty tissue, blood

vessels, and lymph vessels, and breast cancer can start in any of these parts. The most common types of breast cancer are invasive ductal carcinoma and invasive lobular carcinoma, which start in the ducts and lobules, respectively. Breast cancer is caused by changes in the DNA of breast cells, leading to uncontrolled cell growth and the formation of tumors. While the exact cause of most breast cancers is not known, certain risk factors have been identified, including a family history of breast cancer, personal history of breast cancer, certain breast conditions, and lifestyle factors such as obesity and alcohol consumption. Breast cancer can cause various symptoms, such as a breast lump or thickened area, nipple changes, skin changes, or changes in the size or shape of the breast. However, not all breast cancers cause noticeable symptoms, and some may be detected only through mammograms or other screening tests. The prognosis for breast cancer depends on several factors, including the stage of the cancer at diagnosis, the type of breast cancer, and the patient's overall health. The five-year survival rate for breast cancer varies depending on the stage at diagnosis, with a 99% survival rate for localized breast cancer, 86% for regional breast cancer, and 30% for distant breast cancer. Prevention and early detection are crucial in reducing the impact of breast cancer. Regular mammograms and clinical breast exams, as well as being aware of any changes in the breasts, can help detect breast cancer early, when it is more treatable. Lifestyle modifications, such as maintaining a healthy weight, limiting alcohol consumption, and engaging in regular physical activity, may also help reduce the risk of breast cancer.

1.3.2 LUNG CANCER:

Lung cancer is a malignant tumor that begins in the lung and is caused by genetic damage to the DNA of cells in the airways, often caused by cigarette smoking or inhaling damaging chemicals. The most common type of lung cancer is non-small cell lung cancer (NSCLC), which accounts for over 80% of lung cancer cases, and includes adenocarcinoma, squamous cell carcinoma, and large-cell carcinoma. Small cell lung cancer (SCLC) is less common but grows more quickly and is harder to treat. Lung cancer is the most diagnosed and deadliest cancer worldwide, with 2.2 million cases in 2020 resulting in 1.8 million deaths.

Most lung cancer cases are caused by tobacco smoking, but the remainder are caused by exposure to hazardous substances like asbestos and radon gas, or by genetic mutations that arise by chance. Consequently, lung cancer prevention efforts encourage people to avoid hazardous chemicals and quit smoking. Quitting smoking both reduces one's chance of developing lung cancer and improves treatment outcomes in those already diagnosed. Early lung cancer often has no symptoms, but when symptoms do arise, they are often nonspecific respiratory problems – coughing, shortness of breath, or chest pain. Treatment for early-stage lung cancer includes surgery to remove the tumor, sometimes followed by radiation therapy and chemotherapy to kill any remaining cancer cells. Later stage cancer is treated with radiation therapy and chemotherapy alongside drug treatments that target specific cancer cells. Even with treatment, only around 20% of people survive five years on from their diagnosis. Survival rates are higher in those diagnosed at an earlier stage, diagnosed at a younger age, and in women compared to men.

1.3.3 CERVICAL CANCER:

Cervical cancer is a type of cancer that originates in the cells of the cervix, the lower part of the uterus that connects to the vagina. It is primarily caused by the human papillomavirus (HPV), a common infection transmitted through sexual contact. Regular screenings, such as Pap tests, are crucial for early detection and prevention. Treatment options for cervical cancer include surgery, chemotherapy, and radiation therapy, depending on the stage of the cancer. The HPV vaccine is recommended for individuals aged 9 to 45 to protect against cervical cancer. Cervical cancer is highly treatable when detected early, with a five-year relative survival rate of over 90% for early-stage cases. It is essential to practice safe sex, get vaccinated against HPV, have routine Pap tests, and seek medical attention if any abnormal symptoms like vaginal bleeding, discharge, or pelvic pain occur.

1.4 WNT PATHWAY

Epigenetic and genetic changes play a crucial role in the development of cancers, enabling cells to undergo excessive proliferation and evade the usual regulatory mechanisms that govern their survival and movement. These alterations primarily occur within the signaling pathways responsible for controlling cell growth, division, death, fate, and motility. Among the numerous pathways implicated in cancer, notable ones include the JAK/STAT pathway, NOTCH signaling pathway, NF κ B pathway, TGF β pathway, and WNT pathway.(Sever et al., 2015).

Aberrant WNT signaling pathway is a common characteristic of tumors and also plays crucial roles in the progression and metastasis of various cancer types (Polakis, et., al.,2012). Multiple lines of evidence indicate that genetic mutations impact WNT pathway components, along with epigenetic mechanisms that regulate the inhibitors of WNT pathway. The rationale for investigating the WNT signaling pathway stems from the extensive evidence and prevalence of this pathway in different cancers. However, it remains relatively understudied within the South Indian population.

WNT, an abbreviation in genetics, stands for Wingless/Integrated. The WNT signaling pathways are signal transduction pathways that initiate with proteins transmitting signals into a cell via receptors located on the cell surface. There are three types of WNT signaling pathways: the canonical WNT pathway, the non-canonical WNT/calcium pathway, and the noncanonical planar cell polarity pathway. All three pathways are activated by the binding of a WNT protein ligand to a frizzled family receptor, which then transmits the signal to the disheveled protein inside the cell. Each pathway regulates gene transcription differently, with the canonical WNT pathway primarily responsible for this function (Zhang et al., 2010). The shape of a cell is determined by the non-canonical planar cell polarity pathway, which regulates the cytoskeleton. On the other hand, the non-canonical WNT/calcium pathway is responsible for regulating the levels of calcium within the cell. The WNT signaling pathway can involve either paracrine or autocrine communication, as described by Nusse in 2005. Initially, the WNT signaling

pathway was recognized for its involvement in carcinogenesis and later for its crucial role in embryonic development. The clinical significance of the WNT pathway was further demonstrated by the discovery of mutations that contribute to various diseases, such as breast and prostate cancer, type 2 diabetes, and others (Komiya et al., 2008).

Signaling pathways are a constant presence in mammalian cells, as highlighted by Logan et al. (2004). The WNT pathway, in particular, has been found to have significant implications in various cellular processes such as differentiation, proliferation, apoptosis, and overall cellular function. Moreover, the WNT signaling pathways have been identified as crucial players in the development, growth, and maintenance of homeostasis in animals (Johnson et al., 2006). WNT is a collection of secreted lipid-modified signaling glycoproteins that form a diverse family. These glycoproteins are approximately 350-400 amino acids long and undergo a specific type of lipid modification known as palmitoleoyl action of serine. This modification occurs in a conserved pattern of serine residues. The role of palmitoleoyl action is crucial as it facilitates the targeting of the WNT protein to the plasma membrane for secretion and enables the WNT protein to bind to its receptor through the covalent attachment of fatty acids.

The WNT proteins exhibit a high degree of conservation across various species, including mice, humans, and drosophila, among others. In humans, several WNT proteins have been identified, such as WNT1, WNT2, WNT2B, WNT3, and so forth. These same proteins are also present in mice. In vertebrates, there are a total of 19 known WNT genes that participate in four distinct signaling pathways. These pathways include the WNT/ β catenin or canonical pathway, the planar cell polarity pathway, the Ca^{+2} /protein kinase A pathway, and the pathway involving protein kinase C (PKC) in muscle myogenesis. (Chen et al., 2005).

1.4.1 WNT Signaling Mechanism

The canonical WNT pathway (or WNT/ β -catenin pathway) is the kind of pathway that causes an assemblage of β -catenin in the cytoplasm and its eventual translocation into the nucleus to act as a transcriptional coactivator of transcription factors that belong to the TCF/LEF family.

