

***IN SILICO IDENTIFICATION AND ANALYSIS OF CANCER  
CAUSING SMAD4 AND ERBB2 GENE VARIANTS***

**PROJECT THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE  
REQUIREMENT OF BACHELOR OF BIOTECHNOLOGY**



**By**

**Meghana Reddy Dropathi– 1601-18-805016**

**Sai Vamsi Erra – 1601-18-805040**

**Under the guidance of**

**Dr. Sanjeeb Kumar Mandal, Asst. Professor**

**DEPARTMENT OF BIOTECHNOLOGY  
CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY (A)  
HYDERABAD, TELANGANA**

**May 2022**



## CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY (A)

Kokapet ( Village), Gandipet, Hyderabad, Telangana-500075. [www.cbit.ac.in](http://www.cbit.ac.in)

Recognized  
Research Centers

Programs Accredited by  
**NBA**  
ASSOCIATION  
ACREDITATION  
COMMISSION

Approved by  
**UGC**

Approved by  
**NAAC**  
A - Grade

All India 132nd Rank in  
**NIIF**  
NATIONAL  
INSTITUTIONAL  
FRAMEWORK

ISO Certified  
9001:2015

COMMITTED TO  
RESEARCH,  
INNOVATION AND  
EDUCATION

**43**  
years

### CERTIFICATE

This is to certify that the project entitled "***In silico identification and analysis of cancer causing smad4 and erbB2 gene variants***" submitted by **Meghana Reddy Dropathi (1601-18-805-016)**, and, **Sai Vamsi Erra(1601-18-805-040)**, in partial fulfilment for the degree of "Bachelor of Technology" in Biotechnology, Chaitanya Bharathi Institute of Technology (CBIT), Hyderabad affiliated to Osmania University is a bonafide record of work carried under the supervision of Dr. Sanjeeb Kumar Mandal, Asst. Professor, Department of Biotechnology, CBIT and the same has not been submitted to any other university or institute for the award of degree or diploma.

### Internal Guide

**Name:** Dr. Sanjeeb Kumar Mandal  
**Signature:**

**Date:**

### Internal Examiner – 1

**Name:** Dr. C. Nagendranatha Reddy  
**Signature:**

### Internal Examiner – 2

**Name:** Dr. Rajasri Yadavalli  
**Signature:**

**Date:**

**Date:**

### External Examiner

**Name:**  
**Signature:**

### HOD

**Name:** Dr. Rajasri Yadavalli  
**Signature:**

**Date:**

**Date:**

## **DECLARATION**

This is to certify that the project entitled "***In silico identification and analysis of cancer causing SMAD4 and erbB2 gene variants***" submitted by **Meghana Reddy Dropathi(1601-18-805-016)**, and **Sai Vamsi Erra (1601-18-805-040)** in partial fulfilment for the degree of "Bachelor of Technology" in Biotechnology, Chaitanya Bharathi Institute of Technology (CBIT), Hyderabad affiliated to Osmania University is a bonafide record of work carried out under the supervision of Dr. Sanjeeb Kumar Mandal, Asst. Professor, Department of Biotechnology, CBIT and the same has not been submitted to any other university or institute for the award of degree or diploma.

Student Name : Meghana Reddy Dropathi  
Roll Number : 1601-18-805-016

Signature :

Student Name : Sai Vamsi Erra  
Roll Number : 1601-18-805-040

Signature :

Date of Submission : 14.05.2022

## **ACKNOWLEDGMENTS**

We, hereby thank the following distinguished personalities who stretched their helping hand in the project.

Firstly, I would like to thank Dr. Sanjeeb Kumar Mandal, Assistant Professor, Department of Biotechnology, CBIT, and Dr. B. Sumithra, Assistant Professor, Department of Biotechnology, CBIT for a keen interest in providing us with all the necessary support required for the successful submission of the thesis. We are grateful to the Principal of CBIT, Dr. P. Ravinder Reddy, and all the Directors as well as the Director (Academics) for motivating us to carry out our project work in CBIT and also providing Infrastructure and a variety of additional resources that are required for my research.

I would also like to acknowledge the constant support of Dr. C. Nagendranatha Reddy, Project Coordinator, and Assistant Professor, Department of Biotechnology, CBIT for his unifying support in providing us with all the necessary guidelines for the project.

We are greatly indebted to Dr. Rajasri Yadavalli our HOD Project Coordinator and Assistant Professor, Department of Biotechnology, CBIT for her constant support and interest which drove me for doing this project.

We would also want to thank all of our teaching and non-teaching personnel at CBIT's Department of Biotechnology for their continuous support and contributions.

Finally, we owe a debt of gratitude to our family members, who have always been a continual source of inspiration and support.

## **CONTENTS**

<b>S. No.</b>	<b>Content</b>	<b>Page Number</b>
i	Abbreviations .....	I
ii	List of Figures .....	II
iii	List of Tables .....	III
	Abstract .....	1
	Keywords .....	1
	Graphical Abstract .....	2
1	Aims & Objectives .....	3
2	Significance Or Novelty of the Work.....	3
3	Introduction .....	4
4	Review of Literature .....	5
5	Tools, Softwares, databases, and methods.....	18
6	Results .....	21
7	Discussion .....	53
8	Conclusions .....	56
9	Future Prospects .....	57
10	References .....	58

## **ABBREVIATIONS**

CUPSAT	-	Cologne University Protein Stability Analysis Tool
ERBB2	-	Erb-B2 Receptor Tyrosine Kinase 2
HOPE	-	Have (y)Our Protein Explained
NCBI	-	National Centre for Biotechnology Information
PANTHER	-	Protein Analysis Through Evolutionary Relationships
PHYRE2	-	Protein Homology/AnalogY Recognition Engine
POLYPHEN 2	-	Polymorphism Phenotyping v2
PROVEAN	-	Protein Variation Effect Analyzer
PTM	-	Post Translational Modification
RMSD	-	Root Mean Square Deviation
SIFT	-	Sorting Intolerant from Tolerant
SMAD4	-	Mothers against decapentaplegic homolog 4
SNP	-	Single Nucleotide Polymorphism
TM SCORE	-	Template Modelling score

## LIST OF FIGURES

<b>Figure No</b>	<b>Title</b>	<b>Page No</b>
Figure 1	Graphical Abstract	2
Figure 2	Snapshot of dbSNP NCBI Home page	6
Figure 3	Snapshot of SIFT results page	7
Figure 4	Snapshot of Polyphen-2 submission page	7
Figure 5	Snapshot of PredictSNP home page	8
Figure 6	Snapshot of SNPs&GO homepage	9
Figure 7	Snapshot of PROVEAN homepage	10
Figure 8	Snapshot of PANTHER homepage	11
Figure 9	Snapshot of INTERPRO home page	12
Figure 10	Snapshot of ConSurf homepage	13
Figure 11	Snapshot of Phyre2 homepage	14
Figure 12	Snapshot of Swiss-Model homepage	15
Figure 13	Snapshot of HOPE homepage	16
Figure 14	CONSURF results	22
Figure 15	Docked wild SMAD4 with ALPHA L-FUCOSE	40
Figure 16	Docked W509R mutant SMAD4 with ALPHA L-FUCOSE	41
Figure 17	Docked I383K mutant SMAD4 with ALPHA L-FUCOSE	41
Figure 18	Docked R361C mutant SMAD4 with ALPHA L-FUCOSE	42
Figure 19	Docked Y353S mutant SMAD4 with ALPHA L-FUCOSE	42
Figure 20	Docked C363R mutant SMAD4 with ALPHA L-FUCOSE	43
Figure 21	Docked G491V mutant SMAD4 with ALPHA L-FUCOSE	43
Figure 22	Docked L533P mutant SMAD4 with ALPHA L-FUCOSE	44
Figure 23	Docked L533R mutant SMAD4 with ALPHA L-FUCOSE	44
Figure 24	Docked W524L mutant SMAD4 with ALPHA L-FUCOSE	45
Figure 25	Docked G386D mutant SMAD4 with ALPHA L-FUCOSE	45
Figure 26	CONSURF results	47
Figure 27	Docked wild erbB2 with X4Z	51
Figure 28	Docked D493H MUTANT erbB2 with X4Z	52
Figure 29	Docked E368K mutant erbB2 with X4Z	52

## LIST OF TABLES

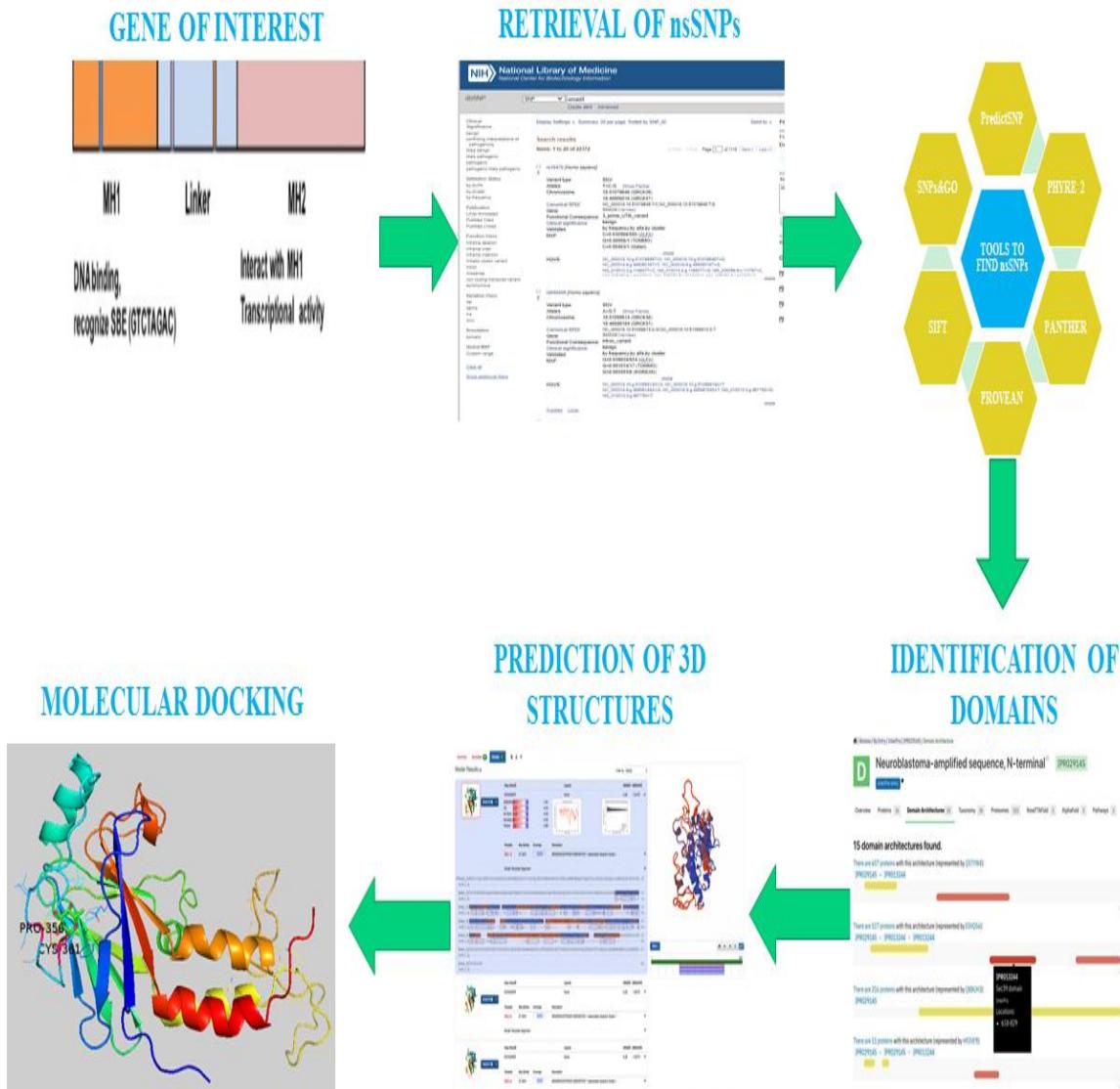
<b>Table No</b>	<b>Title</b>	<b>Page No</b>
<b>Table 1</b>	Identification of deleterious nsSNPs by six in silico programs	21
<b>Table 2</b>	Examination of evolutionary conservation of amino acids of SMAD4 by CONSURF	24
<b>Table 3</b>	TM score and RMSD predictions for nsSNPs in SMAD4	25
<b>Table 4</b>	Various parameters of 23 nsSNPs investigated by Swiss Model	26
<b>Table 5</b>	CUPSAT predictions on the effect of nsSNPs on protein stability	26
<b>Table 6</b>	Project Hope was used to investigate the structural effects of 23 nsSNPs on the SMAD4 protein.	27
<b>Table 7</b>	Missense3D tool was used to predict the structural influence of 23 nsSNPs in the SMAD4 protein.	36
<b>Table 8</b>	Docking with ALPHA L-FUCOSE	40
<b>Table 9</b>	Identification of deleterious nsSNPs by five in silico programs	46
<b>Table 10</b>	CUPSAT predictions on the effect of nsSNPs on protein stability	46
<b>Table 11</b>	Consurf investigated the evolutionary conservation of erbB2 amino acids.	49
<b>Table 12</b>	TM score and RMSD predictions for nsSNPs in erbB2	50
<b>Table 13</b>	Various parameters of 2 nsSNPs investigated by Swiss Model	50
<b>Table 14</b>	Structural effect of 5 nsSNPs over erbB2 protein using Project Hope	50
<b>Table 15</b>	Docking with X4Z	51

## **ABSTRACT**

Genetic variations in SMAD4 AND ERBB2 genes have been related to various cancers such as colorectal cancer. In this work, we have identified functional SNPs of both genes. Various computational tools were used to predict the deleterious nsSNPs which will impact the function and protein structure leading to the malfunction of the protein hence leading to cancer. Computational analysis was performed by using 6 tools that are SIFT, PROVEAN, PANTHER, POLYPHEN2, PredictSNP, and SNPs&GO. Six in silico SNP prediction techniques classified just 23 and 10 SNPs for the SMAD4 and ERBB2 genes, respectively, from all the SNPs that were found in the NCBI database. For the gene SMAD4, all nsSNPs were decreasing the stability whereas for the ERBB2 gene out of ten only eight were decreasing the stability. For the SMAD4 gene out of 23 nsSNPs, only 19 were highly conserved whereas, for the ERBB2 gene out of 10 nsSNPs, only 6 were highly conserved. The 3D structures of human and mutant SMAD4 and ERBB2 were generated by using the Swiss Model. Further analysis of nsSNPs of SMAD4 and ERBB2 was done to find which nsSNPs are stabilizing. It was found for SMAD4 only 10 out of 23 nsSNPs were stabilizing and for ERBB2, 2 out of 5 nsSNPs (that lie on the domain) were stabilizing. Furthermore, for the SMAD4 gene, only five were showing a certain change in structure according to MISSENSE 3D Tool. Post-translational modification sites have also been analyzed. Five nsSNPs of SMAD4 and 2 nsSNPs of ERBB2 have been found should be deserving of additional investigation in functional investigations to better understand their impact on metabolic phenotype occurrence.

**Keywords:** nsSNPs, SMAD 4, ERBB2, POLYPHEN 2, PANTHER, SIFT, PROVEAN, SNPs&GO, PredictSNP, MISSENSE 3D, SWISS MODEL

## SCHEMATIC REPRESENTATION OF OVERALL WORK



**Figure 1:** Graphical Abstract

## **1. AIMS AND OBJECTIVES**

To use in silico approaches to anticipate and analyse deleterious SNPs, as well as to assess their impact on the functionality and structure of the proteins with this in mind, the current study attempted to achieve the following goals:

1. Identification of deleterious nsSNPs
2. Determination of the stability of protein structures
3. Evolutionary conservation analysis
4. Structures of wild type and mutant models were investigated
5. Determination of protein structural stability
6. Molecular docking analysis

## **2. SIGNIFICANCE OR NOVELTY OF THE WORK**

The genes chosen for such studies have not been worked on and the studies used to evaluate SNPs, predictSNPs, etc. are novel bioinformatics tools. Most of nsSNPs have not been studied except for R361C (Aretz et al., 2007), G352E (Ray et al., 2022), and Y353S (Mafficini et al., 2022) in SMAD4.