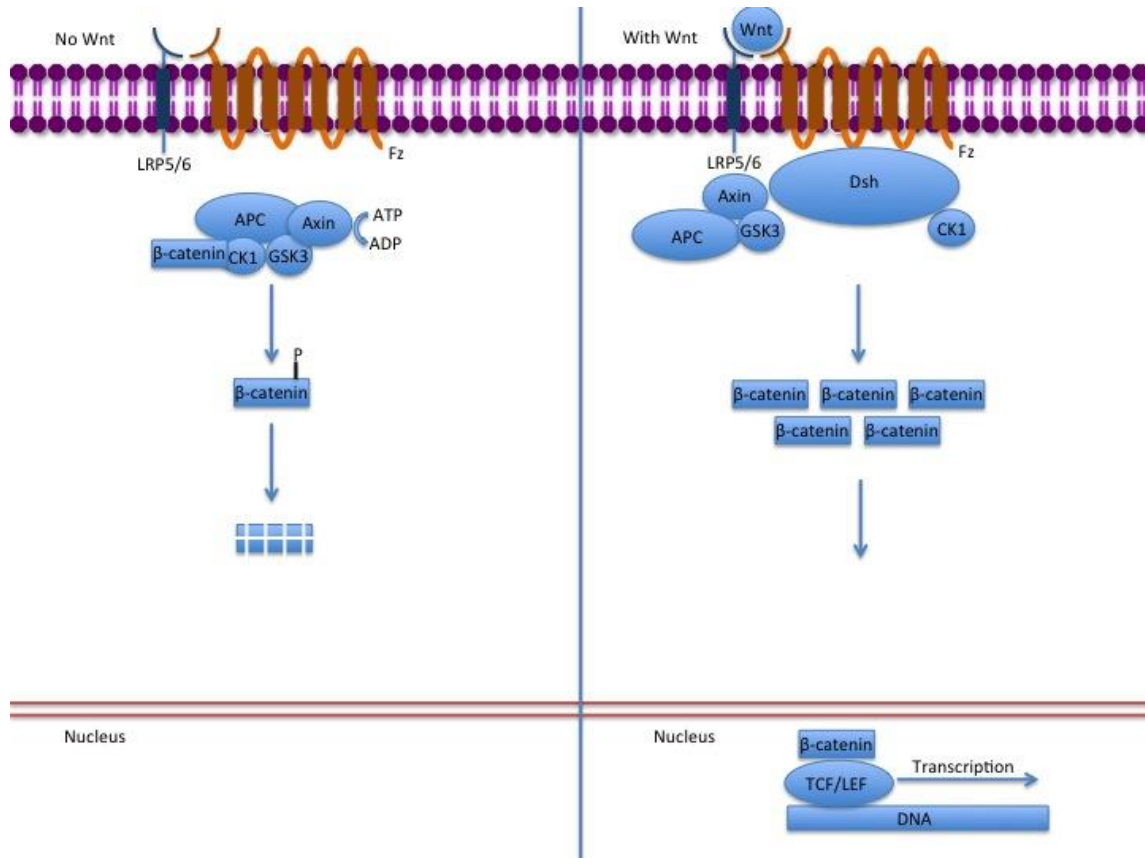


Figure 1.1: WNT pathway mechanism

Without WNT, β -catenin would not accumulate in the cytoplasm since a destruction complex would normally degrade it. This destruction complex includes the following proteins: Axin, adenomatous polyposis coli (APC), protein phosphatase 2A (PP2A), glycogen synthase kinase 3 (GSK3) and casein kinase 1 α (CK1 α) (Minde et al., 2013). This destruction complex degrades β -catenin through ubiquitination by targeting it, which eventually needs to be digested by proteasome. (Rao et al., 2010). However when WNT binds LRP5/6 and Fz the destruction complex function gets disrupted. This is because the WNT causes the translocation of negative WNT regulator, Axin, and the destruction complex to the plasma membrane. The other proteins in the destruction complex through phosphorylation subsequently bind axin to the cytoplasmic tail of LRP5/6. The stability levels of Axin decrease when de-phosphorylated. Dsh then becomes activated via phosphorylation and its DIX and PDZ domains inhibit the GSK3 activity of the destruction complex, which then allows beta catenin to accumulate and localize to the nucleus subsequently inducing a cellular response via gene transduction alongside the TCF/LEF (Staal et al., 2000) transcription factors

(MacDonald et al., 2009). Of the many genes present in the WNT a lot of mutations that are single nucleotide polymorphisms are discovered.

1.4.2 FASN INVOLVED IN WNT PATHWAY

The FASN gene encodes a protein called FASN (Fatty Acid Synthase) which is a multifunctional enzyme responsible for the synthesis of long-chain saturated fatty acids from acetyl-CoA and malonyl-CoA. It encompasses multiple functional domains including β -ketoacyl synthase (KS), acetyl/malonyl transacylase (AT/MT), dehydratase (DH), enoyl reductase (ER), and thioesterase (TE). FASN plays a crucial role in lipid metabolism and is highly expressed in tissues involved in fatty acid synthesis such as the liver, adipose tissue, and lactating mammary gland.

FASN is implicated in various physiological processes including energy storage, membrane formation, and cell signaling. Dysregulation of FASN expression and activity has been associated with various diseases including cancer, obesity, and metabolic syndrome. In cancer, FASN is often overexpressed and contributes to tumor growth, survival, and metastasis by providing building blocks for membrane biosynthesis and promoting oncogenic signaling pathways. The human FASN gene is located on chromosome 17q25.3 and spans approximately 59,992 base pairs of genomic DNA, consisting of 43 exons. FASN protein is a large polypeptide of 2,504 amino acids, featuring multiple functional domains crucial for fatty acid synthesis. The enzyme is predominantly localized in the cytoplasm where it catalyzes the sequential reactions involved in de novo fatty acid synthesis.

FASN inhibition has emerged as a potential therapeutic strategy for cancer and metabolic diseases. Several small molecule inhibitors targeting FASN have been developed and are being investigated for their efficacy in preclinical and clinical studies. Additionally, genetic variations in the FASN gene, including single nucleotide polymorphisms (SNPs), may influence FASN activity and contribute to disease susceptibility or treatment response. Computational analysis of FASN SNPs can provide insights into their potential functional effects and their

relevance to disease mechanisms.

1.5 SINGLE NUCLEOTIDE POLYMORPHISMS

A single nucleotide polymorphism typically short-noted to SNP is a substitution of any single nucleotide that occurs at any specific position in the genome. Each and every variation is present to some appreciable degree within a population.(Miller et al., 2005). For example, at a specific base position in the human genome, the C nucleotide may appear in most individuals, but in a minority of population, the position is occupied by an A which indicates that there is a SNP at this specific position, and the two possible nucleotide variations – C or A – are said to be alleles for this position. SNPs result in a wide range of diseases (Altshuler et al., 2000). A single base mutation in the APOE (apolipoprotein E) gene is associated with a lower risk for Alzheimer's disease. Single nucleotide polymorphisms might fall within coding sequences of genes, non coding regions of genes. Sometimes also seen in the regions between genes (intergenic regions). The SNPs within a coding sequence do not necessarily change the amino acid sequence of the protein due to degeneracy of genetic code (Delpont et al., 2008). The SNPs in the coding region are of two types. They are synonymous and nonsynonymous SNPs. Synonymous SNPs are those which do not affect the protein sequence, while nonsynonymous SNPs change the amino acid sequence of protein. The nonsynonymous SNPs again are of two types : missense and nonsense.

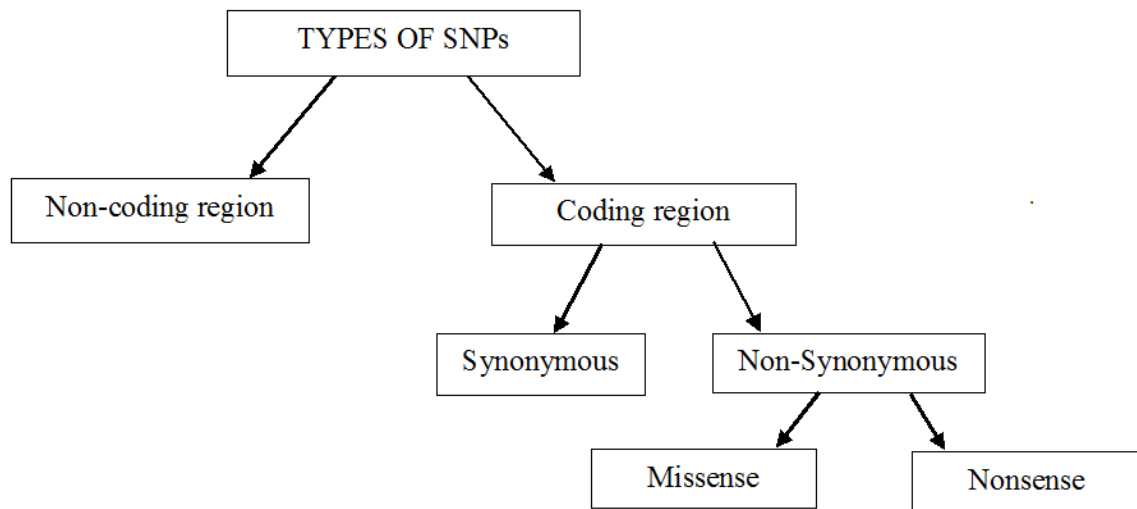


Figure 1.2: Types of SNPs

CHAPTER 2

LITERATURE SURVEY

2.1 WNT PATHWAY AND ITS IMPORTANCE

Noguti (et.al.,2012) stated that the WNT pathway holds significant significance in the progression, expansion, and sustenance of cells, tissues, and organisms. The WNTs encompass a vast assortment of remarkably preserved growth factors linked to numerous functionalities. Gaining comprehension of WNT signaling in cancer and the manner in which this pathway governs cellular conduct will contribute to the advancement of novel small molecular and pharmaceutical agents for the treatment of these ailments as well as various other diseases.

Dabhi (et.al.,2014) stated that the study identified potential functional non-synonymous single nucleotide polymorphisms (nsSNPs) in the TNF- α gene, focusing on nsSNPs rs4645843 and rs1800620. Computational analysis predicted these nsSNPs to be damaging, with decreased stability observed upon mutation. Structural analysis suggested that variants P84L and A94T may disrupt amino acid interactions and hydrogen bond networks, potentially contributing to functional deviations in the TNF- α protein. These findings provide insight into the genetic basis of inflammatory diseases and may aid in further research and therapeutic development.

Akhtar (et.al.,2019) stated that the study identified several potentially damaging nonsynonymous SNPs (nsSNPs) in the TOX3 gene associated with breast cancer. Notably, nsSNPs rs368713418 (A266D), rs751141352 (P273S, P273T), and rs200878352 (A275T) were identified as particularly deleterious. Additionally, premature stop codon producing SNPs and regulatory SNPs were also characterized. Our findings suggest that these SNPs could play a crucial role in breast cancer development and may have implications for personalized medicine and gene therapy. Further experimental validation in model organisms and cell cultures is warranted to fully understand their functional significance.

The study of Anasta(et.al.,2012) reported that the WNT signaling pathways and their downstream targets play a crucial role in controlling multiple aspects of cancer development, such as tumor formation, growth, cellular aging, apoptosis, differentiation, and spread. Despite the challenges in targeting WNT pathways directly, advancements in drug development methods and innovative technologies have paved the way for identifying compounds capable of modulating WNT signaling in experimental models, thereby laying the groundwork for potential clinical trials in humans.

2.2 WNT PATHWAY MECHANISM

Hocking (et.al.,1999) reported the relationship between WNTsignaling, the regulation of β -catenin stability, and oncogenesis and highlighted the importance of understanding how WNT signals modulate different cellular processes.

Boutros (et.al.,2002) reported the WNT pathways are involved in the control of gene expression, cell behavior, cell adhesion, and cell polarity. In addition, they often operate in combination with other signaling pathways. The WNT/ β catenin pathway is the best studied of the WNT pathways and is highly conserved through evolution. In this pathway, WNTsignaling inhibits the degradation of β -catenin, which can regulate transcription of a number of genes. Some of the genes regulated are those associated with cancer and other diseases.

AIM AND SCOPE

3.1 AIM AND OBJECTIVE

3.1.1 Aim :

The aim of this study is to predict the effect of mutations of genes involved in the WNT pathway and their structural and functional impact on their corresponding protein.