### **3. INTRODUCTION**

In recent years, more genomes have been sequenced, leading to an exponential increase in knowledge about the variations that occur in the genome. Breakthroughs in sequencing and genotyping technologies, as well as economies of scale, have reduced their relative costs. Gene discovery for both monogenic and complicated illnesses has also increased. Bioinformatics databases and tools for collecting and analysing genetic data have grown in number, size, and breadth. Improved genetic data analysis is being performed to answer problems that cannot be solved by conventional methods. Nucleotide changes occur as a result of drug interactions, illness susceptibility, and cancer development. Nucleotide substitutions, insertions, duplications, and fusions are all possible nucleotide changes that can occur in a genome. A single-nucleotide substitution occurs when a nucleotide is substituted with one of the other three types of nucleotides. Single nucleotide mutations or SNPs are terms used to describe these types of alterations (Matsuda, 2017). In order for a base location in genomic DNA with sequence alternatives to be classed as an SNP, the least recurrent allele must have a frequency of 1% or more. The majority of SNPs are biallelic. SNPs are formed with a low frequency of single nucleotide substitutions and mutation bias are the explanations for this (Vignal et al., 2002) As a result, SNPs are divided into transversions and transitions. SNPs with transitions are more common than those with transversions. The genome has 10 million or more SNPs, with one every 100–300 base pairs (Srinivasan et al., 2016). In cancer research, SNP and haplotype analysis has a wide range of clinical and public health implications, as well as cancer research (Erichsen and Chanock, 2004).

## **4. REVIEW OF LITERATURE**

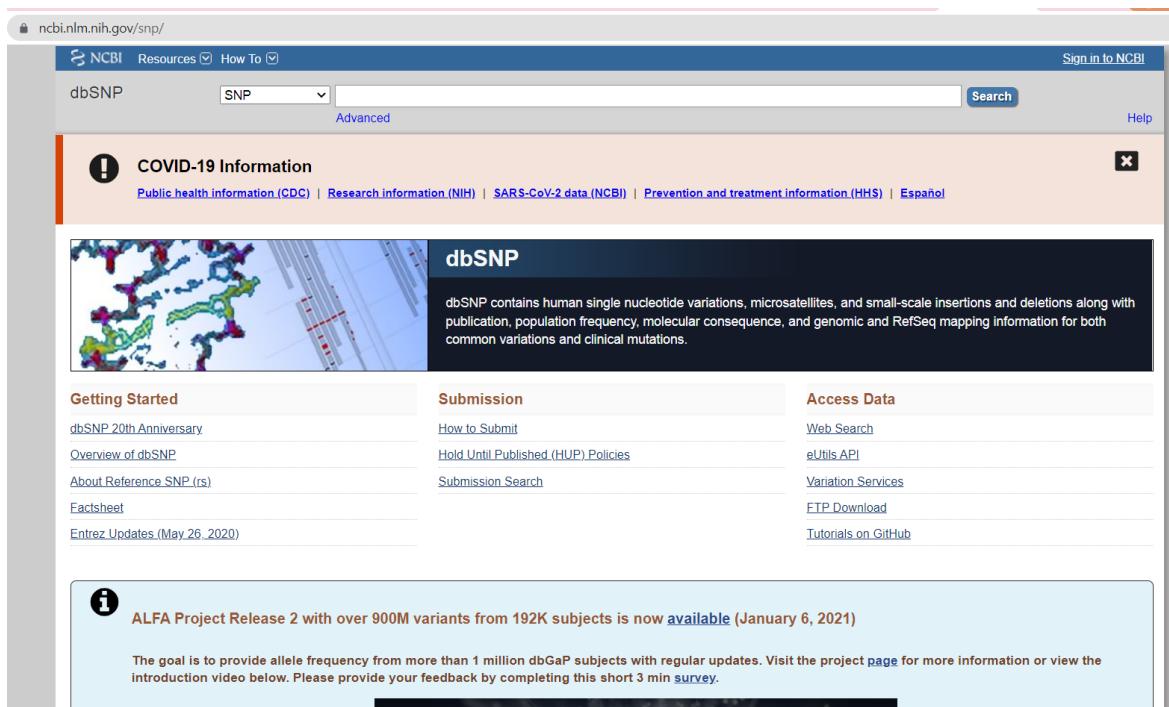
### **4.1 SNPs & THEIR IMPORTANCE**

Genetic diversity occurs both within and between populations, resulting in polymorphisms that can be connected to a hereditary trait being influenced by factors present in the environment. SNPs are variations in the genetic code. Although multiallelic SNPs occur, they mostly contain alternative bases and require 1% or more population frequency. SNPs are the most commonly occurring variations that occur in a genome we come across. They're commonly employed to look at genetic differences between people and groups (Vallejos-Vidal et al., 2020). SNPs in the coding (exons), intergenic, and noncoding (introns) regions of the genome could play a role in genome alterations (Ahmad et al., 2018). Because of their extensive occurrence, improved analysis, cheap genotyping costs, and we can also perform association studies using statistical and computational methodologies, The most important biomarkers for illness diagnosis and prognosis are SNPs (Srinivasan et al., 2016). As a result, SNPs have recently acquired recognition as critical drivers in disease-association studies.

### **4.2 TOOLS, SOFTWARES, AND DATABASES BEING USED**

#### **4.2.1 dbSNP DATABASE**

Individual phenotypic features, like a person's proclivity for complicated illnesses like heart disease and cancer, are caused by sequence changes at designated places within genomes. Sequence variants can be utilised for gene mapping, population structure definition, and functional research as aids for studying variation in different organisms. dbSNP is a publicly accessible database of basic genetic variations. This group of polymorphisms deletions or insertions, comparable element insertions, and repeat variants of bases in any number. Please keep in mind that any sort of variation can be used in place of SNP. Each dbSNP entry contains information on the polymorphism's sequence, the rate at which an amino acid occurs in the sequence, and the experimental method(s), procedures, and factors used to quantify the variance. dbSNP accepts variant inputs from any type of species and from any genome (Kitts and Sherry, 2011).



**Figure 2:** Snapshot of dbSNP NCBI Home page

#### 4.2.2 SIFT

SIFT is a protein function prediction technique that is frequently used to select nonsynonymous and missense variations. A protein may be tolerant of an amino acid change while still functioning normally, or it may be intolerant. SIFT determines if an amino acid change is tolerable or harmful to protein function. SIFT considers protein conservation and the severity of amino acid changes in homologous sequences. Many diseases, mutations, and genetic research have already employed it. The new protocol, which is presented below, enhances current features.

This technique explains how to quickly generate SIFT predictions for nonhuman species using a single graphics processing unit, as well as how to obtain predictions on a large number of protein sequences (GPU). SIFT 4G is a quicker form of SIFT that allows us to make SIFT predictions for huge groups of organisms (Vaser et al., 2016).

**SIFT results (dbSNP)**

Processing... If your browser times out before results are shown, html results can be seen at [https://sift.bii.a-star.edu.sg/www/sift/tmp/286e4f41bf\\_dbSNP.html](https://sift.bii.a-star.edu.sg/www/sift/tmp/286e4f41bf_dbSNP.html) and tsv results at [https://sift.bii.a-star.edu.sg/www/sift/tmp/286e4f41bf\\_dbSNP.tsv](https://sift.bii.a-star.edu.sg/www/sift/tmp/286e4f41bf_dbSNP.tsv). Both files are stored for 24 hours before being deleted.

Done.

SNP	ORGANISM/BUILD	CHR	COORDINATE	REF ALLELE	ALT ALLELE	AMINO ACID CHANGE	GENE NAME	GENE ID	TRANSCRIPT ID	PROTEIN ID	REGION	SIFT SCORE	SIFT MEDIAN	NO OF SEQS AT POSITION	SIFT PREDICTION
rs80338963	Homo_sapiens/GRCh37.74	18	48591918	C	A	R265S	SMAD4	ENSG00000141646	ENST00000588745	ENSP00000464901	CDS	0.001	2.51	218	DELETERIOUS
rs80338963	Homo_sapiens/GRCh37.74	18	48591918	C	A	R361S	SMAD4	ENSG00000141646	ENST00000342988	ENSP00000341551	CDS	0.027	2.46	141	DELETERIOUS
rs80338963	Homo_sapiens/GRCh37.74	18	48591918	C	A	R361G	SMAD4	ENSG00000141646	ENST00000398417	ENSP00000381452	CDS	0.027	2.46	141	DELETERIOUS
rs80338963	Homo_sapiens/GRCh37.74	18	48591918	C	G	R361G	SMAD4	ENSG00000141646	ENST00000342988	ENSP00000341551	CDS	0.001	2.46	141	DELETERIOUS
rs80338963	Homo_sapiens/GRCh37.74	18	48591918	C	G	R361G	SMAD4	ENSG00000141646	ENST00000398417	ENSP00000381452	CDS	0.001	2.46	141	DELETERIOUS
rs80338963	Homo_sapiens/GRCh37.74	18	48591918	C	G	R265G	SMAD4	ENSG00000141646	ENST00000588745	ENSP00000464901	CDS	0.021	2.51	218	DELETERIOUS
rs80338963	Homo_sapiens/GRCh37.74	18	48591918	C	T	R361C	SMAD4	ENSG00000141646	ENST00000342988	ENSP00000341551	CDS	0	2.46	141	DELETERIOUS
rs80338963	Homo_sapiens/GRCh37.74	18	48591918	C	T	R361C	SMAD4	ENSG00000141646	ENST00000398417	ENSP00000381452	CDS	0	2.46	141	DELETERIOUS
rs80338963	Homo_sapiens/GRCh37.74	18	48591918	C	T	R265C	SMAD4	ENSG00000141646	ENST00000588745	ENSP00000464901	CDS	0	2.51	218	DELETERIOUS
rs139569694	Homo_sapiens/GRCh37.74	18	48591943	A	G	N369S	SMAD4	ENSG00000141646	ENST00000342988	ENSP00000341551	CDS	0	2.46	141	DELETERIOUS
rs139569694	Homo_sapiens/GRCh37.74	18	48591943	A	G	N369S	SMAD4	ENSG00000141646	ENST00000398417	ENSP00000381452	CDS	0	2.46	141	DELETERIOUS
rs139569694	Homo_sapiens/GRCh37.74	18	48591943	A	G	N273S	SMAD4	ENSG00000141646	ENST00000588745	ENSP00000464901	CDS	0.02	2.51	218	DELETERIOUS
rs281875324	Homo_sapiens/GRCh37.74	18	48591826	A	G	E330G	SMAD4	ENSG00000141646	ENST00000342988	ENSP00000341551	CDS	0	2.45	142	DELETERIOUS
rs281875324	Homo_sapiens/GRCh37.74	18	48591826	A	G	E330G	SMAD4	ENSG00000141646	ENST00000398417	ENSP00000381452	CDS	0	2.45	142	DELETERIOUS
rs281875324	Homo_sapiens/GRCh37.74	18	48591826	A	G	E234G	SMAD4	ENSG00000141646	ENST00000588745	ENSP00000464901	CDS	0.001	2.51	222	DELETERIOUS
rs377767339	Homo_sapiens/GRCh37.74	18	48591807	T	C	C324R	SMAD4	ENSG00000141646	ENST00000342988	ENSP00000341551	CDS	0	2.45	143	DELETERIOUS
rs377767339	Homo_sapiens/GRCh37.74	18	48591807	T	C	C324R	SMAD4	ENSG00000141646	ENST00000398417	ENSP00000381452	CDS	0	2.45	143	DELETERIOUS
rs377767339	Homo_sapiens/GRCh37.74	18	48591807	T	C	C228R	SMAD4	ENSG00000141646	ENST00000588745	ENSP00000464901	CDS	0	2.51	221	DELETERIOUS
rs377767342	Homo_sapiens/GRCh37.74	18	48591825	G	A	E330K	SMAD4	ENSG00000141646	ENST00000342988	ENSP00000341551	CDS	0	2.45	142	DELETERIOUS
rs377767342	Homo_sapiens/GRCh37.74	18	48591825	G	A	E330K	SMAD4	ENSG00000141646	ENST00000398417	ENSP00000381452	CDS	0	2.45	142	DELETERIOUS

**Figure 3:** Snapshot of SIFT results page

#### 4.2.3 POLYPHEN 2

It is software which can also be accessed through a Web server. With the help of structural and evolutionary conservation considerations it anticipates the influence of amino acid alterations that affect the firmness and working of proteins. It develops conservation profiles by annotating SNPs for function, mapping coding SNPs to gene transcripts hence leading to annotations of amino acid sequences and structural determinations, and annotating SNPs. The likelihood of the missense mutation being deleterious is then calculated using a combination of these attributes. MultiZ genome alignments and UCSC Genome Browser human genome annotations are also included in the package. It can analyse huge amounts of data generated by next-generation sequencing programs (Adzhubei et al., 2013).

**Figure 4:** Snapshot of Polyphen-2 submission page

#### 4.2.4 PredictSNP

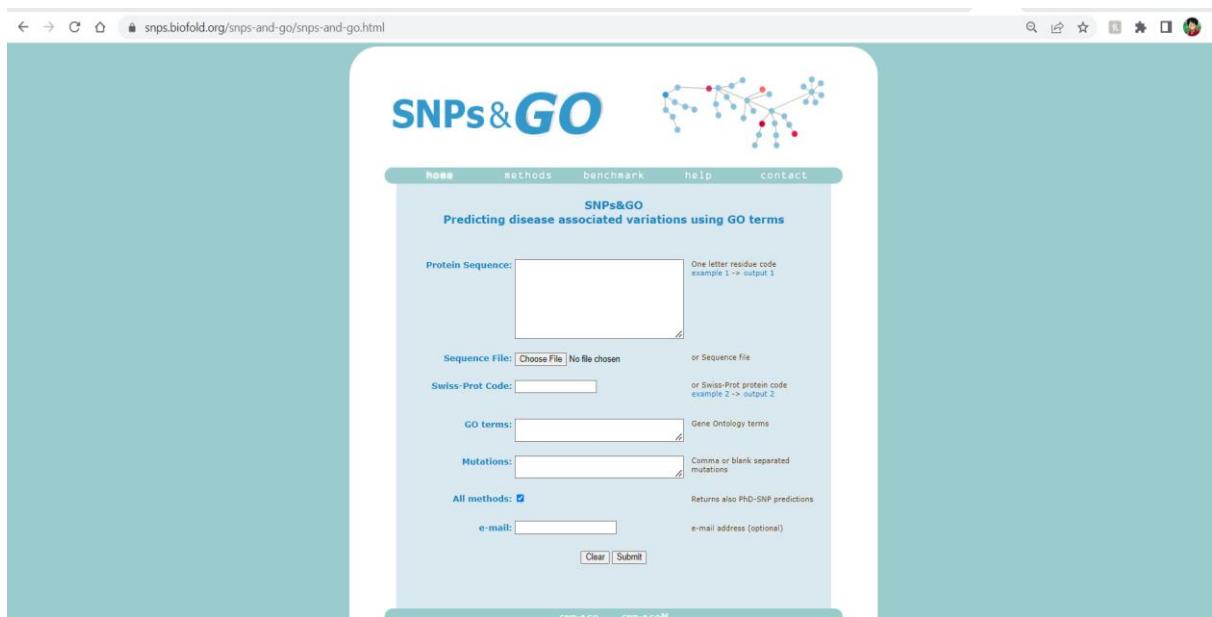
A benchmark dataset with endless mutations was used to conduct a neutral evaluation of eight proven prognosis tools like MAPP etc. The six best-performing aids came together to form PredictSNP, a concurrent classifier that significantly improved prediction accuracy while also returning findings for all variants, demonstrating that concurrent prediction provides a trustworthy and accurate substitute for individual tool prognosis. It can be easily operated by users and it is a web server that provides access to PredictSNP, and annotations from the databases like UniProt, etc (Bendl et al., 2014).



**Figure 5:** Snapshot of PredictSNP home page

#### 4.2.5 SNPs&GO

SNPs&GO is a strategy that uses the annotation of protein function to predict detrimental single amino acid polymorphisms (SAPs). The server uses Support Vector Machines (SVM), and the sequence and/or 3D structure of a protein, as well as a collection of target variants and functional Gene Ontology (GO) terms, are all inputs for a given protein. The server's output shows the likelihood of each protein variant being linked to human diseases (Capriotti et al., 2013).



**Figure 6:** Snapshot of SNPs&GO homepage

#### 4.2.6 PROVEAN

It is an established tool that scores proteins based on alignment (Choi et al., 2012). It is different from other programmes, it can predict changes in protein sequence, insertions, and deletions using the same algorithm it uses.

The PROVEAN online server has been expanded to allow for high-throughput analysis using the PROVEAN tool, whereas the initial web server (Choi et al., 2012) was only capable of single protein searches. The PROVEAN online service can now give predictions for big groups of nucleotide or amino acid changes in the genomes of both humans and mice. By conserving and reusing previously computed homologous known sequences information for query proteins, the current updated version of the single protein query function has greatly increased its run time. (Choi and Chan, 2015).

**Figure 7:** Snapshot of PROVEAN homepage

#### 4.2.7 PANTHER

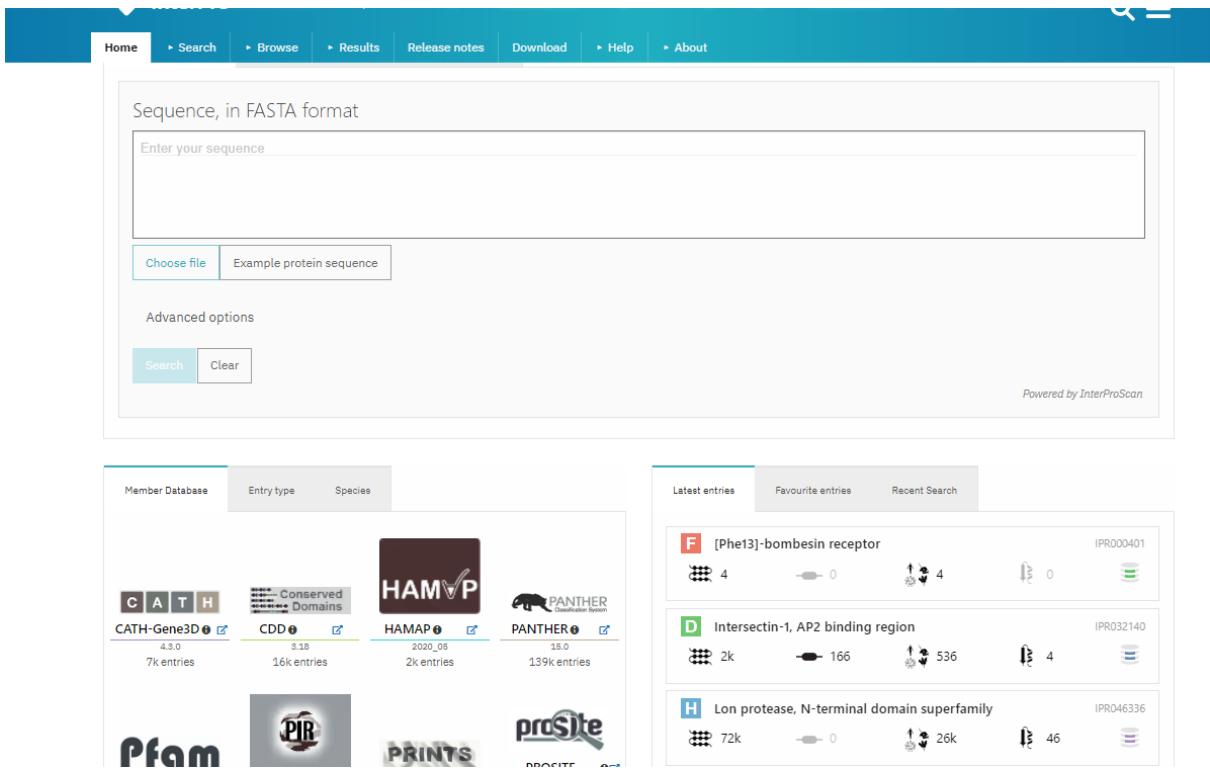
It is a database that gives us links between proteins. It provides us tools for using the classifications to analyse any size of genomes. Proteins are classified using evolutionary categories like its class, family etc and functional groupings. Evolutionary groups come under to the 'natural classification' of sequences based on their inheritance and evolution. Protein families are sets of similar proteins that may be matched to other protein by using different sequence alignment algorithms. Subfamilies are groups of proteins in different species that have evolved by events that occur among species through time via evolutionary tree created for a specific family. There may be multiple subfamilies in each family, and some of them are highly diverse. The identification of subfamilies by PANTHER is a critical component of the annotations' specificity. Proteins are classified according to their functions rather than their families in functional groupings. Functions

are classified in a variety of ways, including Gene's three different properties as well as pathways. Annotating functions can be done in two ways (Ebert et al., 2021).

**Figure 8:** Snapshot of PANTHER homepage

#### 4.2.8 INTERPRO

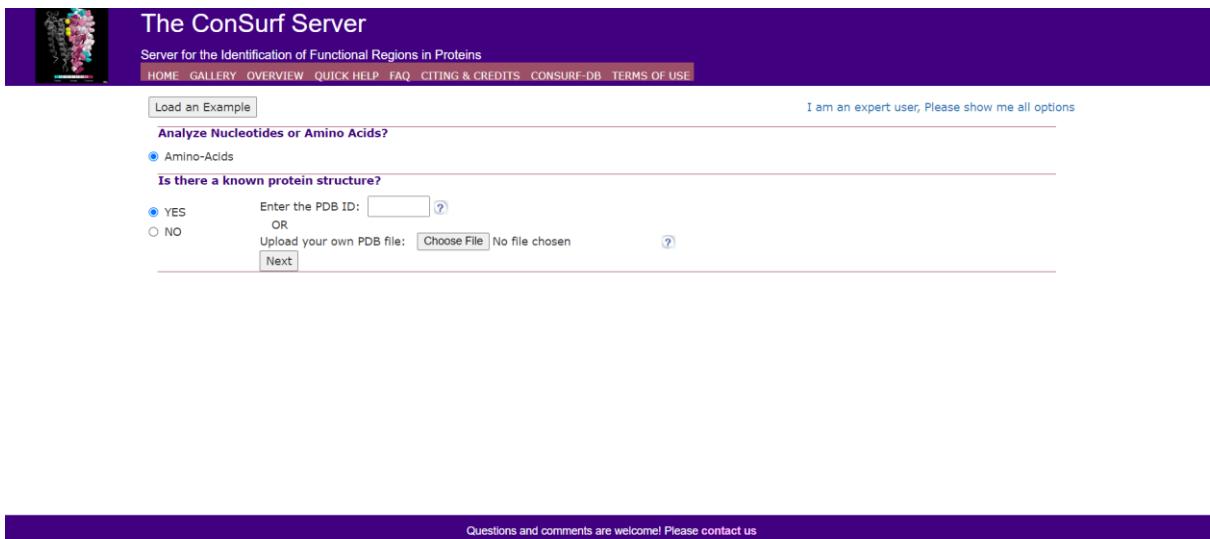
The InterPro is used by many other databases to aggregate prediction models that represent protein domains, families, etc from a variety of sources or databases. Nearly half of the 58,000 models in the source databases correspond to manual InterPro entry. On the website, they have started displaying the remaining un-integrated prediction models. Other modifications include the addition of unequalled UniProtKB proteins to the current match XML files and the inclusion of non-prediction modelled data. ADAN is a database that predicts protein-protein interactions, as well as the SPICE-like databases, have been added to the web interface (Hunter et al., 2009).



**Figure 9:** Snapshot of INTERPRO home page

#### 4.2.9 CONSURF

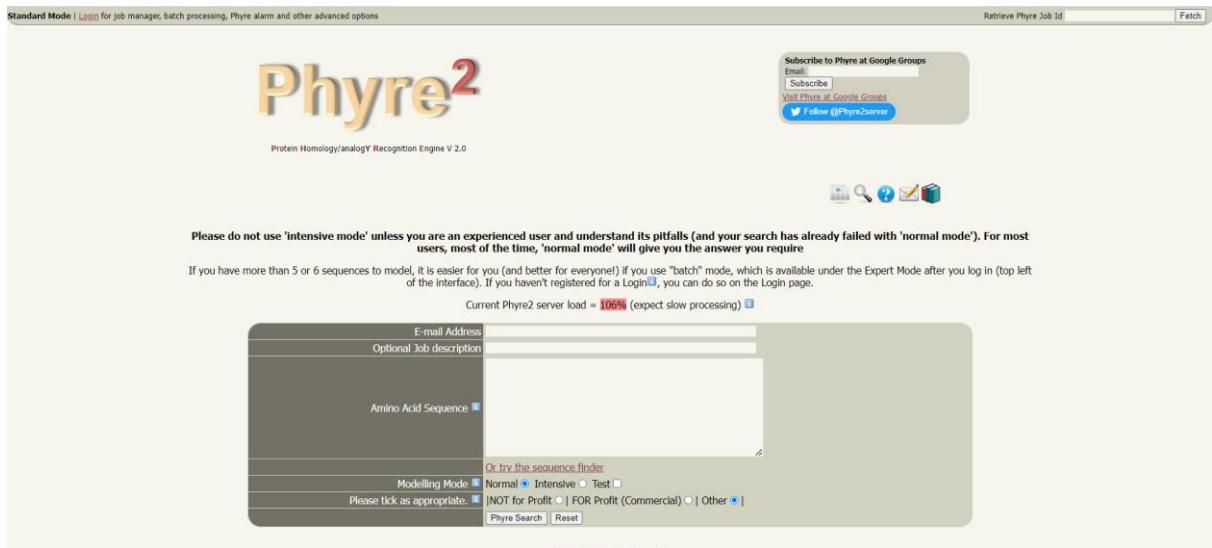
The evolution of protein sequences reveals the link between a protein's natural tendency to change and the general need to maintain structural integrity and function. The web server, which has been active for many years, analyses the amino/nucleic acid inheritance pattern to identify parts of the macromolecule that are important for analysis. The service finds homologues from a protein sequence or structure, infers their MSA, and uses all this information to construct a phylogenetic tree that shows their evolutionary. Using a probabilistic framework, these data are then used to estimate the evolutionary rates. ConSurf now includes the ability to model query proteins, predict the structure of query molecules from sequence, view the biological assembly of the input, map conservation grades onto 2D RNA models, and a more updated representation of the phylogenetic tree (Ashkenazy et al., 2016).



**Figure 10:** Snapshot of ConSurf homepage

#### 4.2.10 PHYRE2

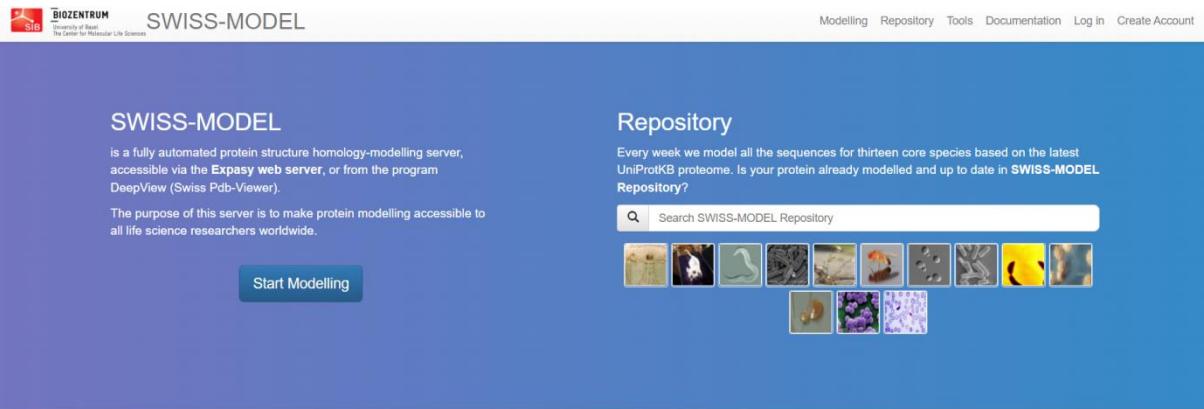
Phyre2 is an online suite of tools for predicting and analysing the function and structural changes of a protein caused by mutations. Phyre2 is a simple and intuitive user interface for cutting-edge protein bioinformatics tools for biologists. The initial server version, Phyre, has been replaced by Phyre2. We discuss Phyre2 in this upgraded protocol, which employs powerful remote detection methods to create 3D structures that anticipate which places ligands bind to and assess the impact of amino-acid changes on a user's protein sequence. A straightforward interface guides users through the results to the level of detail they desire. After inputting a protein sequence, this methodology will guide a user through the process of understanding their models' secondary and tertiary structure, domain composition, and model quality. It can also be used for locating a protein structure, submitting large quantities of sequences at once, and running searches for difficult-to-model proteins. It takes between 30 minutes and 2 hours to complete a typical structural forecast (Kelley et al., 2015).



**Figure 11:** Snapshot of Phyre2 homepage

#### 4.2.11 SWISS-MODEL

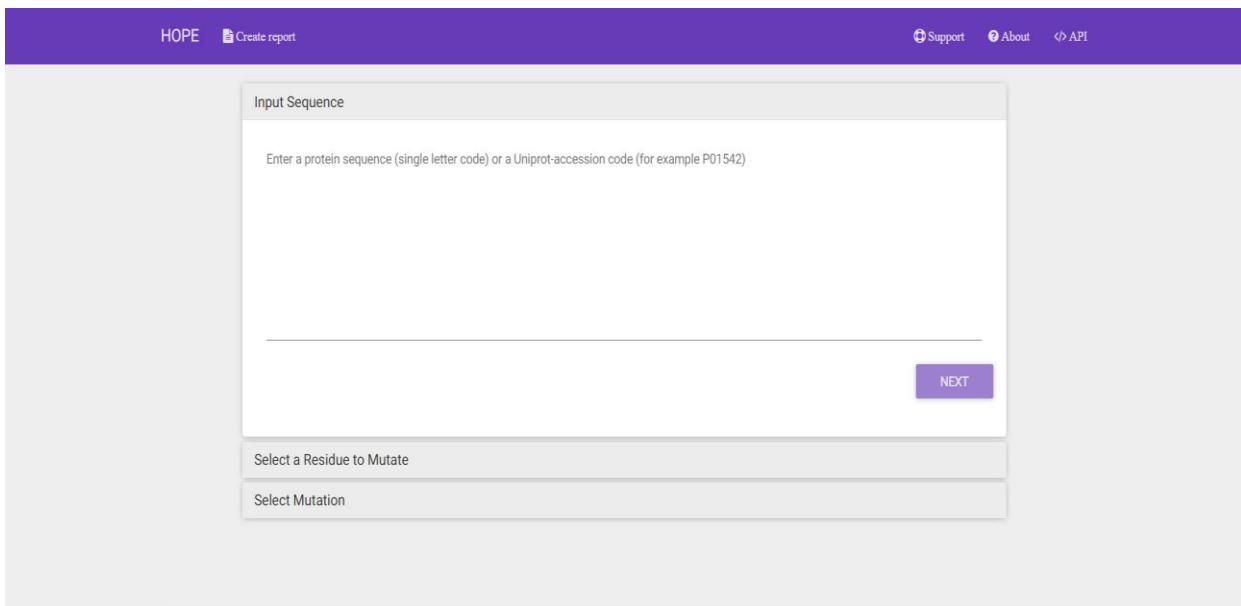
SWISS-MODEL is a service that predicts and provides three-dimensional (3D) protein structure comparative modelling. It is frequently used as it is a free web-based computer-based modelling facility today, having pioneered the field of automated modelling in 1993. The server processes many user queries for 3D Models every year. In its Wide Web interface allows for many degrees of user interaction: in the 'initial approach mode,' only a protein sequence can be given as input to generate a 3D model. The server handles templates, alignment, and prediction of models totally automatically. The modelling method in 'alignment mode' predicts structures according to the template given by the user. DeepView (Swiss-PdbViewer), an integrated sequence-to-structure workbench, can perform complicated modelling jobs. In the EVA-CM project, the reliability of SWISS-MODEL is constantly tested, and the results are provided to the users via email. Its server is constantly being upgraded in order to ensure that the professional information is successfully integrated into a user-friendly server (Schwede et al., 2003).



**Figure 12:** Snapshot of Swiss-model homepage

#### 4.2.12 Project HOPE

It is an online tool that analyses mutations automatically. HOPE was created to explain the molecular information of mutant residues compared to wild residues generated by human protein mutations. HOPE is similar to the aforementioned systems in this regard (PolyPhen, SIFT, ALAMUT). HOPE, in which information is collected utilising Web services and DAS servers. HOPE gathers data from a variety of sources that contain 3D structures and also uses UniProt database. This information is recorded in a PostgreSQL-based information system for each protein. These data are processed using a decision system, which studies the implications of the variations on the overall protein. This study is offered in a format tailored to the needs of human genetics researchers. We put HOPE to the test on a set of mutations that we had personally examined before. HOPE outperformed a skilled protein structure bioinformatician in every scenario that was uncomplicated (Venselaar et al., 2010).



**Figure 13:** Snapshot of HOPE homepage

#### 4.2.13 AUTODOCK VINA

It is exhibited AutoDock Vina, a revolutionary molecular docking and virtual screening software. According to our research on the trained data set used in the development of AutoDock 4, AutoDock Vina outperforms the molecular docking tool we previously developed (AutoDock 4) by a factor of two, while also improving the working of binding modes. On multi-core devices, parallelism is used to boost speed by leveraging multithreading (Trott and Olson, 2010).