3.1.2 Objective :

- ✓ To identify the genes linked with the WNT pathway.
- ✓ To analyze the mutations in genes.
- ✓ To find their structural and functional impact on the protein.
- ✓ To find the protein – Ligand interactions

3.2 SCOPE :

WNT signaling pathway is a common feature of tumors and also plays important roles in tumor progression and metastasis of many cancer types. Various lines of evidence suggest that genetic defects affect WNT pathway components, as well as epigenetic mechanisms that modulate the suppressors of WNT pathway. The reason behind considering the WNT signalling pathway is due to the wide evidence and presence of this pathway in various cancers.

From the literature survey, it is known the mutations in WNT pathway genes are linked with cancer. On the basis of this, our present study was focused on the prediction of the deleterious mutations involved in the gene FASN of the WNT pathway using computational tools. In the disease condition, the particular gene mutations can possibly be used as targets. Further, experimental analysis should be performed for validating the association of these mutations with the disease.

CHAPTER 4

MATERIALS AND METHODS

4.1. MINING OF DATA

The information about the genes is collected from Entrez Gene from the National Center for Biological Information (NCBI) database. The information of SNP (reference sequence ID) and protein sequence (accession number) of the genes were retrieved from NCBI dbSNP (<https://www.ncbi.nlm.nih.gov/>) .

4.2. IDENTIFICATION OF DELETERIOUS NSSNPS

To ascertain the functional wallop (deleterious, damaging or neutral), coding nsSNPs were examined using four different tools. This assured the accuracy of results at the time of screening and likely noxious variations were selected and further subjected to other analyses.

4.2.1 PANTHER (*Protein Analysis through Evolutionary Relationship*)

The PANTHER tool was used to classify the proteins on the basis of their evolutionary relationship, molecular functions and interaction with other proteins (<http://www.pantherdb.org/tools.>) (Tang and Thomas, 2016). The PANTHER classification System was contrived to classify proteins and their genes in order to facilitate high-throughput analysis. Analysis was based on substitution position specific evolutionary conservation score, calculated from alignment of different evolutionarily related proteins. Protein FASTA sequence along with AAS (amino acid substitution) is submitted as a query.

4.2.2 MUTPRED 2

MutPred is a computational tool designed to analyze the impact of variations in protein sequences, particularly missense variants in humans. MUTPRED2 aims to predict whether a given mutation in a protein-coding gene will affect protein structure, function, or interactions, and to what extent.

4.2.3 POLYPHEN 2.0 (*Polymorphism Phenotyping*)

Functional effects of coding nsSNPs can be analyzed by using PolyPhen 2.0 (Adzhubei et al., 2010), freely available web-based tool. It is used for predicting the Functional Effect of Human Missense Mutations. Coding nsSNPs are categorized into three groups such as benign, probably damaging and possibly damaging based on the PSIC (position-specific independent count) score. Protein accession no followed by amino acid substitution represented in each row and submitted as query.

4.2.4 SIFT (*Sorting Intolerant from Tolerant*)

SIFT tool (Sim et al., 2012) is used to predict whether an amino acid substitution affects protein function based on sequence homology and the physical properties of amino acids. SIFT can be applied to naturally occurring non synonymous polymorphisms and laboratory-induced missense mutations. Reference SNP ids (rs ids) can be used as input.

4.2.5 SNP AND GO

SNPs&GO (Bendl et al., 2014) is a web server designed for predicting human disease-related single point protein mutations. This method utilizes support vector machines to analyze protein sequences, protein sequence profiles, and protein function, achieving an accuracy of 82% and a Matthews correlation coefficient of 0.63. SNPs&GO integrates information from protein sequences, profiles, and functions to predict whether a mutation is disease-related or not. It was trained on over 33,000 mutations and tested using a rigorous cross-validation process, demonstrating improved performance with increased input complexity. By incorporating data from the Gene Ontology annotation, specifically the LGO score derived from the GO database, SNPs&GO enhances its ability to distinguish disease-related mutations from neutral ones. The method's performance improves as more information is included in the analysis, highlighting the effectiveness of support vector machines in capturing correlations within diverse datasets. SNPs&GO stands out as one of the top classifiers for predicting the disease relevance of protein mutations based on in-house benchmarking studies.

4.3 ANALYZING THE EFFECTS OF NSSNPS ON STABILITY OF PROTEIN

DynaMut, (Carlos HM Rodrigues et al., 2018) a web server implementing two distinct, well-established normal mode approaches, which can be used to analyze and visualize protein dynamics by sampling conformations and assess the impact of mutations on protein dynamics and stability resulting from vibrational entropy changes. DynaMut integrates our graph-based signatures along with normal mode dynamics to generate a consensus prediction of the impact of a mutation on protein stability. We demonstrate our approach outperforms alternative approaches to predict the effects of mutations on protein stability and flexibility (P-value < 0.001), achieving a correlation of up to 0.70 on blind tests. DynaMut also provides a comprehensive suite for protein motion and flexibility analysis and visualization via a freely available, user-friendly web server at <http://biosig.unimelb.edu.au/dynamut/>.

4.4 HOMOLOGY MODELING AND STRUCTURAL ANALYSIS

To analyze the quality of predicted models and their possible 3D structure, Homology modeling was performed. Since, the crystal structures of selected proteins except FASN protein were not present in the PDB (Protein data bank) therefore, homology modeling approach with SWISS-MODEL was carried out to predict their possible 3d Structures (Biasini et al., 2014). Stereo chemical properties and Ramachandran plot analysis were carried out by the PROCHECK server for the structural validation of predicted models. Other structural properties like RMSD, free energy, energy minimization etc. were carried out and visualized under the Swiss-PDB viewer, PyMOL. VADAR (volume, area dihedral angle reporter) server was used to calculate H-bond energy (Willard et al., 2003).

4.5 COMPUTATION OF ACCESSIBLE SURFACE AREA (ASA)

VADAR is a freely available web server which can be used for quantitative appraisal of the quality of protein structures (Willard et al., 2003). To analyze the structural parameters for both individual residue and the entire protein, it uses PDB formatted files or PDB accession numbers as an input. ASA was computed and compared for the native and mutant model with the help of this server.

4.6 PROTEIN –LIGAND DOCKING

Protein –ligand docking of the proteins was performed using an online tool, dock Thor. Dock Thor is a web server that offers a virtual screening platform for drug design and repurposing. Its primary function is to predict protein-ligand interactions and calculate binding affinities. To achieve this, Dock Thor utilizes a scoring function that takes into account intermolecular interaction energy, intramolecular interaction energy, and the torsional term of the ligand. During the docking process, the generated docking poses are clustered using DTStatistics, an in-house tool. The top energy-poses from each cluster are then selected as representatives and made accessible to users. The affinity prediction and ranking of different ligands are carried out using a linear model that is not tailored specifically for the Dock Score scoring function (Barboza JJ et.al.,2023) Dock Score is a collection of empirical scoring functions developed by the research group. It incorporates essential terms for protein-ligand binding, including intermolecular interaction terms, torsional entropy term, protein-ligand lipophilic interaction term, and polar solvation term. The scoring function was trained and tested using a diverse set of high-quality 3D structures associated with varying molecular weights, number of rotatable bonds, and relevant therapeutic targets for drug design. Dock Thor is capable of predicting the binding affinity of small macrocyclic drugs, which are provided to users in their original conformation. Additionally, a separate dataset containing an ensemble of conformations of small macrocyclic drugs from e-Drug 3D is also available to users. Furthermore, a separate dataset specifically for e-Drug3D macrocycle drugs with a backbone size smaller than 20 heavy atoms is provided. In terms of charge assignment, algorithms are typically employed to distribute the net charge of a molecule among its constituent atoms. Most docking methods assume that the protonation state and charge distribution in molecules remain constant. However, it is crucial to evaluate free torsions, protonation states, and charge assignments for successful docking. The protonation states of the target's amino acid residues can be determined through various methods

CHAPTER 5

RESULTS AND DISCUSSION

The Panther tool is based on sequence homology and the Panther-PSEP method. Scores are based on hidden Markov model families. The effect of the nsSNP is given in terms of probably damaging, possibly damaging and probably benign (> 450 my; 200my to 450my; time < 200my respectively). Polyphen2 tool performs functional annotation of SNPs, maps coding SNPs to gene transcripts. Based on these properties, it gives Position specific independent counts (PSIC) scores to each substitution the values of which are PSIC < 0.05 – Benign; PSIC >0.05 – Probably Damaging & Possibly Damaging. The output of MutPred2 includes a general score (g), which represents the probability that an amino acid substitution is pathogenic. This score is derived as the average of the scores generated by all neural networks within MutPred2. Interpreted as a probability, a score threshold of 0.50 indicates the likelihood of pathogenicity. SIFT tool is based on sequence homology, scores assessment based on position – specific scoring matrices with Dirichlet priors. Score ranges from 0 to 1, where score < =0.05 is damaging and score > 0.05 is tolerated.

The nsSNPs that could be deleterious were evaluated based on the similarity of prediction by at least 3 tools of the four used. Based on the prediction, a total number of 34 nsSNPs for FASN were identified as deleterious .