#### 4.2.14 PyMOL

PyMOL is a molecular graphics application that can visualise proteins, nucleic acids, small molecules etc in 3D on any platform. It also has molecular editing and can also produce movies or videos of protein interaction. This is a software based on the coding language Python, was built to increase PyMOL's functionality and make drug creation easier. In order to gain a thorough understanding of PyMOL's useful tools and their functions. For visualisation and analysis improvements, drug screening, protein-ligand modelling, molecular simulations, there are unique technologies available (Yuan et al., 2017).

#### **4.2.15 CUPSAT**

It's an internet-based tool for analysing and assessing the stability of protein structures caused by single-position amino acid changes. This programme calculates the differences in the Gibb's free enthalpy of unfolding between mutant and wild-type proteins. It uses a Protein Data Bank-formatted protein structure as well as the position of the changed amino acid. The output contains about the mutation site, its structural characteristics, and detailed information about changes in protein stability for all different amino acid substitutions. It works on the principle of thermal denaturation (Parthiban et al., 2006).

## **5. TOOLS, SOFTWARES, DATABASES AND METHODS**

### **5.1 TOOLS& SOFTWARES**

1. SIFT
2. PANTHER
3. PROVEAN
4. POLYPHEN
5. PredictSNP
6. SNPs&GO
7. SWISS-MODEL
8. PHYRE 2
9. INTERPRO
10. I-MUTANT
11. CONSURF
12. CUPSAT
13. PROJECT HOPE
14. PYMOL
15. AUTODOCKVINA
16. NETPHOS 3.1
17. GPS 5.0
18. GPS-SUMO
19. GPS-MSP 1.0
20. BDM-PUB

### **5.2 DATABASES**

1. dbSNP DATABASE

### **5.3 METHODS**

#### **5.3.1 Retrieving nsSNPs**

NCBI dbSNP database (<https://www.ncbi.nlm.nih.gov/projects/SNP>) was used to acquire data on SNPs like rsid etc for both genes.

### **5.3.2 Identifying the most deleterious SNPs**

SIFT(<https://sift.bii.a-star.edu.sg/>),

PANTHER(<http://www.pantherdb.org/tools/csnpScoreForm.jsp>), PolyPhen2

(<http://genetics.bwh.harvard.edu/pph2/>),

SNPs&GO(<https://snps.biofold.org/snps-and-go/sn> ). This ensured the results' precision and consistency, and we classified those SNPs as deleterious because they were found to be harmful by all six systems.

### **5.3.3 Identification of nsSNPs on the domains of SMAD 4 &erbB2**

We used the InterPro tool (<https://www.ebi.ac.uk/interpro/>) to locate nsSNPs on the domains of SMAD4 and erbB2, which help to recognise motifs and domains of a protein and thus identify the functional defined domains of a protein by using various databases.

### **5.3.4 Analyzing protein evolutionary conservation**

The above was determined using ConSurf (<https://consurf.tau.ac.il>). It operates by comparing homologous sequences' phylogenetic similarities. We further analysed the nsSNPs.

### **5.3.5 Visualization of the effects of nsSNPs through 3D protein modelling**

We utilised three distinct online tools to predict the 3D models of the mutant protein: SWISS-MODEL (<https://swissmodel.expasy.org/>) and Phyre2 (<https://www.sbg.bio.ic.ac.uk/phyre2>). We used Phyre2 to predict the SMAD4 and erbB2 3D structure and function of SMAD4 and erbB2. The templates were chosen based on the scores provided by Phyre2. Swiss-Model was created using the SMAD4 templates d1khxa and erbB2 templates d1n8zc2andc5wnoA. The choice of our templates predicts the accuracy of our models. The proportion of sequence identity between query and template must be larger than 30–40% for a high accuracy model. PHYRE2 provided the template modelling-score (TM-score) and root mean square deviation. The TM-score gives you a score between 0 and 1. If it is 1 indicating perfect structure matching. Greater RMSD, on the other hand, indicates that the native and mutant structures are more different.

Finally, SWISS-MODEL was used to predict the three-dimensional structures of wild and mutant types.

### **5.3.6 Analyzing the effect of the nsSNPs on protein stability**

CUPSAT is used to stabilise the proteins (<http://cupsat.tu-bs.de/> ).

### **5.3.7 Prediction of nsSNP structural effects**

With the help of HOPE (<https://www3.cmbi.umcn.nl/hope>) we figured out how nsSNPs affect the overall protein. HOPE is a web-based programme that discovers mutations' structural consequences. Q13485 (SMAD4 UniProt-Accession Code) and B4DTR1 were used (UniProt-Accession Code of erbB2). Missense 3D tool was put into work (<https://www.sbg.bio.ic.ac.uk/missense3d/>) to guarantee that our results were accurate and stringent. The structural alterations caused by point mutations were predicted by Missense 3D.

### **5.3.8 Molecular Docking**

The COACH tool (<https://zhanggroup.org/COACH/>) was utilised to identify the ligands used for docking. We used the Autodock Vina software to perform molecular docking to see how detrimental point mutations affected the binding affinity of SMAD4 and erbB2. PYMOL was used to visualise the docked complexes.

### **5.3.9 Different post-translational modification sites prediction**

PTMs have an impact on a number of important biological processes, including cell signalling, metabolic pathways, and so on. The web-server ModPred (<https://www.modpred.org/>) is used to identify putative methylation locations in SMAD4 and erbB2 PROTEIN. NetPhos 3.1 (<https://www.cbs.dtu.dk/services/NetPhos/>) and GPS 3.0 (<https://gps.biocuckoo.cn/>) were used to anticipate probable phosphorylation sites. Apart from phosphorylation sites, we used GPS-MSP 1.0 (<https://msp.biocuckoo.org/>) to predict possible methylation sites, ubiquitylation sites using UbPred (<https://www.ubpred.org>) and BDM-PUB (<https://www.bdmpub.biocuckoo.org>), and SUMOylation sites using GPS-SUMO (<https://sumosp.bioc>).

## **6. RESULTS**

### **For SMAD4 gene:**

#### **6.1 SNP annotation**

We used the NCBI dbSNP database to find SMAD4 SNPs, which comprised a total of 22,372 SNPs. 477 SNPs were nsSNPs and the SNPs that were synonymous are 273 in the coding sequence. Only nsSNPs were chosen for further analysis in this study.

#### **6.2 Identification of deleterious nsSNPs**

We used six different in silico nsSNP prediction algorithms for this (SIFT, PANTHER, PolyPhen-2, PROVEAN, SNPs&GO, and PredictSNP). In all computational techniques, 23 nsSNPs were projected to be harmful SNPs out of 477.

**Table 1:** Identification of deleterious nsSNPs by six in silico programs

SNP Id	AA Change	SIFT	POLYPHEN 2	PredictSNP	SNPs&GO	PROVEAN	PANTHER
rs80338963	R361S	DELETERIOUS	DAMAGING	DELETERIOUS	DISEASE	DELETERIOUS	DAMAGING
rs80338963	R361G	DELETERIOUS	DAMAGING	DELETERIOUS	DISEASE	DELETERIOUS	DAMAGING
rs80338963	R361C	DELETERIOUS	DAMAGING	DELETERIOUS	DISEASE	DELETERIOUS	DAMAGING
rs139569694	N369S	DELETERIOUS	DAMAGING	DELETERIOUS	DISEASE	DELETERIOUS	DAMAGING
rs281875324	E330G	DELETERIOUS	DAMAGING	DELETERIOUS	DISEASE	DELETERIOUS	DAMAGING
rs377767339	C324R	DELETERIOUS	DAMAGING	DELETERIOUS	DISEASE	DELETERIOUS	DAMAGING
rs377767342	E330K	DELETERIOUS	DAMAGING	DELETERIOUS	DISEASE	DELETERIOUS	DAMAGING
rs377767345	G352E	DELETERIOUS	DAMAGING	DELETERIOUS	DISEASE	DELETERIOUS	DAMAGING
rs377767346	Y353S	DELETERIOUS	DAMAGING	DELETERIOUS	DISEASE	DELETERIOUS	DAMAGING
rs377767347	R361H	DELETERIOUS	DAMAGING	DELETERIOUS	DISEASE	DELETERIOUS	DAMAGING
rs377767347	R361L	DELETERIOUS	DAMAGING	DELETERIOUS	DISEASE	DELETERIOUS	DAMAGING
rs377767348	C363R	DELETERIOUS	DAMAGING	DELETERIOUS	DISEASE	DELETERIOUS	DAMAGING
rs377767350	L364W	DELETERIOUS	DAMAGING	DELETERIOUS	DISEASE	DELETERIOUS	DAMAGING

rs377767353	R380K	DELETERIOUS	DAMAGING	DELETERIOUS	DISEASE	DELETERIOUS	DAMAGING
rs377767355	I383K	DELETERIOUS	DAMAGING	DELETERIOUS	DISEASE	DELETERIOUS	DAMAGING
rs377767367	G491V	DELETERIOUS	DAMAGING	DELETERIOUS	DISEASE	DELETERIOUS	DAMAGING
rs377767369	W509R	DELETERIOUS	DAMAGING	DELETERIOUS	DISEASE	DELETERIOUS	DAMAGING
rs377767371	G510V	DELETERIOUS	POSSIBLY DAMAGING	DELETERIOUS	DISEASE	DELETERIOUS	DAMAGING
rs377767375	W524L	DELETERIOUS	DAMAGING	DELETERIOUS	DISEASE	DELETERIOUS	DAMAGING
rs377767381	L533V	DELETERIOUS	DAMAGING	DELETERIOUS	DISEASE	DELETERIOUS	DAMAGING
rs377767382	L533P	DELETERIOUS	DAMAGING	DELETERIOUS	DISEASE	DELETERIOUS	DAMAGING
rs377767382	L533R	DELETERIOUS	DAMAGING	DELETERIOUS	DISEASE	DELETERIOUS	DAMAGING
rs121912580	G386D	DELETERIOUS	DAMAGING	DELETERIOUS	DISEASE	DELETERIOUS	DAMAGING

### **6.3 Identification of nsSNPs on the domains of SMAD4**

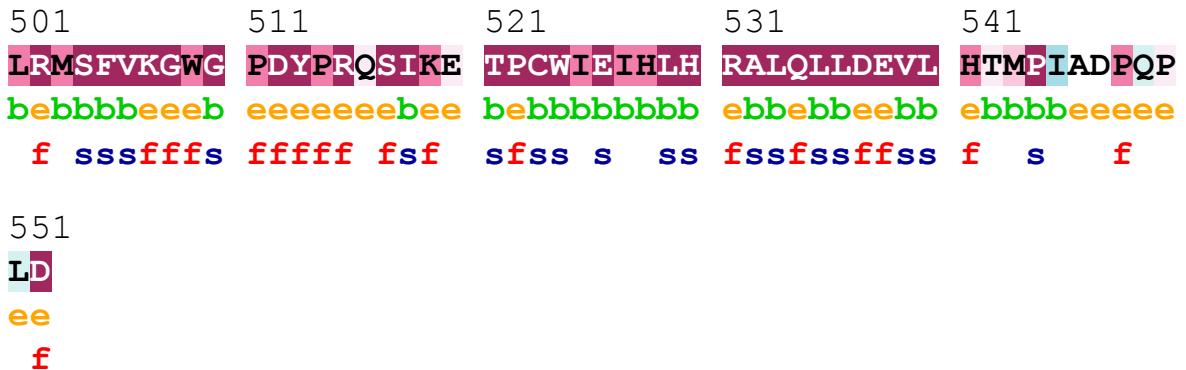
It showed that SMAD4 has 2 functional domains, MH1 (18-142) and MH2 domain (321-530), and showed that all 23 nsSNPs are located on the MH2 domain.

## 6.4 Evolutionary conservation analysis

It identified structural and functional residues of the 23 high-risk nsSNPs in the SMAD4 protein using evolutionary conservation and solvent accessibility. We discovered that R361S, R361G, R361C, N369S, E330G, E330K, Y353S, R361H, R361L, R380K, W509R, and G386D residues are highly conserved and functional, whereas C363R, L364W, I383K, G352E, G491V, G510V, W524L, L533V, L53 These 21 residues are all well conserved. C324R is also expected to be saved and buried.

## ConSurf Results

51	61	71	81	91
<b>KDELDSLITA</b>	<b>ITTNGAHPSK</b>	<b>CVTIQRTLDG</b>	<b>RLQVAGRKGF</b>	<b>PHVIYARLWR</b>
eeeebeebbeb	bbbeeeeeeee	bbebeeeebee	ebebbeeeeb	bbbbbbbbbbe
<b>ffffsfssfs</b>	<b>sfff f f</b>	<b>sfsffffsff</b>	<b>f fssffffs</b>	<b>fss sss f</b>
101	111	121	131	141
<b>WPDLHKNELK</b>	<b>HVKYCQYAFD</b>	<b>LKCDSVCVNP</b>	<b>YHYERVVSPG</b>	<b>IDLSGLTLQS</b>
beebbeeeeeee	ebebbebbbe	bebeebbbb	eebeebbbeb	bbbbbbbbbbee
<b>ffsfffffff</b>	<b>f f s sf</b>	<b>f f sssss</b>	<b>ff ff f</b>	
151	161	171	181	191
<b>NAPSSMMVKD</b>	<b>EYVHDFEGQP</b>	<b>SLSTEGHSIQ</b>	<b>TIQHPPSNRA</b>	<b>STETYSTPAL</b>
eeeeeebbee	eeeeeeeeeee	eeeeeeeeeee	eebeeeeeeee	eeeeeeeeeee
201	211	221	231	241
<b>LAPSESNATS</b>	<b>TANFPNIPVA</b>	<b>STSQPASILG</b>	<b>GSHSEGLLQI</b>	<b>ASGPQPGQQQ</b>
eeeeeeeeeee	eeebeebbeee	eeeeeeeeeebe	beeeeebbbeb	eeeeeeeeeee
251	261	271	281	291
<b>NGFTGQPATY</b>	<b>HHNSTTTWTG</b>	<b>SRTAPYTPNL</b>	<b>FHHQNNGHLQH</b>	<b>HPPMPHPGH</b>
eeeeeeeeeee	eeeeeeebee	eeeeeeeeeee	eeeeeeeeeee	eeeeeeeeeee
301	311	321	331	341
<b>YWPVHNELAf</b>	<b>QPPISNHPAP</b>	<b>EYWCSIAYFE</b>	<b>MDVQVGETFK</b>	<b>VPSSCPIVTV</b>
ebbeeeeeeb	eeeeeeeeeee	ebbbbbbbee	bebebeeebe	eeeebebbbbb
f	f	s	sf	sf f f s s
351	361	371	381	391
<b>DGYVDPSGGD</b>	<b>RFCLGQLSNV</b>	<b>HRTEAIERAR</b>	<b>LHIGKGVQLE</b>	<b>CKGEGDVWVR</b>
ebbbbbeeee	ebbbbebeeb	eebebbeebe	bebeeebebe	beeeeebbbe
fsfss ff	fssssfsffs	ffs s fsf	sfsfff fsf	ffff f
401	411	421	431	441
<b>CLSDHAVFVQ</b>	<b>SYYLDREAGR</b>	<b>APGDAVHKIY</b>	<b>PSAYIKVFDL</b>	<b>RQCHROMQQQ</b>
bbeeebbbbb	bbbbeeeeeee	beeebbbebb	ebbbbebbbeb	eebeeeeeeee
s fffsssss	s ssffffff	fffssssfss	f s sfs	f ff f
451	461	471	481	491
<b>AATAQAAAAAA</b>	<b>QAAAVAGNIP</b>	<b>GPGSVGGIAP</b>	<b>AISLSAAAGI</b>	<b>GVDDLRLLCI</b>
bbebebbbeb	ebbebbebbe	beeeebbb	bbbbbbbbb	bbeebbebbb
f f	ff	ff	s	ssffffssss



**Figure 14:** CONSURF results

**Table 2:** Examination of evolutionary conservation of amino acids of SMAD4 by Consurf

SNP Id	Residue and Position	Conservation score	Prediction
rs80338963	R361S	9	Greatly exposed and conserved (F)
rs80338963	R361G	9	Greatly exposed and conserved (F)
rs80338963	R361C	9	Greatly exposed and conserved (F)
rs139569694	N369S	9	Greatly exposed and conserved (F)
rs281875324	E330G	9	Greatly exposed and conserved (F)
rs377767339	C324R	7	Conserved and buried
rs377767342	E330K	8	Greatly exposed and conserved (F)
rs377767345	G352E	9	Greatly buried and exposed (S)
rs377767346	Y353S	9	Greatly exposed and conserved (F)
rs377767346	R361L	9	Greatly exposed and conserved (F)
rs377767348	C363R	9	Greatly buried and exposed (S)
rs377767350	L364W	9	Greatly buried and exposed (S)
rs377767353	R380K	9	Greatly exposed and conserved (F)
rs377767355	I383K	9	Greatly buried and exposed (S)
rs377767367	G491V	8	Greatly buried and exposed (S)
rs377767369	W509R	8	Greatly exposed and conserved (F)
rs377767371	G510V	9	Greatly buried and exposed (S)
rs377767375	W524L	9	Greatly buried and exposed (S)
rs377767381	L533V	9	Greatly buried and exposed (S)
rs377767382	L533P	9	Greatly buried and exposed (S)
rs377767382	L533R	9	Greatly buried and exposed (S)
rs121912580	G386D	9	Greatly buried and conserved (F)

## 6.5 Structure analysis of wild type and mutant models

The natural and 23 mutations of the SMAD4 protein were three dimensionally modelled using the Phyre2 homology simulation platform. We selected [1khx](#) template for MH2 domain as all nsSNPs lie on MH2 domain. The 23 detrimental nsSNPs were each changed in the templates' native sequence, and 3D models for all of the mutants were predicted.

On the basis of TM-score and RMSD scores, Phyre2 was utilised to compare the structural similarities of normal and mutant models. The RMSD value was high in all of the mutant models reported in the MH2 domain. The presence of structural changes between the wild type and mutant models is shown by a larger RMSD value. SWISS MODEL was used to explore the 3D structures of the 23 nsSNPs in the SMAD4 protein in order to investigate protein solvation and torsion. The d1khxa template was used to look at the nsSNPs.

**Table 3:** TM score and RMSD predictions for nsSNPs in SMAD4

Serial No.	AA substitution	Domain	Tm Score	RMSD
1.	R361S	MH2 domain	0.89	1.244
2.	R361G		0.83	1.324
3.	R361C		0.83	1.327
4.	N369S		0.89	1.244
5.	E330G		0.89	1.244
6.	C324R		0.89	1.247
7.	E330K		0.89	1.247
8.	G352E		0.89	1.244
9.	Y353S		0.89	1.244
10.	R361H		0.89	1.244
11.	R361L		0.89	1.244
12.	C363R		0.89	1.245
13.	L364W		0.89	1.244
14.	R380K		0.89	1.244
15.	I383K		0.89	1.245
16.	G491V		0.89	1.244
17.	W509R		0.89	1.244
18.	G510V		0.89	1.244
19.	W524L		0.89	1.244
20.	L533V		0.89	1.244
21.	L533P		0.89	1.244
22.	L533R		0.89	1.244
23.	G386D		0.89	1.244

**Table 4:** Various parameters of 23 nsSNPs investigated by the Swiss Model

<b>SNPs</b>	<b>QMEN</b>	<b>Cb</b>	<b>All atom</b>	<b>Solvation</b>	<b>Torsion</b>	<b>Template</b>
Wild Type	-3.72	-2.3	-1.64	-2.80	-2.28	d1khxa
R361S	-3.89	-2.34	-1.77	-2.80	-2.44	d1khxa
R361G	-4.18	-2.51	-1.76	-2.78	-2.71	d1khxa
R361C	-3.78	-2.41	-1.77	-2.74	-2.33	d1khxa
N369S	-3.86	-2.35	-1.62	-2.77	-2.42	d1khxa
E330G	-3.52	-1.28	-1.63	-2.62	-2.34	d1khxa
C324R	-3.84	-2.26	-1.7	-2.84	-2.39	d1khxa
E330K	-3.75	-2.32	-1.84	-2.73	-2.32	d1khxa
G352E	-3.43	-1.58	-1.61	-2.81	-2.12	d1khxa
Y353S	-3.66	-2.21	-1.63	-2.87	-2.21	d1khxa
R361H	-4.03	-2.41	-1.79	-2.70	-2.60	d1khxa
R361L	-3.41	-1.39	-1.63	-2.74	-2.16	d1khxa
C363R	-3.67	-2.24	-1.55	-2.90	-2.20	d1khxa
L364W	-3.37	-2.39	-1.73	-2.83	-1.88	d1khxa
R380K	-3.86	-2.27	-1.65	-2.84	-2.41	d1khxa
I383K	-3.32	-1.12	-1.48	-2.88	-2.08	d1khxa
G491V	-4.28	-1.33	-1.71	-2.91	-3.01	d1khxa
W509R	-3.79	-2.25	-1.65	-2.77	-2.37	d1khxa
G510V	-4.33	-2.22	-1.68	-2.80	-2.93	d1khxa
W524L	-3.25	-1.37	-1.53	-2.62	-2.06	d1khxa
L533V	-3.69	-2.14	-1.66	-2.86	-2.26	d1khxa
L533P	-4.01	-2.08	-1.61	-2.89	-2.59	d1khxa
L533R	-3.98	-1.90	-1.72	-2.80	2.63	d1khxa
G386D	-3.44	-1.35	-1.64	-2.82	-2.17	d1khxa

### 6.6 Determination of protein structural stability

13 of the 23 nsSNPs were shown to be destabilising the protein and ten were stabilising the protein. These nsSNPs have been taken further for analysis.

Out of 13, 6 were shown to create a structural damage according to MISSENSE3D tool.

**Table 5:** CUPSAT predictions on the effect of nsSNPs on protein stability

<b>Amino acid change</b>	<b>Stabilising or destabilising</b>	<b>Predicted <math>\Delta\Delta G</math></b>
R361S	Destabilising	-4.27
R361G	Destabilising	-6.6
R361C	Stabilising	2.85
N369S	Destabilising	-0.65
E330G	Destabilising	-2.51
C324R	Destabilising	-0.34
E330K	Destabilising	-2.52
G352E	Destabilising	-1.36
Y353S	Stabilising	0.58
R361H	Destabilising	-0.74

R361L	Destabilising	-0.95
C363R	Stabilising	1.5
L364W	Destabilising	-2.35
R380K	Destabilising	-0.17
I383K	Stabilising	7.4
G491V	Stabilising	10.91
W509R	Stabilising	4.84
G510V	Destabilising	-8.69
W524L	Stabilizing	1.59
L533V	Destabilising	-0.64
L533P	Stabilizing	0.72
L533R	Stabilizing	0.07
G386D	Stabilizing	1.68

### 6.7 Structural effect of point mutation on SMAD4

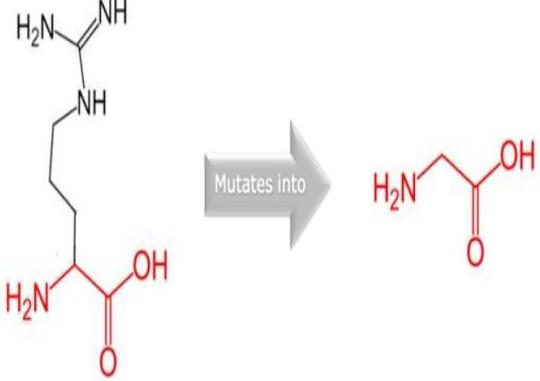
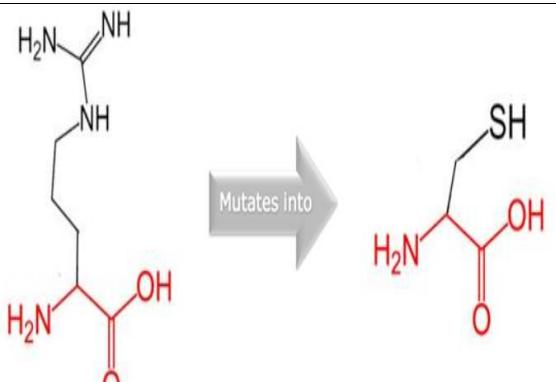
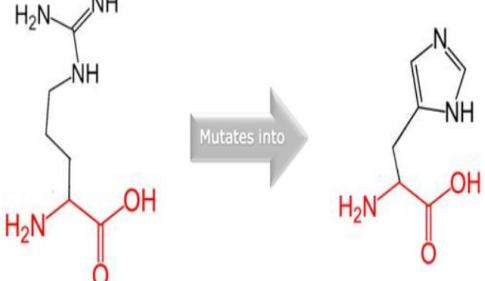
Larger C324R, E330K, G352E, C363R, L364W, I383K, G491V, G510V, L533R, and G386D residues are more water repellent than wild type residues, according to the Project HOPE website, and these changes in size and hydrophobicity can disrupt H-bond interactions with neighbouring molecules.

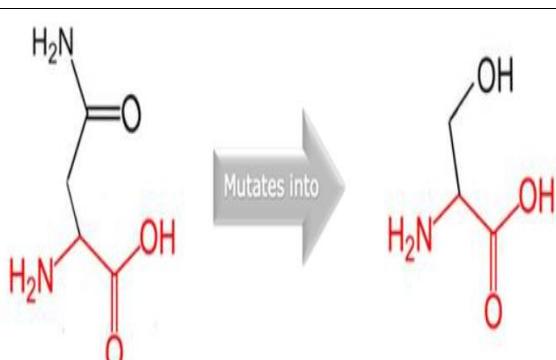
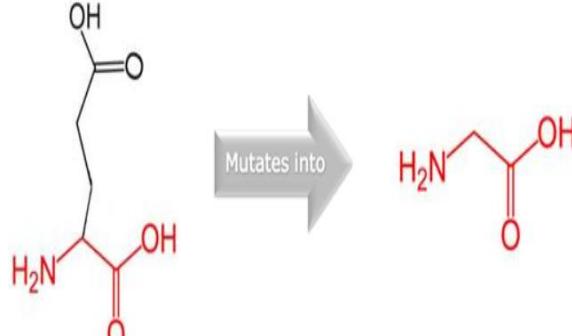
The substitution substitutes glycine that was initially situated in a bend curvature, according to the Missense3D tool. Furthermore, the side-chain and main-chain H-bond(s) produced by wild type were disrupted by E330G, E330K, C363R, and W524L, which were buried with different amino acids.

Out of the 13 nsSNPs which were destabilising only six (E330G, C324R, E330K, G352E, G510V, L533V) of them were found to create a structural change according MISSENSE3D tool.

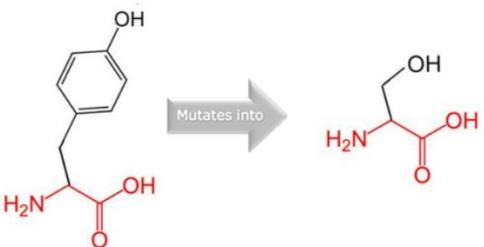
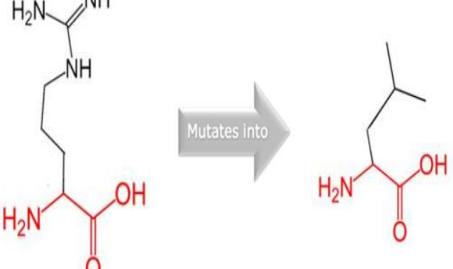
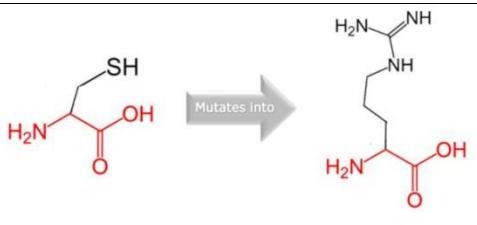
**Table 6:** Structural effect of 23nsSNPs over SMAD4 protein using Project Hope

Residue	Structure	Properties
R361S		<p>When compared to the size of the wild-type residue, the mutant residue is significantly smaller.</p> <p>The charge possessed by the mutant-type residue is NEUTRAL whereas the wild-type residue has a net POSITIVE charge.</p> <p>The mutant-type residue shows the property of hydrophobicity.</p> <p>In the UniProt domain MH2, the mutation was detected. Because of the mutation, a new amino acid is</p>

		introduced which has completely new properties that can cause inactivation or altered functioning of the domain
R361G		<p>When compared to the size of the wild-type residue, the mutant residue is significantly smaller.</p> <p>The charge possessed by the mutant-type residue is NEUTRAL whereas the wild-type residue has a net POSITIVE charge.</p> <p>The mutant-type residue shows the property of hydrophobicity.</p> <p>In the UniProt domain MH2, the mutation was detected. Because of the mutation, a new amino acid is introduced which has completely new properties that can cause inactivation or altered functioning of the domain</p>
R361C		<p>When compared to the size of the wild-type residue, the mutant residue is significantly smaller.</p> <p>The charge possessed by the mutant-type residue is NEUTRAL whereas the wild-type residue has a net POSITIVE charge.</p> <p>The mutant-type residue shows the property of hydrophobicity.</p> <p>In the UniProt domain MH2, the mutation was detected. Because of the mutation, a new amino acid is introduced which has completely new properties that can cause inactivation or altered functioning of the domain</p>
R361H		<p>When compared to the size of the wild-type residue, the mutant residue is significantly smaller.</p> <p>The charge possessed by the mutant-type residue is NEUTRAL whereas the wild-type residue has a net POSITIVE charge.</p> <p>The mutant-type residue shows the property of hydrophobicity.</p>

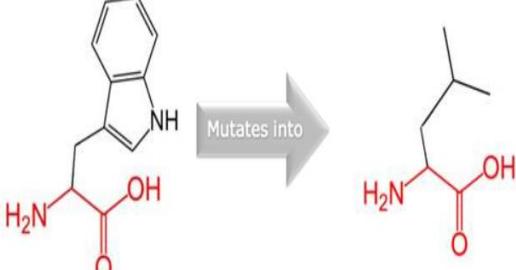
		In the UniProt domain MH2, the mutation was detected. Because of the mutation, a new amino acid is introduced which has completely new properties that can cause inactivation or altered functioning of the domain
N369S		<p>When compared to the size of the wild-type residue, the mutant residue is significantly smaller.</p> <p>The charge possessed by the mutant-type residue is NEUTRAL whereas the wild-type residue has a net POSITIVE charge.</p> <p>The mutant-type residue shows the property of hydrophobicity.</p> <p>In the UniProt domain MH2, the mutation was detected. Because of the mutation, a new amino acid is introduced which has completely new properties that can cause inactivation or altered functioning of the domain.</p>
E330G		<p>When compared to the size of the wild-type residue, the mutant residue is significantly smaller.</p> <p>The charge possessed by the mutant-type residue is NEUTRAL whereas the wild-type residue has a net POSITIVE charge.</p> <p>The mutant-type residue shows the property of hydrophobicity.</p> <p>In the UniProt domain MH2, the mutation was detected. Because of the mutation, a new amino acid is introduced which has completely new properties that can cause inactivation or altered functioning of the domain.</p>

C324R		<p>When compared to the size of the wild-type residue, the mutant residue is significantly larger.</p> <p>The charge possessed by the mutant-type residue is POSITIVE whereas the wild-type residue has a net NEUTRAL charge.</p> <p>The wild-type residue shows the property of hydrophobicity more when compared to mutant-type.</p> <p>In the UniProt domain MH2, the mutation was detected. Because of the mutation, a new amino acid is introduced which has completely new properties that can cause inactivation or altered functioning of the domain.</p>
E330K		<p>When compared to the size of the wild-type residue, the mutant residue is significantly larger.</p> <p>The charge possessed by the mutant-type residue is POSITIVE whereas the wild-type residue has a net NEUTRAL charge.</p> <p>The wild-type residue shows the property of hydrophobicity more when compared to mutant-type.</p> <p>In the UniProt domain MH2, the mutation was detected. Because of the mutation, a new amino acid is introduced which has completely new properties that can cause inactivation or altered functioning of the domain.</p>
G352E		<p>When compared to the size of the wild-type residue, the mutant residue is significantly larger.</p> <p>The charge possessed by the mutant-type residue is POSITIVE whereas the wild-type residue has a net NEUTRAL charge.</p> <p>The wild-type residue shows the property of hydrophobicity more when compared to mutant-type.</p>

		<p>In the UniProt domain MH2, the mutation was detected. Because of the mutation, a new amino acid is introduced which has completely new properties that can cause inactivation or altered functioning of the domain.</p>
Y353S		<p>When compared to the size of the wild-type residue, the mutant residue is significantly smaller.</p> <p>The charge possessed by the mutant-type residue is NEUTRAL whereas the wild-type residue has a net POSITIVE charge.</p> <p>The mutant-type residue shows the property of hydrophobicity.</p> <p>In the UniProt domain MH2, the mutation was detected. Because of the mutation, a new amino acid is introduced which has completely new properties that can cause inactivation or altered functioning of the domain.</p>
R361L		<p>When compared to the size of the wild-type residue, the mutant residue is significantly smaller.</p> <p>The charge possessed by the mutant-type residue is NEUTRAL whereas the wild-type residue has a net POSITIVE charge.</p> <p>The mutant-type residue shows the property of hydrophobicity.</p> <p>In the UniProt domain MH2, the mutation was detected. Because of the mutation, a new amino acid is introduced which has completely new properties that can cause inactivation or altered functioning of the domain.</p>
C363R		<p>When compared to the size of the wild-type residue, the mutant residue is significantly larger.</p> <p>The charge possessed by the mutant-type residue is POSITIVE whereas the wild-type residue has a net NEUTRAL charge.</p>