TABLE: 5.1: IDENTIFICATION OF FUNCTIONALLY SIGNIFICANT SNPS

RS-IDS	MUTATION	SIFT SCORE	PANTHER PROBABILITY	POLYPHEN SCORE	MUTPRED SCORE	SNP & GO	SNP & GO Score
rs148217171	V2066M	0.006	probably damaging	1	0.523	Disease	0
rs369778189	R2026C	0.03	probably damaging	0.091	0.744	Disease	2
rs374709802	Y1924D	0	probably damaging	0.865	0.792	Disease	0
rs144918967	V1913M	0.002	probably damaging	1	0.366	Disease	0
rs142766883	R1907H	0.003	probably damaging	0.999	0.422	Disease	0
rs376251096	A1798V	0.011	probably damaging	1	0.175	Disease	0
rs373722639	G1781S	0	probably damaging	1	0.623	Disease	6
rs112806006	P1713L	0.002	probably	0.992	0.264	Disease	0

			damaging			se	
rs369077167	R1574H	0.003	probably damaging	0.665	0.785	Disea se	4
rs148696950	R995C	0.003	probably damaging	1	0.724	Disea se	6
rs374021258	E945D	0.013	probably damaging	1	0.556	Disea se	5
rs371010865	G882D	0.015	probably damaging	0.971	0.863	Disea se	1
rs1140612	H764D	0.005	probably damaging	1	0.804	Disea se	7
rs150531044	A756T	0	probably damaging	0.999	0.409	Disea se	0
rs199640764	A641P	0.007	probably damaging	0.951	0.954	Disea se	6
rs200141857	R633C	0.009	possibly damaging	1	0.516	Disea se	0
rs149454216	G592S	0.007	possibly damaging	1	0.844	Disea se	6
rs376743967	A590T	0.013	probably damaging	1	0.637	Disea se	2
rs143978140	G571E	0.035	probably damaging	1	0.805	Disea se	2
rs148219546	I561T	0.004	probably damaging	1	0.723	Disea se	5
rs368896751	R522C	0.007	probably damaging	0.995	0.587	Disea se	1
rs368756460	R468H	0	Probably damaging	0.999	0.813	Disea se	5
rs140748143	S275L	0.036	probably damaging	1	0.212	Disea se	6
rs1140610	A244T	0.003	probably damaging	1	0.698	Disea se	6
rs144193556	D216N	0	probably damaging	0.999	0.701	Disea se	8
rs373243965	G210S	0	probably damaging	0.17	0.821	Disea se	5
rs150605488	P194L	0	probably damaging	0.937	0.853	Disea se	8
rs370854158	D158E	0.004	probably damaging	0.997	0.748	Disea se	9
rs200618103	A141T	0.01	probably damaging	1	0.826	Disea se	3
rs372842189	G93R	0.003	probably damaging	1	0.804	Disea se	5
rs199926630	G46R	0.004	probably damaging	0.889	0.884	Disea se	6
rs201676557	G8V	0	probably damaging	1	0.924	Disea se	6

5.2 EFFECTS OF NSSNPS ON STABILITY OF PROTEIN

The identified deleterious nsSNPs were subjected to stability analysis using I-Mutant2.0. I Mutant determines the effect of given nsSNPs in terms of increasing or decreasing the protein stability. based on the unfolding free energy change (DDG) and the reliability index which can be computed only when the sign of the stability change is predicted and is evaluated from the output of the SVM O as $RI = 20 * \text{abs}(O - 0.5)$. Based on the output, 17 nsSNPs of FASN were found to decrease the stability. The effects of the selected mutations on protein stability are listed in table 5.2.1

5.2 .1: Effects of selected nsSNPs of FASN using I-MUTANT 2.0

I-MUTANT 2.0

RS -IDS	SUBSTITUTION	STABILITY	DDG	RELIABILITY INDEX
rs1140612	H764D	DECREASE	-1.09	3
rs150531044	A756T	DECREASE	-1.36	7
rs199640764	A641P	DECREASE	-1.61	4
rs200141857	R633C	DECREASE	-0.13	8
rs149454216	G592S	DECREASE	-1.31	9
rs376743967	A590T	DECREASE	-1.04	9
rs143978140	G571E	DECREASE	-0.64	3
rs148219546	I561T	DECREASE	-3.58	9
rs368896751	R522C	DECREASE	-1.42	9
rs368756460	R468H	DECREASE	-1.55	7
rs1140610	A244T	DECREASE	-2.98	9
rs144193556	D216N	DECREASE	-1.27	2
rs373243965	G210S	DECREASE	-0.88	7
rs370854158	D158E	DECREASE	-0.55	6

rs200618103	A141T	DECREASE	-1.78	9
rs372842189	G93R	DECREASE	-1.79	6
rs199926630	G46R	DECREASE	-0.28	4
rs201676557	G8V	DECREASE	-0.05	1

5.3 EFFECTS OF NSSNPS ON STABILITY OF PROTEIN USING DYNAMUT

The results obtained from the DynaMut web server showcase its efficacy in analyzing protein dynamics and evaluating the impact of mutations on protein stability and dynamics. DynaMut employs two distinct normal mode approaches, integrating graph-based signatures with normal mode dynamics, to yield a consensus prediction of mutation effects on protein stability. Notably, DynaMut has demonstrated superior performance compared to alternative methods, achieving a correlation of up to 0.70 on blind tests. For a single mutation, DynaMut provides results across three panels: $\Delta\Delta G$ Predictions, Interatomic Interactions, and Deformation and Fluctuation Analysis. These results are presented in a concise table format, detailing mutation specifics such as residue type, wild-type residue, mutant residue, $\Delta\Delta G$, and prediction confidence. Additionally, the server offers a comprehensive suite for protein motion and flexibility analysis and visualization. Validation studies underscore the reliability of DynaMut, revealing a Pearson's correlation of $r = 0.67$ and $RMSE = 1.31$ kcal/mol under 10-fold cross-validation. This performance surpasses that of individual methods used in consensus prediction. Furthermore, in a non-redundant blind test scenario, DynaMut achieved a Pearson's correlation coefficient of 0.70 ($RMSE = 1.45$) for both forward and hypothetical reverse mutations, affirming its robustness in predicting mutation effects on protein stability.

5.3.1 : Effects of selected nsSNPs of FASN on protein stability using Dynamut

MUTATION	$\Delta S_{vib}(\text{delta_vibrational_entropy})$	$\Delta\Delta G(\text{ddg_prediction})$	$\Delta\Delta G_{ENCOM}(\text{delta_stability_encom})$	$\Delta\Delta G_{mcsm}$	$\Delta\Delta G_{sdm}$	$\Delta\Delta G_{duet}$
----------	--	--	--	-------------------------	------------------------	-------------------------

A590T	-0.31	0.8	0.248	-1.079	-2.54	-1.214
P491S	0.002	0.012	-0.002	-2.253	-0.55	-2.104
R490H	0.146	-1.094	-0.117	-1.594	-0.89	-1.816
V527M	-0.486	-0.633	0.389	-0.396	-0.61	-0.306
I561T	0.258	-2.646	-0.206	-2.843	-3.46	-3.169
R522C	0.289	-0.845	-0.231	-1.129	-0.44	-1.05
A756T	-0.188	-1.349	0.15	-2.096	-2.88	-2.375
R633C	0.711	-0.453	-0.569	-0.785	-0.66	-0.792
A641P	-0.455	-0.971	0.364	-0.518	-4.13	-1.126
G571E	-0.113	-0.844	0.091	-0.784	-2.42	-1.067
R468H	-0.74	-0.559	0.592	-1.479	-0.56	-1.529
A691T	-0.133	-0.898	0.106	-1.35	-2	-1.395
G592S	-0.076	-1.669	0.061	-0.717	-3.73	-1.217
L771M	-0.16	0.813	0.128	-0.969	-0.99	-1.028
H614Y	0.063	-0.202	-0.05	0.868	-0.11	0.922
P817S	-0.09	0.672	0.072	-0.524	-0.55	-0.276
H764D	0.457	0.017	-0.366	-0.928	-1.2	-0.985

5.4 PAIRWISE STRUCTURAL ANALYSIS OF NATIVE AND MUTANT MODELS

Homology modeling approach was used for the prediction of the possible 3D structure for the native and mutated models of FASN, by using SWISS – MODEL individually (Figure 5.4.1). By using the PDB viewer, RMSD values ,TM scores were calculated. The mutant model having high RMSD value shows high variation in structure RMSD values of native and mutant models were provided in the following table 5.4.1

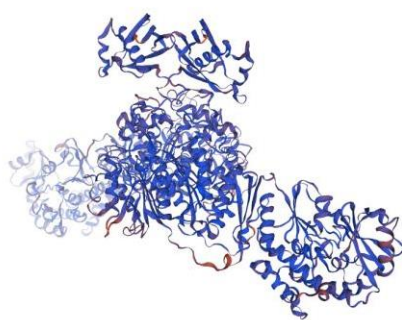
5.4.1 : Pairwise structure alignment Value results

MUTATION	Chain	RMSD	TM-score	Identity	Equivalent Residues	Sequence Length	Modelled Residues
A590T	A	1.42	0.87	100%	851	961	961
P491S	A	1.42	0.87	100%	851	961	961
R490H	A	1.42	0.87	100%	851	961	961
V527M	A	1.42	0.87	100%	851	961	961
I561T	A	1.42	0.87	100%	851	961	961
R522C	A	1.42	0.87	100%	851	961	961
A756T	A	1.42	0.87	100%	851	961	961
R633C	A	1.42	0.87	100%	851	961	961
A641P	A	1.42	0.87	100%	851	961	961
G571E	A	1.42	0.87	100%	851	961	961
R468H	A	1.42	0.87	100%	851	961	961
A691T	A	1.42	0.87	100%	851	961	961
G592S	A	1.42	0.87	100%	851	961	961
L771M	A	1.42	0.87	100%	851	961	961
H614Y	A	1.42	0.87	100%	851	961	961
P817S	A	1.42	0.87	100%	851	961	961
H764D	A	1.42	0.87	100%	851	961	961

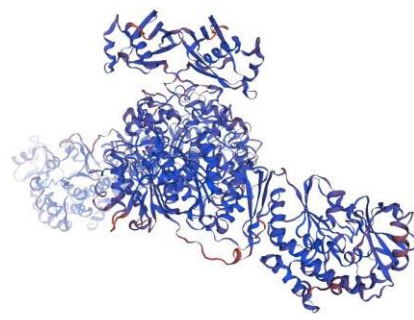
5.5 HOMOLOGY MODELING OF NATIVE AND MUTANT MODELS

Homology modeling approach was used for the prediction of the possible 3D structure for the native and mutated models of FASN, by using SWISS – MODEL individually (Figure 5.5.1) .Our analysis revealed significant structural alterations

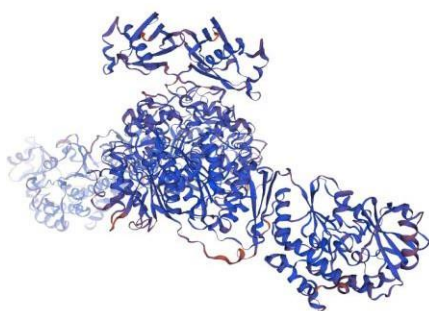
in the mutant models compared to their native counterparts. These alterations were observed in key regions implicated in protein stability, ligand binding, and enzymatic activity. By elucidating these structural changes, our study provides valuable insights into the molecular mechanisms underlying the phenotypic effects associated with the studied mutations. Furthermore, the comparison between native and mutant models highlighted specific residues and structural motifs crucial for maintaining protein integrity and functionality.