		<p>The wild-type residue shows the property of hydrophobicity more when compared to mutant-type.</p> <p>In the UniProt domain MH2, the mutation was detected. Because of the mutation, a new amino acid is introduced which has completely new properties that can cause inactivation or altered functioning of the domain.</p>
L364W	<p>Mutates into</p>	<p>When compared to the size of the wild-type residue, the mutant residue is significantly larger.</p> <p>The charge possessed by the mutant-type residue is POSITIVE whereas the wild-type residue has a net NEUTRAL charge.</p> <p>The wild-type residue shows the property of hydrophobicity more when compared to mutant-type.</p> <p>In the UniProt domain MH2, the mutation was detected. Because of the mutation, a new amino acid is introduced which has completely new properties that can cause inactivation or altered functioning of the domain.</p>
R380K	<p>Mutates into</p>	<p>When compared to the size of the wild-type residue, the mutant residue is significantly smaller.</p> <p>The charge possessed by the mutant-type residue is NEUTRAL whereas the wild-type residue has a net POSITIVE charge.</p> <p>The mutant-type residue shows the property of hydrophobicity.</p> <p>In the UniProt domain MH2, the mutation was detected. Because of the mutation, a new amino acid is introduced which has completely new properties that can cause inactivation or altered functioning of the domain.</p>

I383K	<p>The diagram illustrates the mutation of Isoleucine (I) to Lysine (K). On the left, the structure of Isoleucine (I) is shown: a central carbon atom bonded to an amino group (<math>\text{H}_2\text{N}-</math>), a carboxyl group (<math>-\text{COOH}</math>), a methyl group (<math>\text{CH}_3-</math>), and a hydrogen atom (<math>\text{H}-</math>). An arrow labeled "Mutates into" points to the structure of Lysine (K) on the right: a central carbon atom bonded to an amino group (<math>\text{H}_2\text{N}-</math>), a carboxyl group (<math>-\text{COOH}</math>), a propyl side chain (<math>\text{CH}_2\text{CH}_2\text{CH}_3-</math>), and a hydrogen atom (<math>\text{H}-</math>).</p>	<p>When compared to the size of the wild-type residue, the mutant residue is significantly larger.</p> <p>The charge possessed by the mutant-type residue is POSITIVE whereas the wild-type residue has a net NEUTRAL charge.</p> <p>The wild-type residue shows the property of hydrophobicity more when compared to mutant-type.</p> <p>In the UniProt domain MH2, the mutation was detected. Because of the mutation, a new amino acid is introduced which has completely new properties that can cause inactivation or altered functioning of the domain.</p>
G491V	<p>The diagram illustrates the mutation of Glutamate (E) to Valine (V). On the left, the structure of Glutamate (E) is shown: a central carbon atom bonded to an amino group (<math>\text{H}_2\text{N}-</math>), a carboxyl group (<math>-\text{COOH}</math>), a long-chain side chain (<math>\text{CH}_2\text{CH}_2\text{CH}_2\text{CH}_2\text{CH}_2\text{CH}_3-</math>), and a hydrogen atom (<math>\text{H}-</math>). An arrow labeled "Mutates into" points to the structure of Valine (V) on the right: a central carbon atom bonded to an amino group (<math>\text{H}_2\text{N}-</math>), a carboxyl group (<math>-\text{COOH}</math>), a methyl group (<math>\text{CH}_3-</math>), and a hydrogen atom (<math>\text{H}-</math>).</p>	<p>When compared to the size of the wild-type residue, the mutant residue is significantly larger.</p> <p>The charge possessed by the mutant-type residue is POSITIVE whereas the wild-type residue has a net NEUTRAL charge.</p> <p>The wild-type residue shows the property of hydrophobicity more when compared to mutant-type.</p> <p>In the UniProt domain MH2, the mutation was detected. Because of the mutation, a new amino acid is introduced which has completely new properties that can cause inactivation or altered functioning of the domain.</p>
W509R	<p>The diagram illustrates the mutation of Tryptophan (W) to Arginine (R). On the left, the structure of Tryptophan (W) is shown: a central carbon atom bonded to an amino group (<math>\text{H}_2\text{N}-</math>), a carboxyl group (<math>-\text{COOH}</math>), a long-chain side chain containing a tryptophan ring (<math>\text{CH}_2\text{CH}(\text{NH}_2)\text{C}_6\text{H}_4\text{CH}_2-</math>), and a hydrogen atom (<math>\text{H}-</math>). An arrow labeled "Mutates into" points to the structure of Arginine (R) on the right: a central carbon atom bonded to an amino group (<math>\text{H}_2\text{N}-</math>), a carboxyl group (<math>-\text{COOH}</math>), a propyl side chain (<math>\text{CH}_2\text{CH}_2\text{CH}_3-</math>), and a hydrogen atom (<math>\text{H}-</math>).</p>	<p>When compared to the size of the wild-type residue, the mutant residue is significantly smaller.</p> <p>The charge possessed by the mutant-type residue is NEUTRAL whereas the wild-type residue has a net POSITIVE charge.</p> <p>The mutant-type residue shows the property of hydrophobicity.</p>

		<p>In the UniProt domain MH2, the mutation was detected. Because of the mutation, a new amino acid is introduced which has completely new properties that can cause inactivation or altered functioning of the domain.</p>
G510V		<p>When compared to the size of the wild-type residue, the mutant residue is significantly larger.</p> <p>The charge possessed by the mutant-type residue is POSITIVE whereas the wild-type residue has a net NEUTRAL charge.</p> <p>The wild-type residue shows the property of hydrophobicity more when compared to mutant-type.</p> <p>In the UniProt domain MH2, the mutation was detected. Because of the mutation, a new amino acid is introduced which has completely new properties that can cause inactivation or altered functioning of the domain.</p>
W524L		<p>When compared to the size of the wild-type residue, the mutant residue is significantly smaller.</p> <p>The charge possessed by the mutant-type residue is NEUTRAL whereas the wild-type residue has a net POSITIVE charge.</p> <p>The mutant-type residue shows the property of hydrophobicity.</p> <p>In the UniProt domain MH2, the mutation was detected. Because of the mutation, a new amino acid is introduced which has completely new properties that can cause inactivation or altered functioning of the domain.</p>

L533V		<p>When compared to the size of the wild-type residue, the mutant residue is significantly smaller.</p> <p>The charge possessed by the mutant-type residue is NEUTRAL whereas the wild-type residue has a net POSITIVE charge.</p> <p>The mutant-type residue shows the property of hydrophobicity.</p> <p>In the UniProt domain MH2, the mutation was detected. Because of the mutation, a new amino acid is introduced which has completely new properties that can cause inactivation or altered functioning of the domain.</p>
L533P		<p>When compared to the size of the wild-type residue, the mutant residue is significantly smaller.</p> <p>The charge possessed by the mutant-type residue is NEUTRAL whereas the wild-type residue has a net POSITIVE charge.</p> <p>The mutant-type residue shows the property of hydrophobicity.</p> <p>In the UniProt domain MH2, the mutation was detected. Because of the mutation, a new amino acid is introduced which has completely new properties that can cause inactivation or altered functioning of the domain.</p>
L533R		<p>When compared to the size of the wild-type residue, the mutant residue is significantly larger.</p> <p>The charge possessed by the mutant-type residue is POSITIVE whereas the wild-type residue has a net NEUTRAL charge.</p> <p>The wild-type residue shows the property of hydrophobicity more when compared to mutant-type.</p>

		In the UniProt domain MH2, the mutation was detected. Because of the mutation, a new amino acid is introduced which has completely new properties that can cause inactivation or altered functioning of the domain.
G386D	<p>The diagram illustrates the mutation G386D. On the left, the structure of Alanine (G) is shown: a central carbon atom bonded to an amino group (<math>\text{H}_2\text{N}-</math>), a hydrogen atom (<math>\text{H}</math>), a methyl group (<math>\text{CH}_3-</math>), and a carboxylate group (<math>\text{COO}^-</math>). An arrow labeled "Mutates into" points to the structure of Aspartic Acid (D): a central carbon atom bonded to an amino group (<math>\text{H}_2\text{N}-</math>), a hydroxyl group (<math>\text{OH}-</math>), a methyl group (<math>\text{CH}_3-</math>), and a carboxylate group (<math>\text{COO}^-</math>).</p>	<p>When compared to the size of the wild-type residue, the mutant residue is significantly larger.</p> <p>The charge possessed by the mutant-type residue is POSITIVE whereas the wild-type residue has a net NEUTRAL charge.</p> <p>The wild-type residue shows the property of hydrophobicity more when compared to mutant-type.</p> <p>In the UniProt domain MH2, the mutation was detected. Because of the mutation, a new amino acid is introduced which has completely new properties that can cause inactivation or altered functioning of the domain.</p>

**Table 7:** Predicting the structural effect of 23 nsSNPs of SMAD4 protein using Missense 3D tool

AA substitution	PDB ID	Chain	Result Analysis	Detailed Analysis
R361S	1KHX	A	No structural damage detected	
R361G	1KHX	A	No structural damage detected	
R361C	1KHX	A	No structural damage detected	
N369S	1KHX	A	No structural damage detected	
E330G	1KHX	A	H-bond is broken and deeply placed	This substitution breaks all side-chain H-bonds / or side-chain / main-chain H-bonds formed by a hidden GLU

				residue (RSA 3.0 percent ).
C324R	1KHX	A	Buried hydrophilic introduced Buried charge introduced	A buried hydrophobic residue (CYS, RSA 0.0%) is replaced with a hydrophilic residue in this substitution (ARG, RSA 2.4 percent). A buried uncharged residue (CYS, RSA 0.0%) is replaced with a charged residue in this substitution (ARG).
E330K	1KHX	A	Buried charge switch Buried H-bond breakage	This substitution breaks all side-chain H-bonds / or side-chain / main-chain H-bonds formed by a hidden GLU residue (RSA 3.0 percent). A buried negative-charged residue (GLU, RSA 3.0%) is replaced with a positive-charged residue in this substitution (LYS).
G352E	1KHX	A	Gly in a bend	In the bent of the curvature, the position of glycine is replaced by the substitution.
R361H	1KHX	A	No structural damage detected	
R361L	1KHX	A	No structural damage detected	
C363R	1KHX	A	Buried H-bond breakage Buried/exposed switch	This substitution breaks all side-chain H-bonds / or side-chain / main-chain H-bonds formed by a hidden CYS residue (RSA 3.7 percent).  The target variant residue's buried and exposed states are switched as a result of this substitution. ARG is exposed and CYS is hidden in 1KHX (RSA

				3.7 percent) (RSA 17.7 percent ).
L364W	1KHX	A	No structural damage detected	
R380K	1KHX	A	No structural damage detected	
I383K	1KHX		Buried hydrophilic introduced Buried charge introduced	A buried hydrophobic residue (ILE, RSA 0.5%) is replaced by a hydrophilic residue with this substitution (LYS, RSA 0.0 percent ) A buried uncharged residue (ILE, RSA 0.5%) is replaced by a charged residue in this substitution (LYS)
W509R	1KHX		Buried hydrophilic introduced Buried charge introduced	A buried hydrophobic residue (TRP, RSA 2.2%) is replaced with a hydrophilic residue in this substitution (ARG, RSA 2.0 percent). This replacement substitutes a charged residue for a buried uncharged residue (TRP, RSA 2.2%). (ARG).
G510V	1KHX		Clash Disallow phi/psi Gly in a bend	Clash alarm is triggered by this swap. The wild type has a local clash score of 41.44, while the mutant has a local clash score of 59.87. This substitution causes the phi/psi alert to be disabled. For wild-type residues, the phi/psi angles are in the preferred area, while for mutant residues, they are outliers. This substitution takes the place of glycine, which was previously found in a bend curvature.

W524L	1KHX		Buried H-bond breakage	This substitution breaks all side-chain H-bonds / or side-chain / main-chain H-bonds formed by a hidden TRP residue (RSA 5.2 percent ).
L533V	1KHX		No structural damage detected	
L533P	1KHX		No structural damage detected	
L533R	1KHX		No structural damage detected	
G386D	1KHX		Buried charge introduced Buried gly replaced	A buried uncharged residue (GLY, RSA 0.0 percent) is replaced with a charged residue in this substitution (ASP) A buried GLY residue (RSA 0.0 percent) is replaced with a buried ASP residue in this substitution (RSA 0.0 percent)

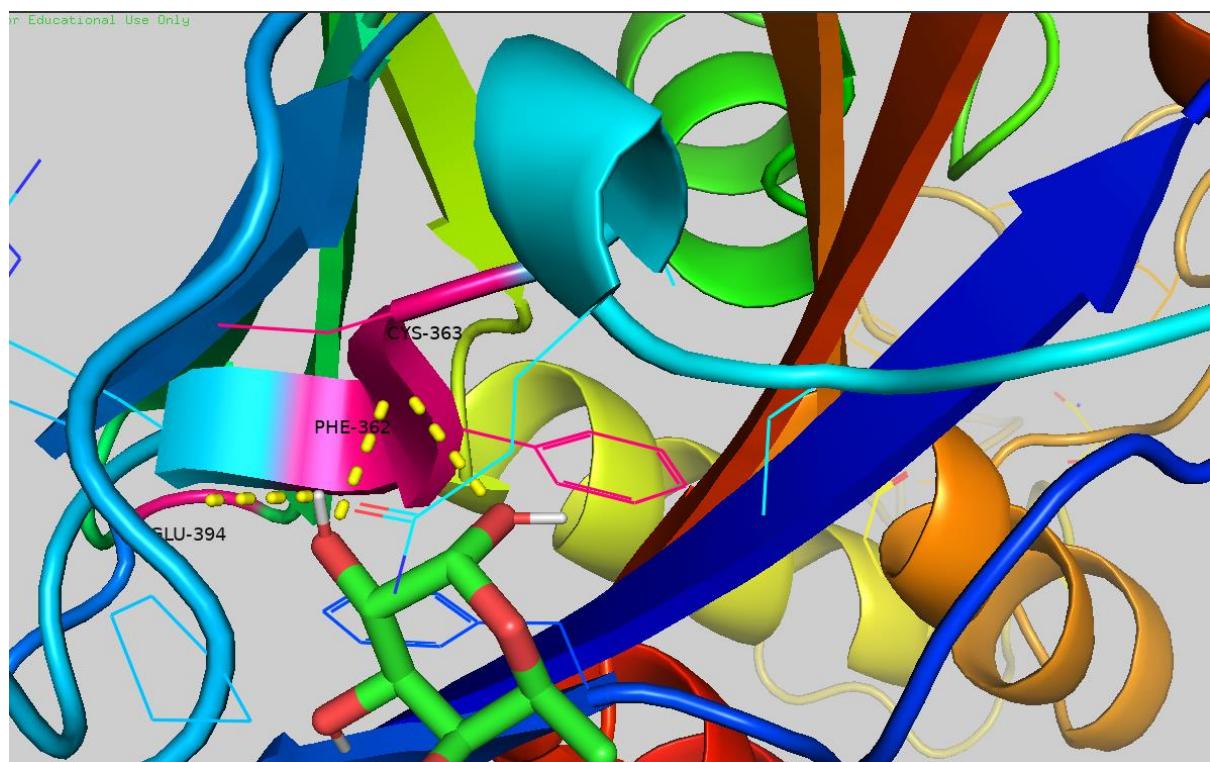
### 6.8 Molecular docking analysis

These studies predicted the binding affinity of the mutated SMAD4 with ALPHA L-FUCOSE. Alpha-L-fucose is an L-fucopyranose having alpha-configuration at the anomeric centre. It has a role as an epitope. It is an enantiomer of an alpha-D-fucose. R361C, Y353S, I383K, W509R, W524L, L533R, and G386D were discovered to improve ligand binding interactions. The binding interactions(kcal/mol) was studied using PYMOL software. With Alpha-L-fucose, the R361C mutation revealed the most significant increase in binding capacity. R361C creates three hydrogen bonds with Alpha-L-fucose, increasing the binding affinity to 5.3 kcal/mol. With Alpha-L-fucose, G491V creates three hydrogen bonds, lowering the binding affinity to 4.6 kcal/mol. Y353S increased the binding affinity to -5.3 kcal/mol. C363R creates three hydrogen bonds with Alpha-L-fucose, lowering the binding interactions to less than 4.4 kcal/mol. I383K created two hydrogen bonds with Alpha-L-fucose, resulting in a binding affinity of 5.3 kcal/mol. W509R increased the binding affinity to 5.3 kcal/mol by creating two hydrogen bonds with Alpha-L-fucose. W524L improved the binding interactions to 5.3 kcal/mol by forming two hydrogen bonds with Alpha-L-fucose. The binding capacity of