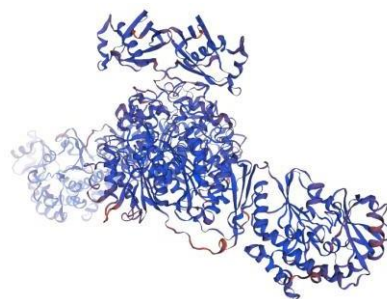
A) A590T



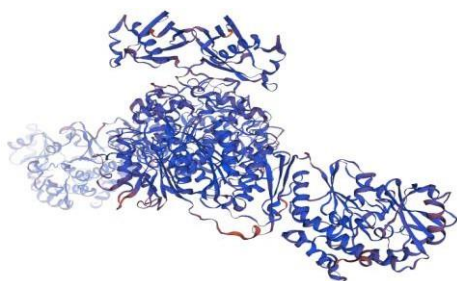
B) P491S



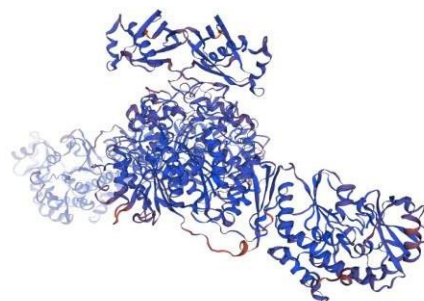
C) R490H



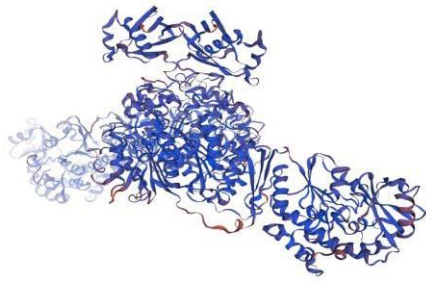
D) V527M



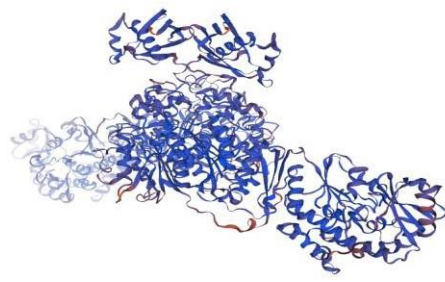
E) I561T



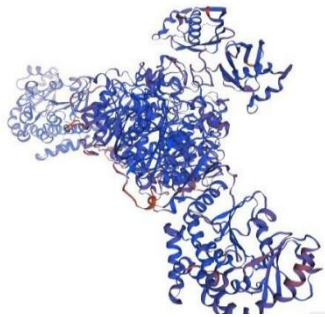
F) R522C



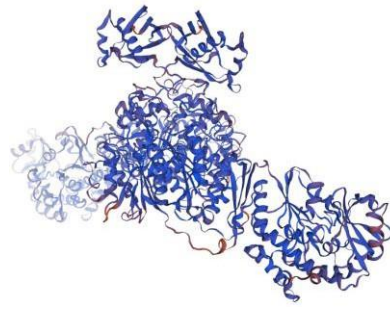
G) A756T



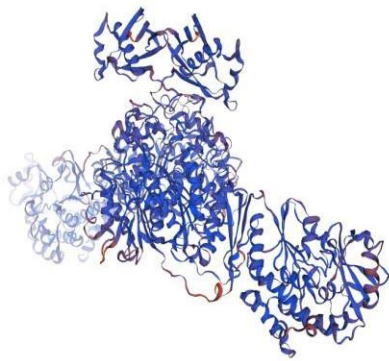
H) R633C



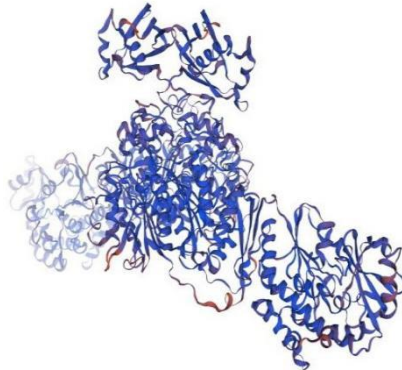
I) A641P



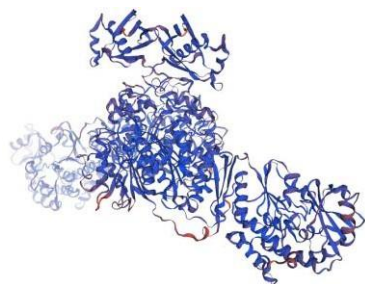
J) G571E



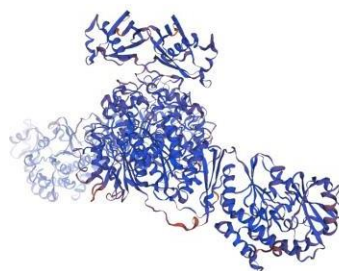
K) R468H



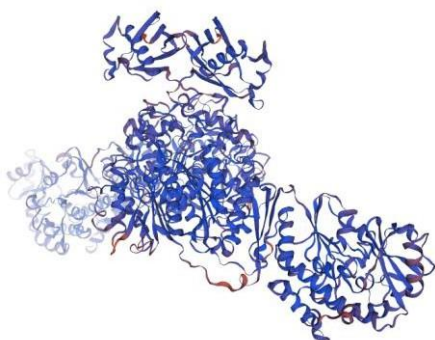
L) A691T



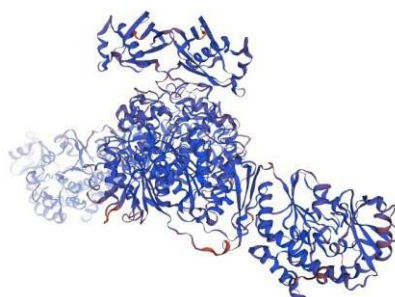
M) G592S



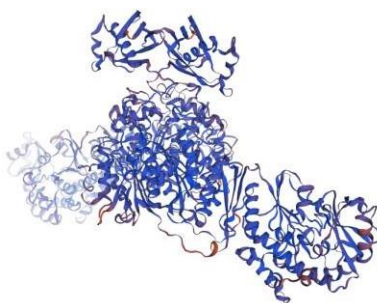
N) L771M



O) H614Y



P) H764D



Q) P817S

Figure 5.5.1: Homology-modeled 3D structures of FASN

5.6 VALIDATION OF NATIVE AND MUTANT MODEL STRUCTURES USING PROCHECK

The PROCHECK algorithm was employed to assess the structural validity and stereochemical properties of the predicted protein models. This algorithm examines the geometry of the protein on a residue-by-residue basis and predicts the stereochemical quality of its structure. The Ramachandran plot, a graphical representation of the protein backbone conformations, was utilized to evaluate the protein's conformational characteristics. Our investigation revealed that the Ramachandran plot statistics indicated that the majority of the predicted structural models, with the exception of the native model, had more than 85% of their residues located in the most favored regions (A, B, L). Additionally, the PROCHECK tool measures the uniqueness of a protein's structural properties using the G factor. The G factor serves as an indicator of how "normal" or "unusual" a particular stereochemical property is. To determine the G factor value, an analysis was conducted on 163 non-homologous, high-resolution protein chains obtained from X-ray crystallography resolved structures, with a resolution of 2.0Å or better, and an R-factor not exceeding 20%. The optimal G factor value for an unusual structure is considered to be less than -0.5, while for a highly unusual structure, it should be less than -1. However, in our study, the G factor value was found to be greater than -0.5. Based on the fact that more than 85% of the residues in the predicted models were located in the favored regions, the modeled structures were deemed to be of good quality. Therefore, the results obtained from the PROCHECK analysis indicate that the predicted models possess typical stereochemical and structural properties, as presented in table 5.4.

Table 5.6.1 Validation of the Structures using PROCHECK

Protein	Model	Ramachandran Plot Statistics				G factor Overall average
		Most favored region (A,B,L)	Additional allowed regions [a, b, l, p]	Generously allowed regions [a, b, l, p]	Disallowed regions (xx)	
FASN	Native	92.60%	8.95%	0.90%	0.90%	0.09
	A590T	90.60%	8.40%	0.60%	0.40%	-0.18
	<u>P491S</u>	90.10%	8.70%	0.70%	0.40%	-0.18
	<u>R490H</u>	91.00%	8.10%	0.40%	0.50%	-0.18
	<u>V527M</u>	90.40%	8.60%	0.60%	0.40%	-0.18
	<u>I561T</u>	90.80%	8.30%	0.40%	0.50%	-0.18
	<u>R522C</u>	89.60%	9.30%	0.60%	0.40%	-0.18
	<u>A756T</u>	90.40%	8.50%	0.60%	0.50%	-0.19
	<u>R633C</u>	90.50%	8.40%	0.60%	0.40%	-0.17
	<u>A641P</u>	90.60%	8.40%	0.60%	0.40%	-0.18
	G571E	90.10%	8.80%	0.70%	0.40%	-0.18
	R468H	90.20%	8.0%	0.60%	0.40%	-0.18
	<u>A691T</u>	90.50%	8.50%	0.50%	0.50%	-0.18
	G592S	90.50%	8.40%	0.60%	0.50%	-0.17
	L771M	90.6%	8.50%	0.40%	0.50%	-0.19
	H614Y	90.70%	8.30%	0.40%	0.50%	-0.18
	P817S	89.80%	9.20%	0.60%	0.40%	-0.18
	H764D	90.40%	8.50%	0.60%	0.50%	-0.19

5.7 ANALYSIS OF ACCESSIBLE SURFACE AREA (ASA) AND SOLVENT ACCESSIBILITY

The VADAR server, utilized for ASA prediction, assesses the solvent accessibility of amino acid residues within a specific protein structure. Amino acid residues are categorized into buried, partially buried, and exposed groups based on their solvent accessibility, representing low, moderate, and high accessibility levels, respectively (Raih et al., 2005). This investigation focused solely on the analysis and comparison of surface-exposed solvent accessible residues and the percentage hydrophobicity of both native and mutant protein models. Consequently, the structural disparity between the predicted native and mutant protein models was evident from this analysis. The ASA calculations for the proteins can be found in Table 5.5.