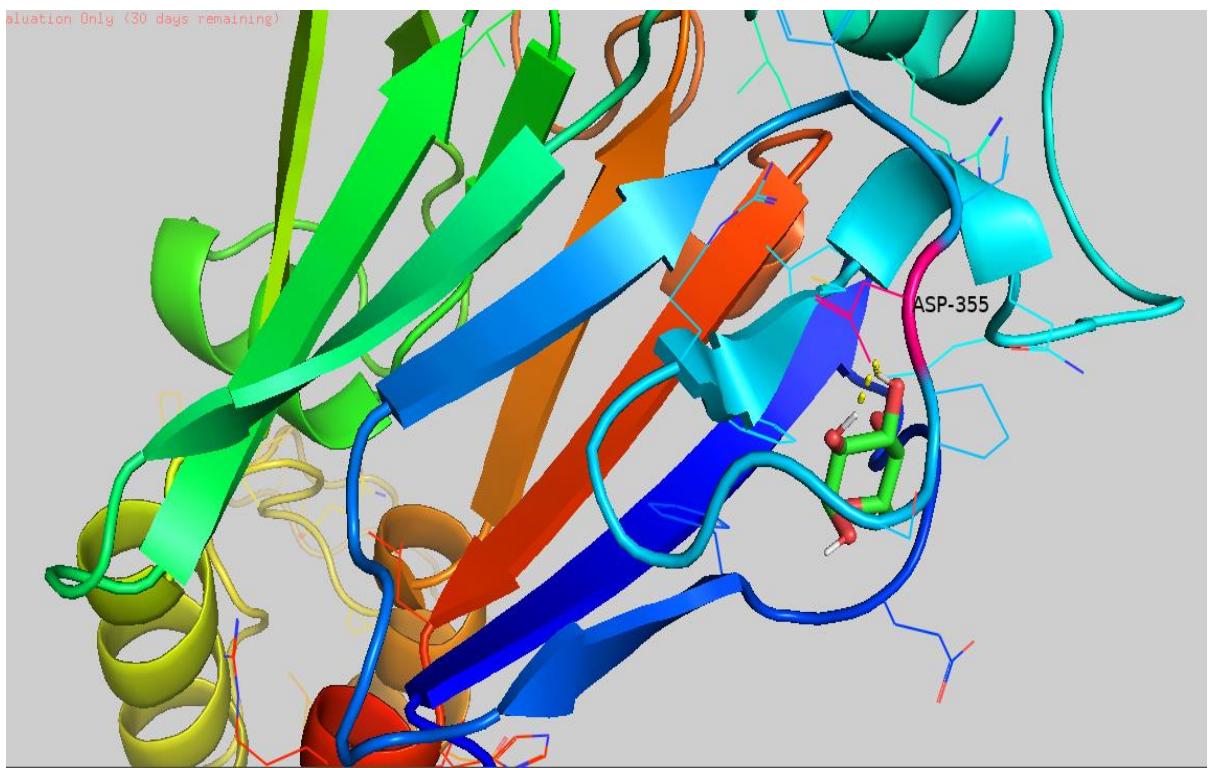
G386D improved to 5.3 kcal/mol by creating two hydrogen bonds with Alpha-L-fucose. L533R improved the binding interactions to 5.3 kcal/mol by creating two hydrogen bonds with Alpha-L-fucose. L533P decreased the binding interactions to 3.9 kcal/mol by forming five hydrogen bonds with Alpha-L-fucose.

**Table 8:** Docking with ALPHA L-FUCOSE

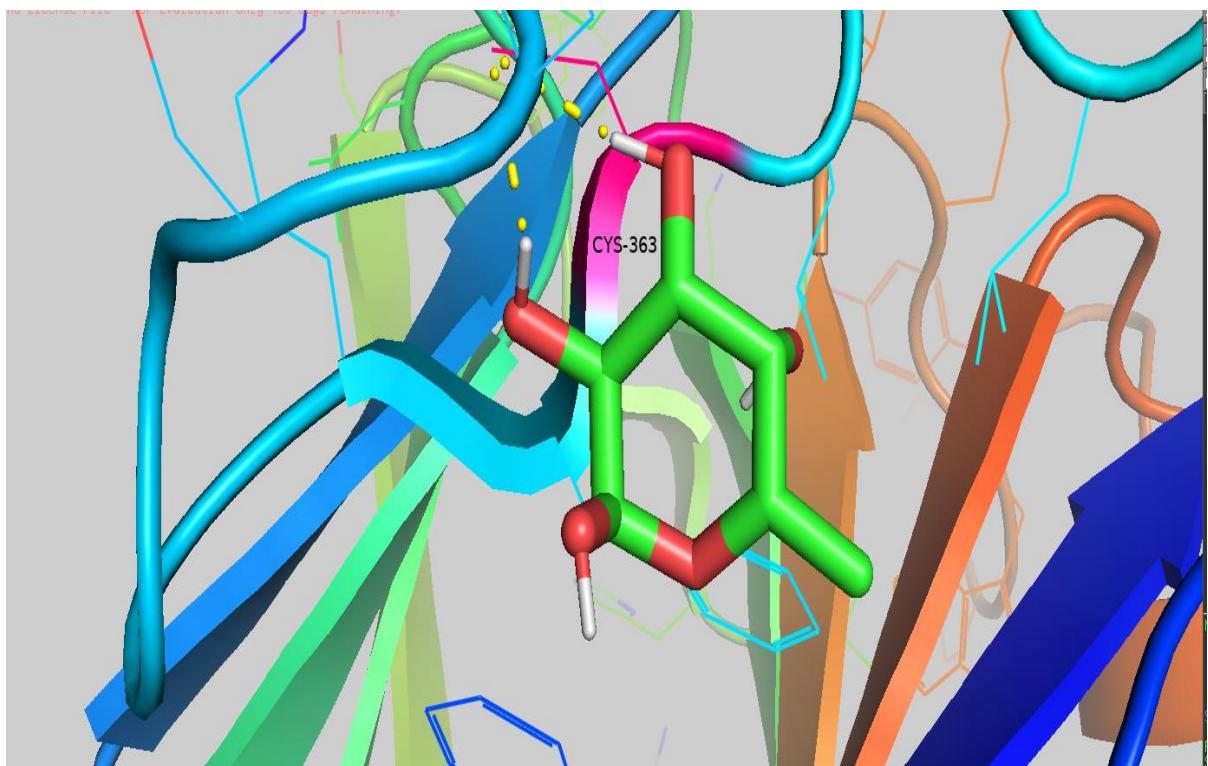
Amino acid change mutant model	Binding affinity(kcal/mol)
WILD	-5.1
R361C	-5.5
Y353S	-5.3
C363R	-4.4
I383K	-5.3
G491V	-4.6
W509R	-5.3
W524L	-5.3
L533P	-3.9
L533R	-5.3
G386D	-5.3



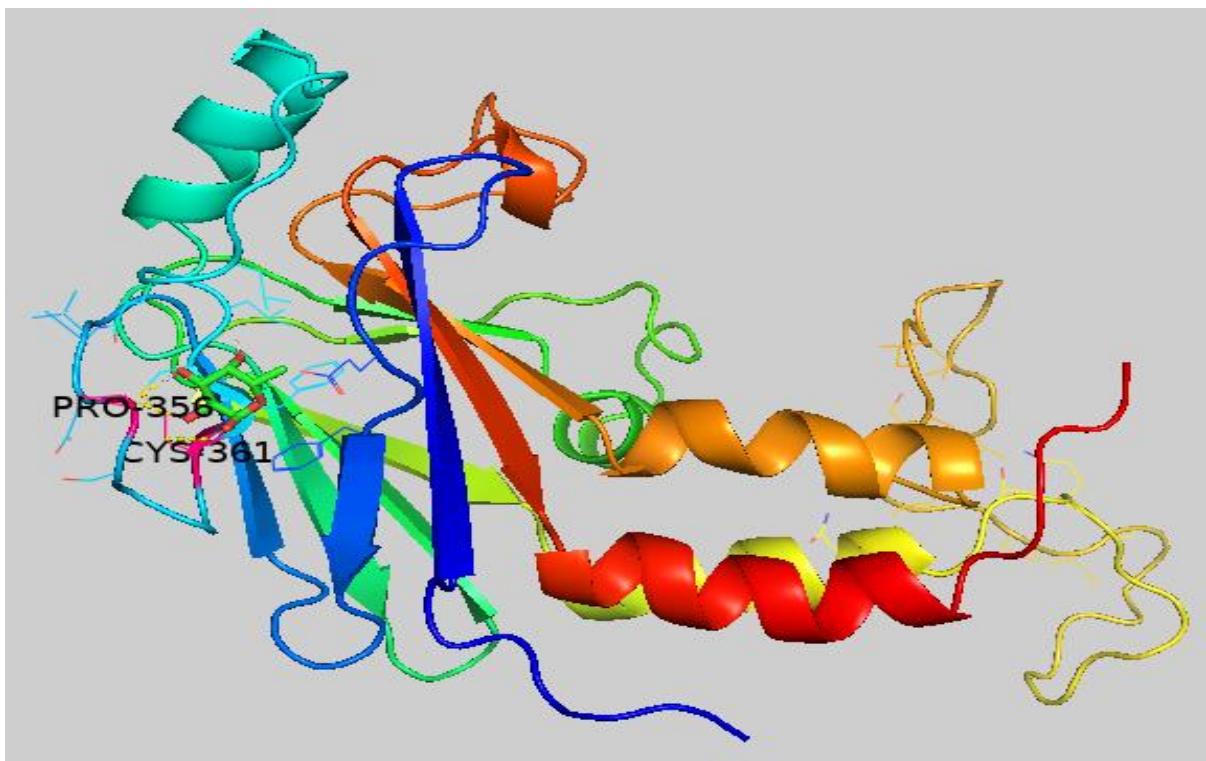
**Figure 15:** Docked wild SMAD4 with ALPHA L-FUCOSE



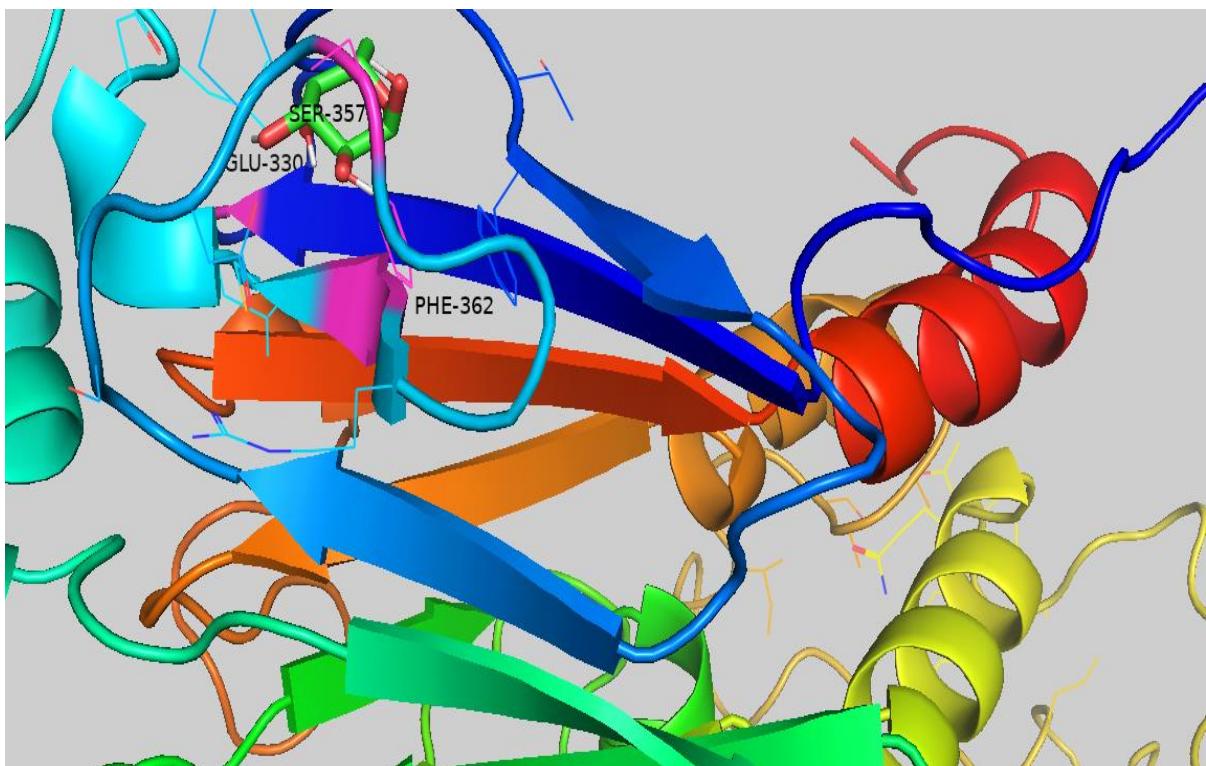
**Figure 16:** Docked W509R mutant SMAD4 with ALPHA L-FUCOSE



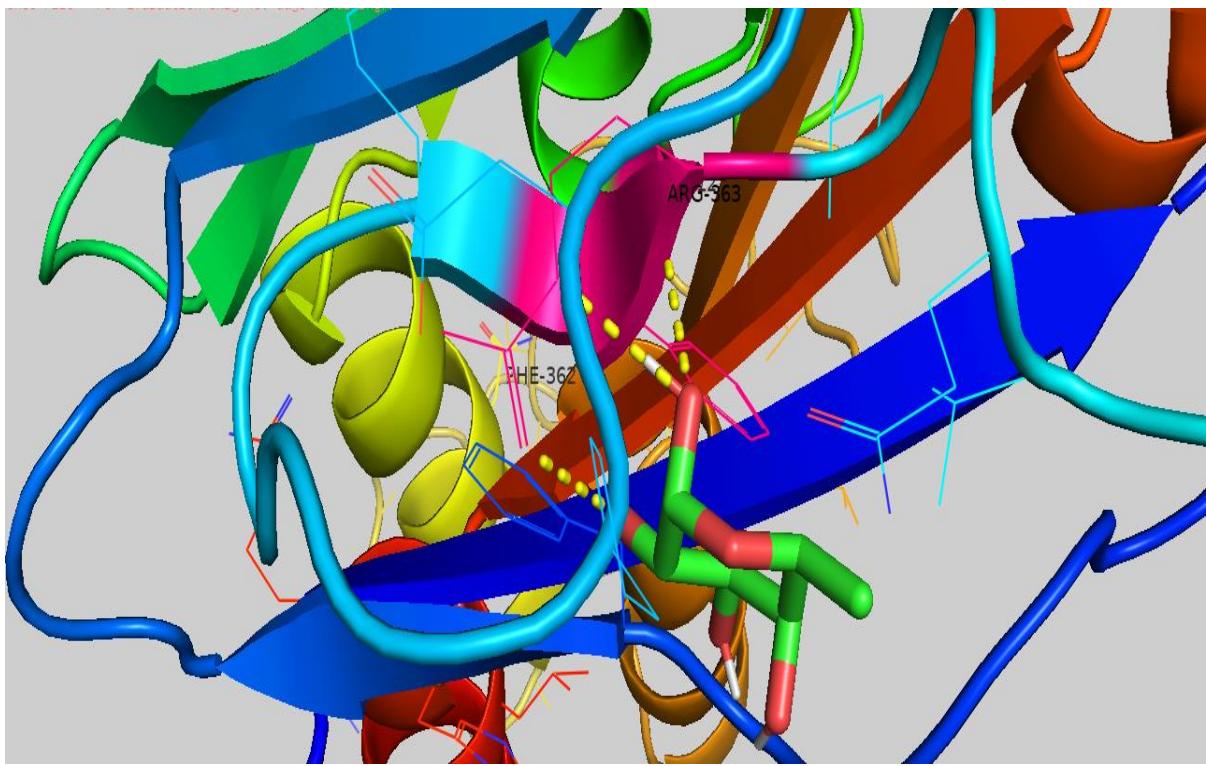
**Figure 17:** Docked I383K mutant SMAD4 with ALPHA L-FUCOSE