Table 5.7. :ASA computations of protein

Protein	Model	Total ASA	Exposed nonpolar ASA	Exposed polar ASA	Exposed charged ASA	% side ASA hydrophobic
	Native	68605.4 Å ²	41384.8 Å ²	13664.1 Å ²	13556.5 Å ²	31.6
FASN	A590T	68605.4 Å ²	41384.8 Å ²	13664.1 Å ²	13556.5 Å ²	31.6
	P491S	68215.4 Å ²	41564.4 Å ²	13764.2 Å ²	13759.5 Å ²	30.5
	R490H	68104.5 Å ²	41489.3 Å ²	13596.0 Å ²	13019.2 Å ²	32.02
	V527M	68770.9 Å ²	41337.5 Å ²	13686.4 Å ²	13747.0 Å ²	31.48
	I561T	68067.3 Å ²	41482.8 Å ²	13567.4 Å ²	13017.1 Å ²	32.0

	R522C	69090.0 Å ²	41156.0 Å ²	13889.7 Å ²	14044.4 Å ²	31.21
	A756T	68552.2 Å ²	41342.5 Å ²	13686.2 Å ²	13523.6 Å ²	31.64
	R633C	68569.0 Å ²	41402.4 Å ²	13743.0 Å ²	13423.6 Å ²	31.78
	A641P	68776.5 Å ²	41301.5 Å ²	13709.2 Å ²	13765.8 Å ²	31.47
	G571E	68875.9 Å ²	41251.9 Å ²	13681.4 Å ²	13942.5 Å ²	31.35
	R468H	67679.5 Å ²	9667.2 Å ²	3511.6 Å ²	3600.6 Å ²	37.71
	A691T	68749.7 Å ²	41439.5 Å ²	13709.9 Å ²	13600.3 Å ²	31.53
	G592S	69178.3 Å ²	41186.8 Å ²	13809.2 Å ²	14182.4 Å ²	30.99
	L771M	68154.1 Å ²	41472.4 Å ²	13571.8 Å ²	13109.8 Å ²	31.93
	H614Y	68094.9 Å ²	41565.5 Å ²	13654.4 Å ²	12874.9 Å ²	31.99
	P817S	69178.3 Å ²	41186.8 Å ²	13809.2 Å ²	14182.4 Å ²	30.99
	H764D	68588.6 Å ²	41382.4 Å ²	13693.8 Å ²	13512.4 Å ²	31.63

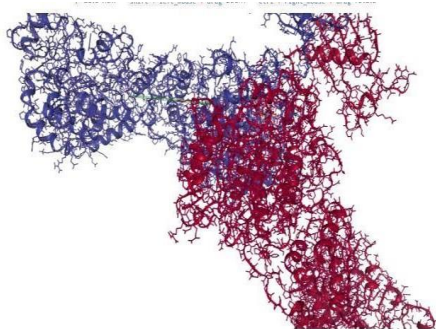
5.8 DOCKTHOR Analysis:

The utilization of DockThor in analyzing the FASN native and mutant models has provided valuable insights into the molecular interactions governing ligand binding. Through our investigation, we have elucidated crucial aspects of ligand-receptor affinity, van der Waals energy, and Electrostatic Ligand-Chemical (ELC) energy, shedding light on the mechanisms underlying ligand recognition and binding specificity. Our findings underscore the significance of structural variations induced by mutations in FASN, revealing their impact on ligand recognition and binding energetics. The comparison between native and mutant models has elucidated distinct patterns in affinity profiles, reflecting altered binding dynamics influenced by structural perturbations. Moreover, the analysis of van der Waals energy has unveiled intricate interplay between ligand geometry and receptor conformation, highlighting the role of non-covalent interactions in stabilizing ligand-receptor complexes.

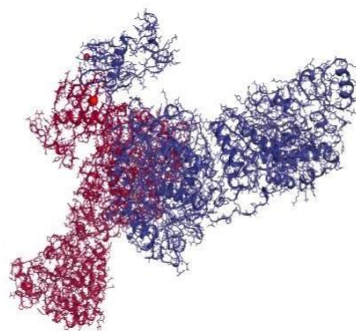
Furthermore, the assessment of ELC energy has provided valuable insights into the electrostatic contributions to ligand binding, uncovering the significance of charge distribution and polarization in modulating binding affinity. The differential ELC energy profiles observed between native and mutant models underscore the impact of structural alterations on electrostatic interactions, thereby influencing ligand recognition and binding specificity. Our study underscores the importance of integrating computational approaches like DockThor with experimental data to unravel the complexities of ligand-receptor interactions. By dissecting the molecular determinants of binding affinity, van der Waals energy, and ELC energy, we have advanced our understanding of FASN-mediated ligand recognition, offering potential avenues for the design of novel therapeutics targeting this crucial enzyme.

Table 5.8.1 : Dockthor RESULTS

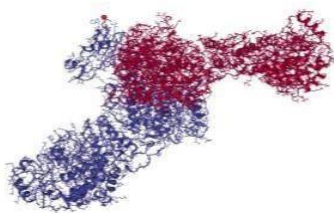
PROTEIN	SUBSTITUTION	Affinity	Total Energy	vdW Energy	Elec. Energy
FASN	NATIVE	-5.754	-158.58	3.433	-31.252
	A590T	-4.542	-37.629	-0.001	-0.057
	P491S	-4.409	-115.35	-0.001	-0.02
	R490H	-9.982	-47.421	39.088	-13.951
	V527M	-9.571	-64.883	50.522	-37.672
	I561T	8.499	-54.152	13.472	-7.082
	R522C	-9.658	-45.282	38.584	-28.86
	A756T	-9.86	-64.252	25.806	-29.773
	R633C	-9.713	-43.978	41.908	-19.28
	A641P	-9.627	-50.847	37.355	-30.201
	G571E	-8.136	-74.992	10.745	-30.854
	R468H	-9.471	-46.883	49.522	-27.572
	A691T	-9.236	-55.992	23.745	-23.854
	G592S	-4.409	-115.35	-0.001	-0.02
	L771M	-9.982	-47.421	39.088	-13.951
	H614Y	-9.571	-64.883	50.522	-37.672
	P817S	8.499	-54.152	13.472	-7.082
	H764D	-10.136	-74.992	10.745	-30.854



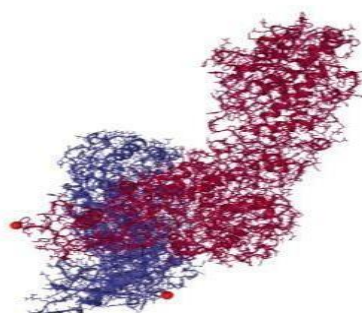
A)R633C



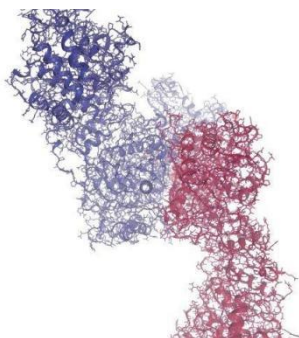
B)A756T



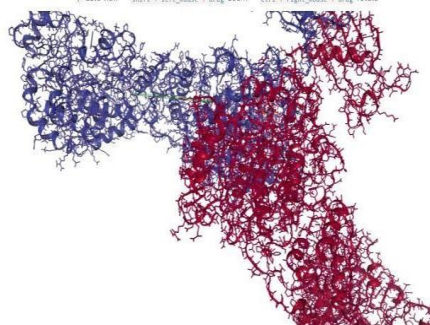
C)R633C



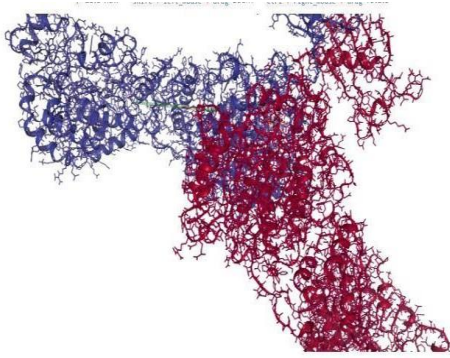
D)A641P



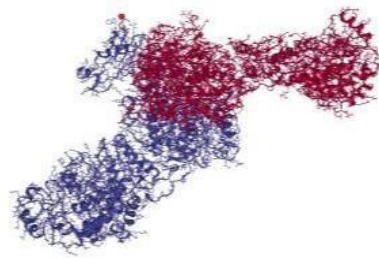
E)P817S



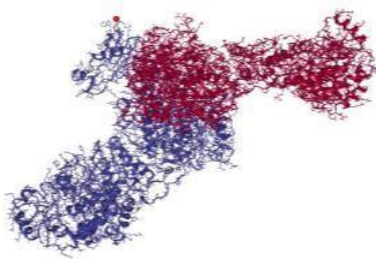
F)V527M



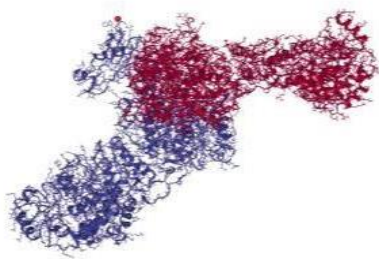
G) L771M



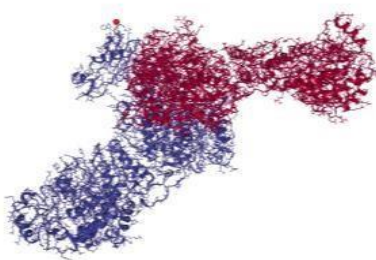
H) G592S



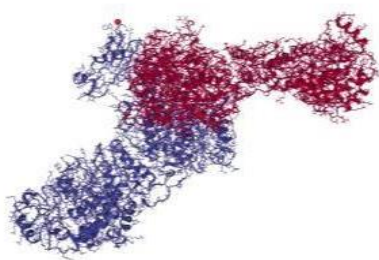
I) A641P



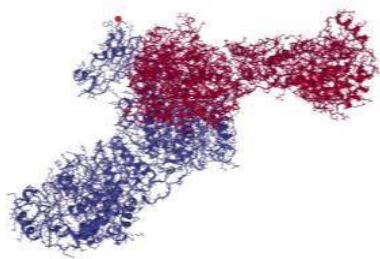
J) G571E



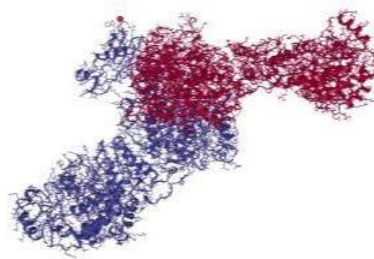
K) R468H



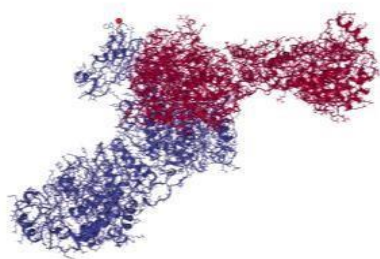
L) A691T



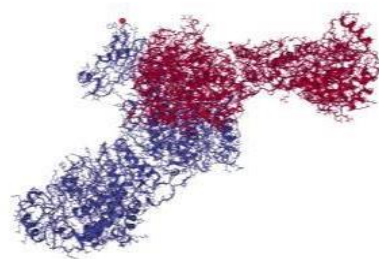
M) G5925



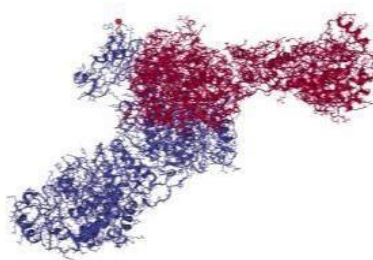
N) L771M



O) H614Y



P) H764D

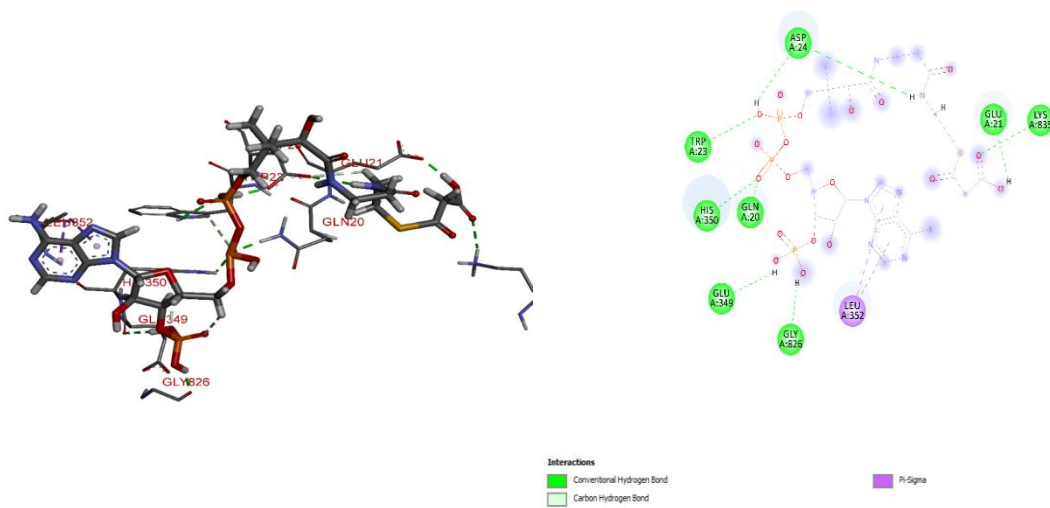


Q) P8175

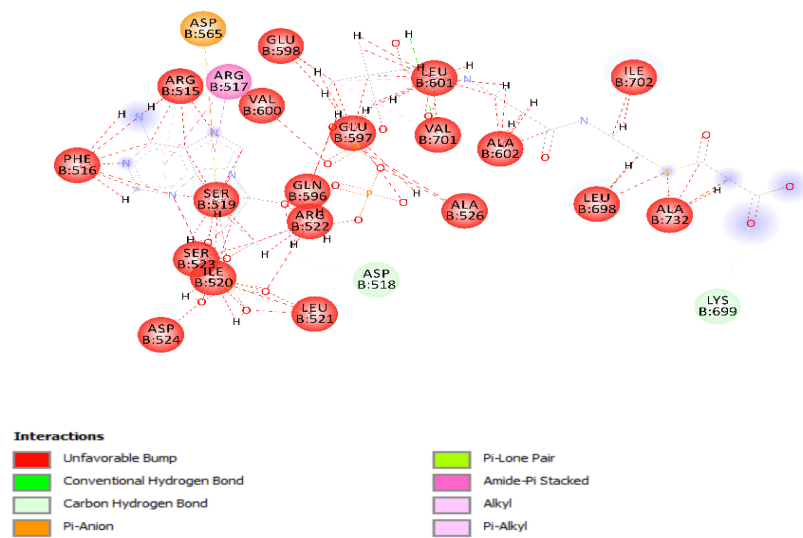
Figure 5.8.2 : FASN Native & Mutant 3D structures using DOCKTHOR

5.9 INTERACTIONS:

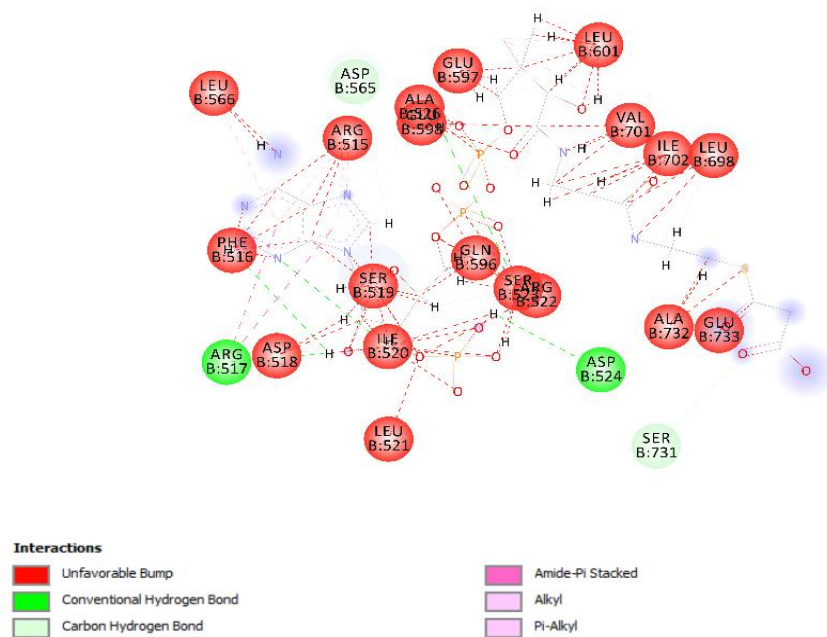
Upon obtaining the docking outcomes from the dockthor server, visualize the ligand interactions using discovery studio. In instances where the amino acids differ during these interactions, it is imperative to acknowledge that such mutations will exacerbate the impact of FASN, consequently resulting in the formation of cancer cells.



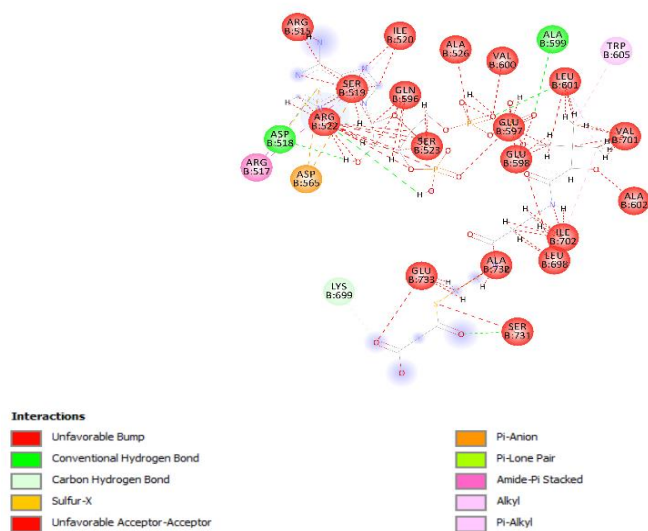
A) NATIVE MODEL-3HHD



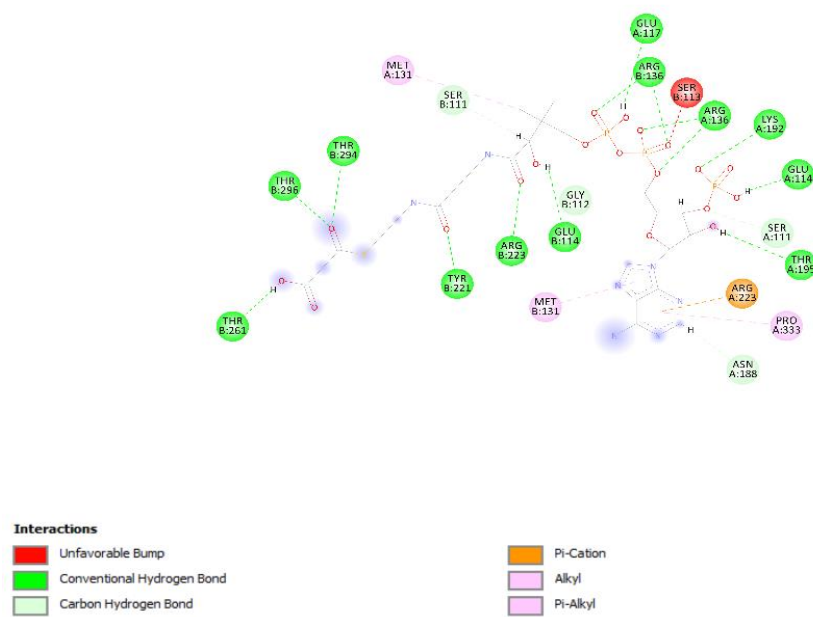
B) MUTANT -R490H



C) MUTANT -A641P



D) MUTANT - R633c



E) MUTANT-L771m

Figure .5.9.1: INTERACTIONS RESULTS

CHAPTER 6

SUMMARY & CONCLUSION:

The current investigation presents findings on three non-synonymous single nucleotide polymorphisms (nsSNPs) in the FASN gene, namely, L771M, R633C, R490H, and A641P. These nsSNPs have been identified as potentially harmful based on the assessment of various computational tools. Consequently, conducting in vivo and in vitro studies on these three nsSNPs may provide valuable insights into the underlying disease mechanisms. Notably, the nsSNPs L771M, R633C, R490H, and A641P are located in the ligand binding site of the FASN protein, suggesting their potential as future drug targets for cancer treatment. Therefore, the computational analysis conducted in this study proves to be both cost-effective and easily manageable in evaluating and monitoring the predicted deleterious nsSNPs in the FASN gene. These findings hold promise for advancing drug discovery and development, potentially leading to targeted therapies for various types of cancer.

CHAPTER 7

REFERENCE

1. Abdul Samad, F., Suliman, B. A., Basha, S. H., Manivasagam, T., & Essa, M. M. (2016). A comprehensive in silico analysis on the structural and functional impact of SNPs in the congenital heart defects associated with NKX2-5 gene—A molecular dynamic simulation approach. *PLoS One*, 11(5), e0153999. DOI:[10.1371/journal.pone.0153999](https://doi.org/10.1371/journal.pone.0153999)
2. Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov, A. S., & Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nat Methods*, 7(4), 248-249. DOI:[10.1038/nmeth0410-248]
3. Akhtar M, Jamal T, Din JU, Hayat C, Rauf M, Ul Haq SM, Sher Khan R, Shah AA, Jamal M, Jalil F. (2019). An in silico approach to characterize nonsynonymous SNPs and regulatory SNPs in human TOX3 gene. *J Genet*, 98, 104. DOI:[10.1007/s12041-019-01216-1]
4. Arifuzzaman, M., Mitra, S., Das, R., Hamza, A., Absar, N., & Dash, R. (2020). In silico analysis of nonsynonymous single-nucleotide polymorphisms (nsSNPs) of the SMPX gene. *Annals of human genetics*, 84(1), 54-71. DOI: [10.1111/ahg.12326](https://doi.org/10.1111/ahg.12326)
5. Arshad, M., Bhatti, A., & John, P. (2018). Identification and in silico analysis of functional SNPs of human TAGAP protein: A comprehensive study. *PloSone*, 13(1), e0188143. DOI:[10.1371/journal.pone.0188143](https://doi.org/10.1371/journal.pone.0188143)

6. Bhagwat, M. (2010). Searching NCBI's dbSNP database. *Curr Protoc Bioinformatics*,32(Unit:1.19)
DOI:[10.1002/0471250953.bi0119s32](https://doi.org/10.1002/0471250953.bi0119s32)
7. Capriotti, E., Calabrese, R., Fariselli, P., Martelli, P. L., Altman, R. B., & Casadio, R. (2013). WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC Genomics*,14(Suppl3),S6.DOI:[10.1186/1471-2164-14-S3-S6]
8. Dabhi, B., & Mistry, K. N. (2014). In silico analysis of single nucleotide polymorphism (SNP) in human TNF- α gene. *Meta gene*, 2, 586-595. DOI: [10.1016/j.mgene.2014.08.003]
9. Dristy TT, Noor AR, Dey P, Saha A.(2023) .Structural analysis and conformational dynamics of SOCS1 gene mutations involved in diffuse large B-cell lymphoma. *Gene*.2023 May 15;864:147293.
doi: 10.1016/j.gene.2023.147293. .
10. Gao S, Feng X, Wu Z, Kajigaya S, Young NS.(2023). CellCallEXT: Analysis of Ligand-Receptor and Transcription Factor Activities in Cell-Cell Communication of Tumor Immune Microenvironment. *Cancers (Basel)*. 2022 Oct 10;14(19):4957. doi: 10.3390/cancers14194957.
11. Guedes IA, Costa LSC, Dos Santos KB, Karl ALM, Rocha GK, Teixeira IM, Galheigo MM, Medeiros V, Krempser E, Custódio FL, Barbosa HJC, Nicolás MF, Dardenne LE.(2021). Drug design and repurposing with DockThor-VS web server focusing on SARS-CoV-2 therapeutic targets and their non-synonym variants. *Sci Rep*. 2021 Mar 10;11(1):5543.
doi: 10.1038/s41598-021-84700-0.

12. Hossain, M. S., Roy, A. S., & Islam, M. S. (2020). In silico analysis predicting effects of deleterious SNPs of human RASSF5 gene on its structure and functions. *Scientific reports*, 10(1), 14542. DOI:[10.1038/s41598-020-71544-9]
13. Kalia, N., Sharma, A., Kaur, M., Kamboj, S. S., & Singh, J. (2016). A comprehensive in silico analysis of non-synonymous and regulatory SNPs of human MBL2 gene. *Springerplus*, 5, 1-14. DOI: [10.1186/s40064-015-1649-9](https://doi.org/10.1186/s40064-015-1649-9)
14. Kaur, R., Singh, J., & Kaur, M. (2017). Structural and functional impact of SNPs in P-selectin gene: A comprehensive in silico analysis. *Open life sciences*, 12(1), 19-33. DOI: [10.1515/biol-2017-0003](https://doi.org/10.1515/biol-2017-0003)
15. Kuhajda, F. P. (2000). Fatty acid synthase and human cancer: new perspectives on its role in tumor biology. *Nutrition*, 16(3), 202–208. DOI: [10.1016/S0899-9007(99)00266-1](https://doi.org/10.1016/S0899-9007(99)00266-1)
16. Kuhajda, F. P. (2006). Fatty acid synthase and cancer: new application of an old pathway. *Cancer Res*, 66(12), 5977–5980. DOI:[10.1158/0008-5472.CAN-05-4673]
17. Marín-Martín FR, Soler-Rivas C, Martín-Hernández R, Rodríguez-Casado A.(2014) A Comprehensive In Silico Analysis of the Functional and Structural Impact of Nonsynonymous SNPs in the ABCA1 Transporter Gene. *Cholesterol*. 2014;6:234. DOI: 10.1155/2014/639751.
18. Menendez, J. A., & Lupu, R. (2007). Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. *Nat. Rev. Cancer*, 7(10), 763–777. DOI: [10.1038/nrc2222](https://doi.org/10.1038/nrc2222)

19. Menendez, J. A., & Lupu, R. (2004). Fatty acid synthase-catalyzed de novo fatty acid biosynthesis: from anabolic-energy-storage pathway in normal tissues to jack-of-all trades in cancer cells. *Arch. Immunol. Ther. Exp.(Warsz.)*,52(6),414-426.
DOI:[10.1023/B:ARCH.0000040890.47755.6e]
20. Menendez JA, Lupu R.(2006). Oncogenic properties of the endogenous fatty acid metabolism: molecular pathology of fatty acid synthase in cancer cells. *Curr Opin Clin Nutr Metab Care*. 2006 Jul;9(4):346-57.
doi: (10.1097/01.mco.0000232893.21050.15.)
21. Messori, A., De Rosa, M., Fadda, V., & Pani, L. (2015). Effectiveness and cost effectiveness of bevacizumab in metastatic colorectal cancer. *J. Clin. Oncol.*,33,3838–3839.DOI:[10.1200/JCO.2015.62.8310]
22. Rajasekaran, R., Doss, C. G. P., Sudandiradoss, C., Ramanathan, K., & Sethumadhavan, R. (2008). In silico analysis of structural and functional consequences in p16INK4A by deleterious nsSNPs associated CDKN2A gene in malignant melanoma. *Biochimie*, 90(10), 1523-1529. DOI: [10.1016/j.biochi.2008.06.005](https://doi.org/10.1016/j.biochi.2008.06.005)
23. Rajasekaran, R., Sudandiradoss, C., Doss, C. G. P., & Sethumadhavan, R. (2007). Identification and in silico analysis of functional SNPs of the BRCA1gene.*Genomics*,90(4),447-452.DOI:[10.1016/j.ygeno.2007.06.005]
24. Rotimi, S. O., Peter, O., Oguntade, O., & Rotimi, O. A. (2018). In silico analysis of the functional non-synonymous single nucleotide polymorphisms in the human CYP27B1 gene. *Egyptian Journal of Medical HumanGenetics*,19(4),367-378.DOI:[10.1016/j.ejmhg.2018.03.003]

25. Vasuki KA, Christy HJ. Comprehensive Study of Human FBXW7 Deleterious nsSNP's Functional Inference and Susceptibility to Gynaecological Cancer. *Appl Biochem Biotechnol*. 2022 Jan;194(1):407-433. doi: (10.1007/s12010-021-03759-7)
26. Wang, H., Xi, Q., & Wu, G. (2016). Fatty acid synthase regulates invasion and metastasis of colorectal cancer via Wnt signaling pathway. *Cancer Med*,5(7),1599-1606.DOI: [10.1002/cam4.672]