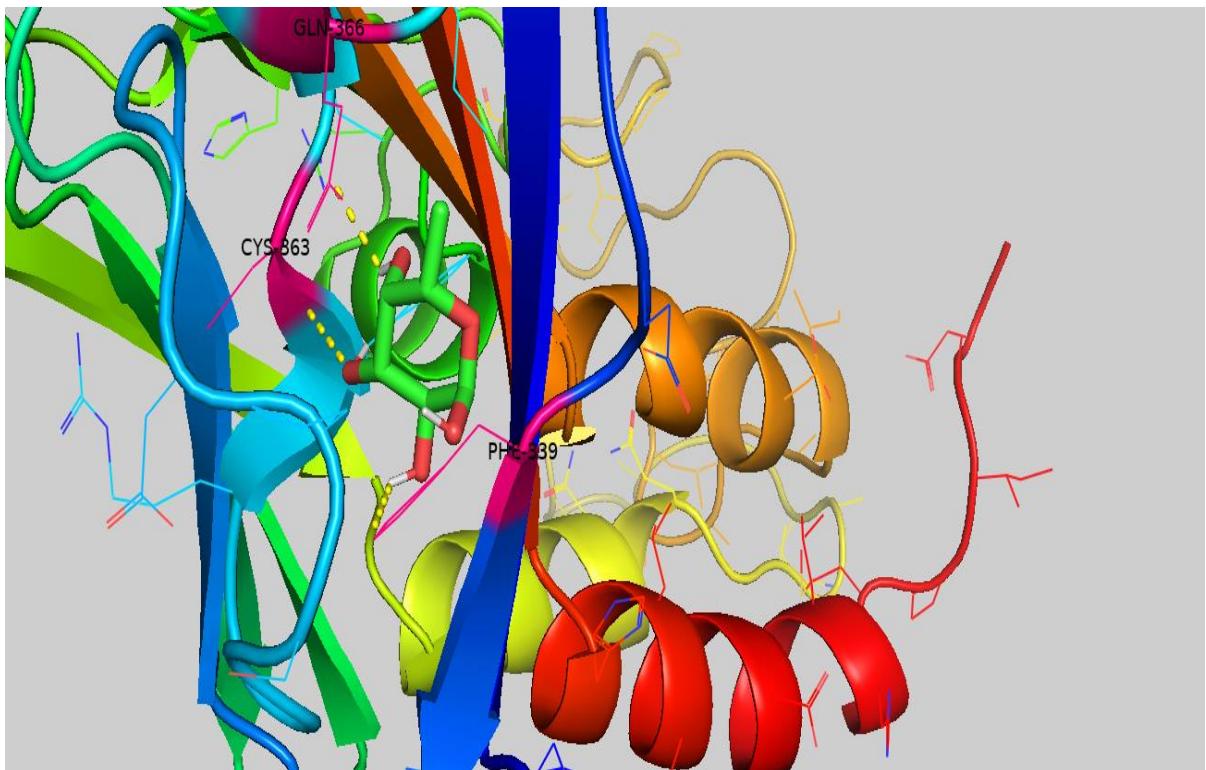
**Figure 18:** Docked R361C mutant SMAD4 with ALPHA L-FUCOSE



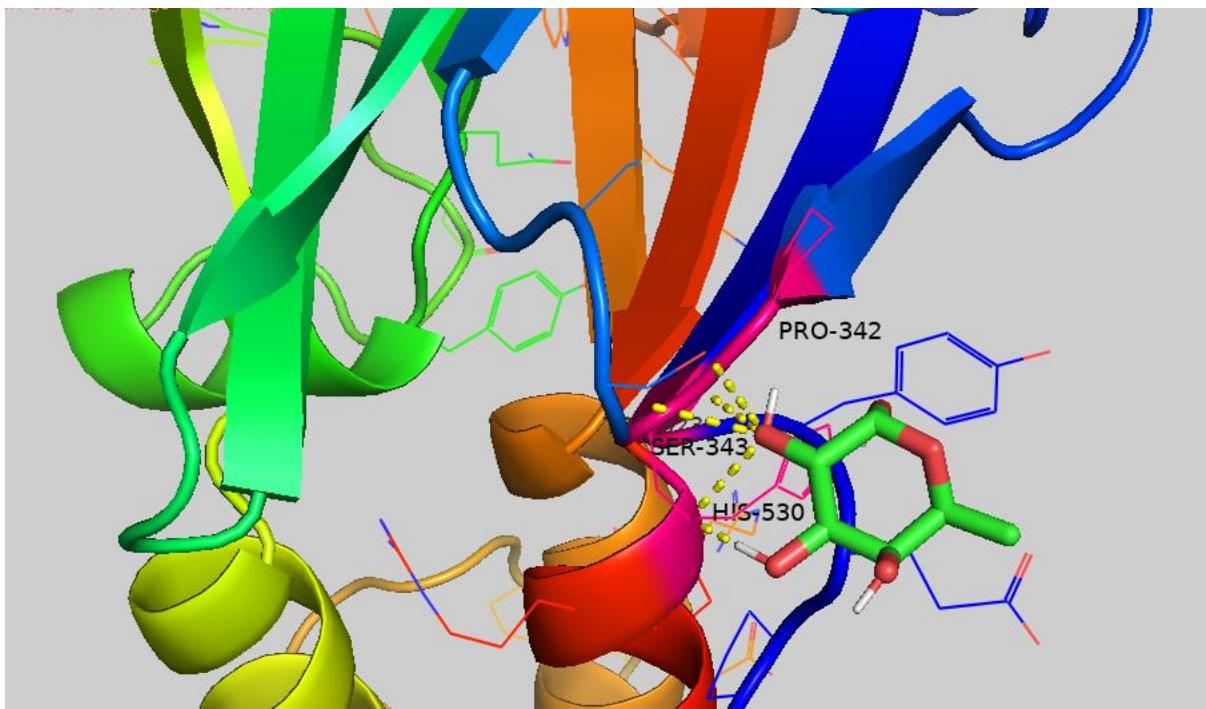
**Figure 19:** Docked Y353S mutant SMAD4 with ALPHA L-FUCOSE



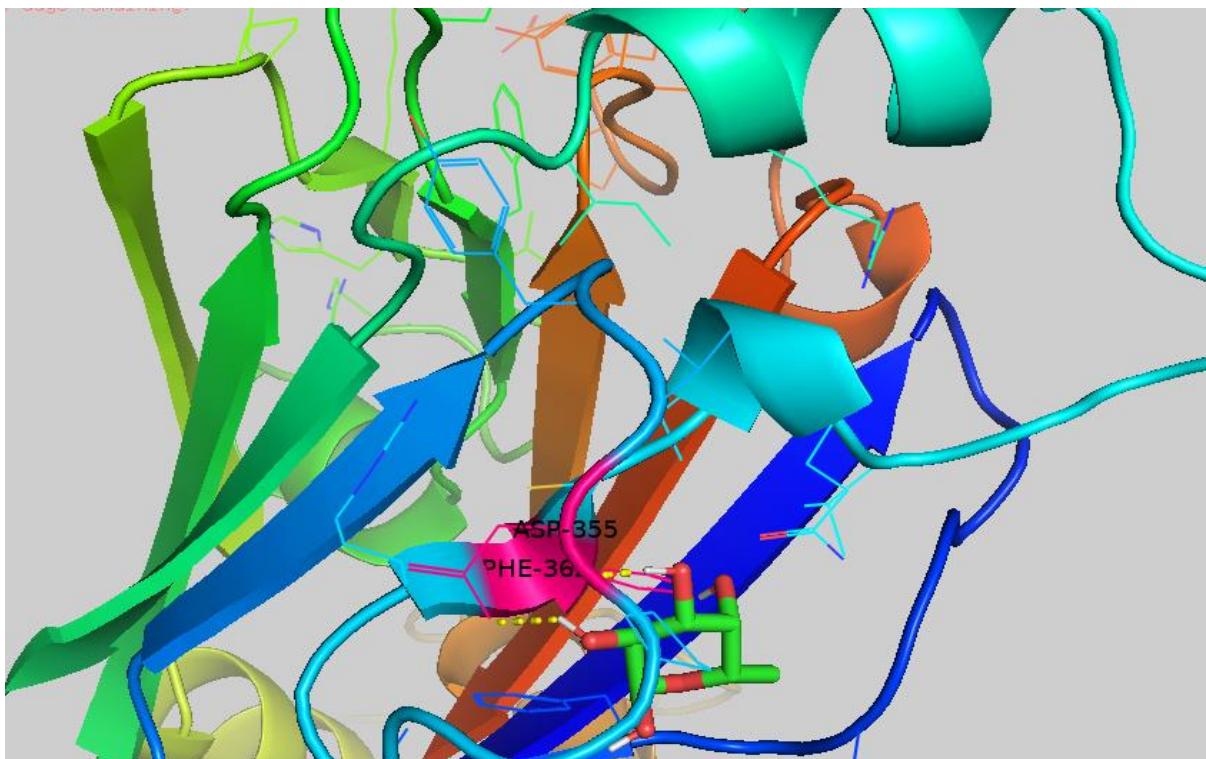
**Figure 20:** Docked C363R mutant SMAD4 with ALPHA L-FUCOSE



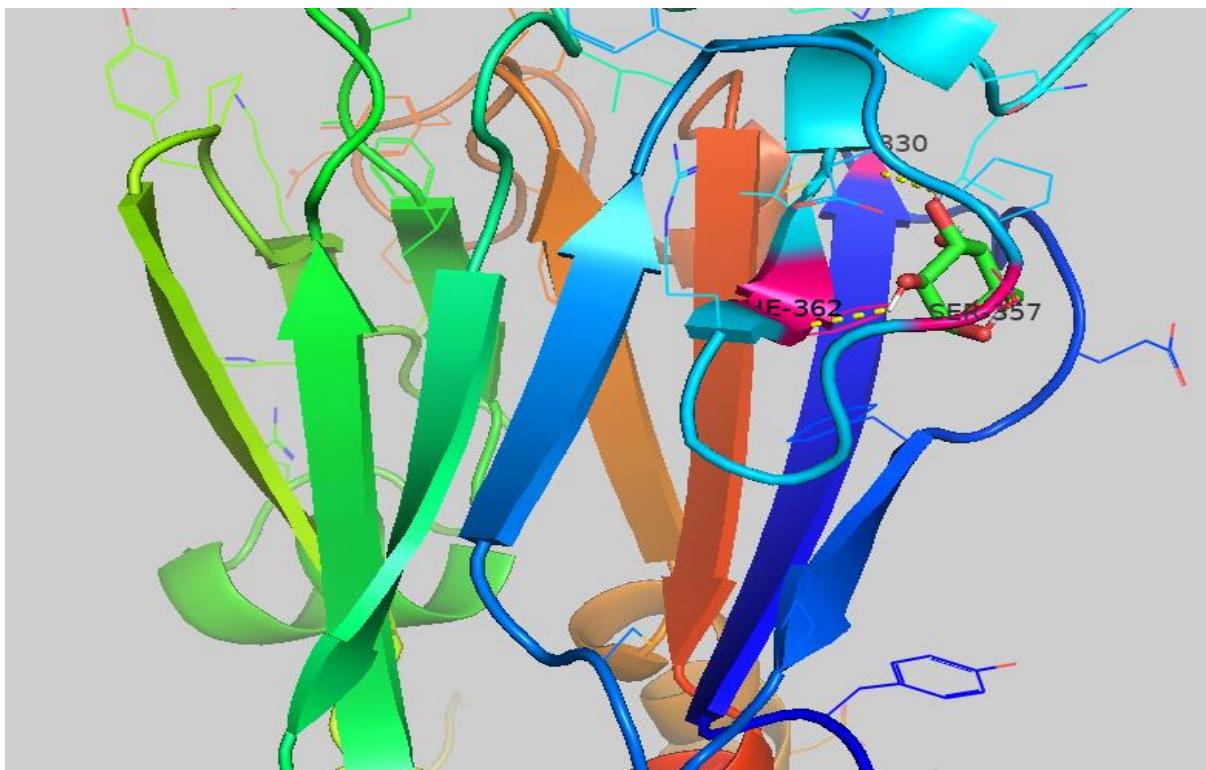
**Figure 21:** Docked G491V mutant SMAD4 with ALPHA L-FUCOSE



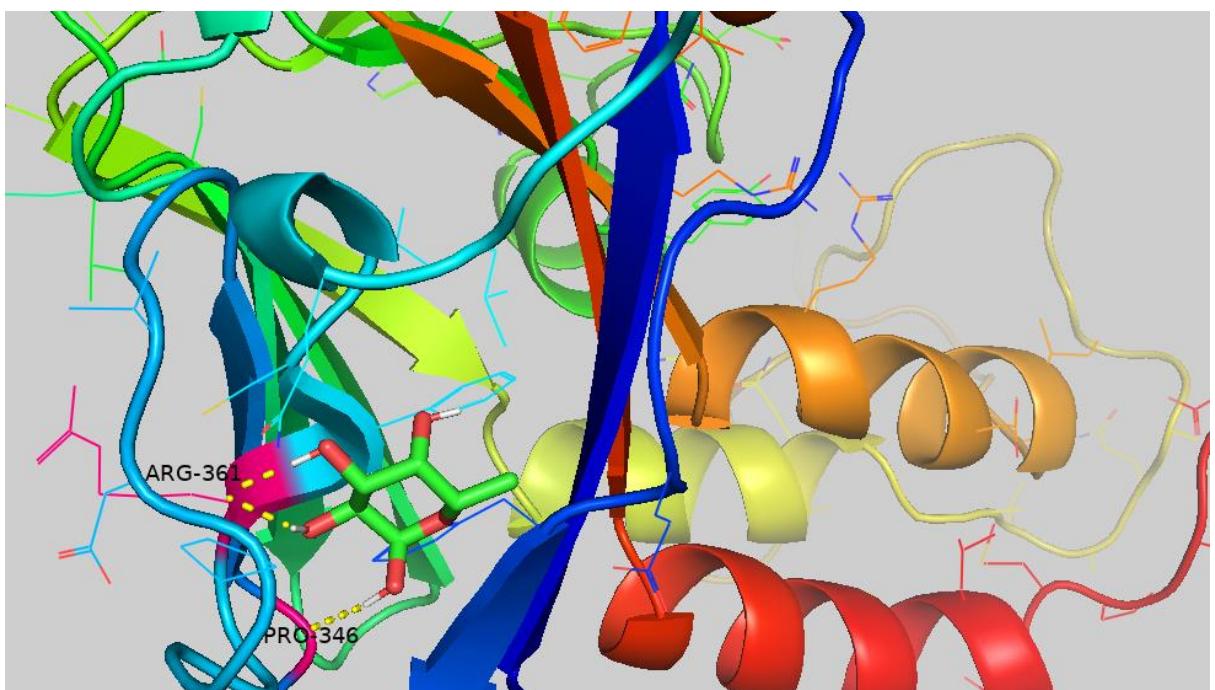
**Figure 22:** Docked L533P mutant SMAD4 with ALPHA L-FUCOSE



**Figure 23:** Docked L533R mutant SMAD4 with ALPHA L-FUCOSE



**Figure 24:** Docked W524L mutant SMAD4 with ALPHA L-FUCOSE



**Figure 25:** Docked G386D mutant SMAD4 with ALPHA L-FUCOSE

## 6.9 Prediction of post-translational modification site

All 23 nsSNPs were analysed with various tools to find PTM sites. Only one Y353S nsSNP has been predicted to undergo phosphorylation.

## **For erbB2 gene:**

### **6.10 SNP annotation**

We used the NCBI dbSNP database to find erbB2 SNPs, which included a total of 16,105 SNPs. In the coding sequence, there were 1104 non-synonymous SNPs and 609 synonymous SNPs. Only nsSNPs were chosen for further analysis in this study.

### **6.11 Identification of deleterious nsSNPs**

Utilization of 6 different in silico nsSNP prediction methods was done. In all computational techniques, 10 out of 1104 nsSNPs were projected to be harmful SNPs.

**Table 9:** Identification of deleterious nsSNPs by five *in silico* applications

SNP Id	AA Change	SIFT	POLYPHEN 2	PredictSNP	PANTHER	PROVEAN
rs28933368	E638K	DELETERIOUS	DAMAGING	DELETERIOUS	DAMAGING	DELETERIOUS
rs61737968	E49A	DELETERIOUS	DAMAGING	DELETERIOUS	POSSIBLY DAMAGING	DELETERIOUS
rs121913468	D493H	DELETERIOUS	DAMAGING	DELETERIOUS	DAMAGING	DELETERIOUS
rs121913470	L479S	DELETERIOUS	DAMAGING	DELETERIOUS	DAMAGING	DELETERIOUS
rs199530208	R223W	DELETERIOUS	DAMAGING	DELETERIOUS	DAMAGING	DELETERIOUS
rs367606199	R402W	DELETERIOUS	DAMAGING	DELETERIOUS	DAMAGING	DELETERIOUS
rs367606199	R156W	DELETERIOUS	DAMAGING	DELETERIOUS	DAMAGING	DELETERIOUS
rs375135008	R621W	DELETERIOUS	POSSIBLY DAMAGING	DELETERIOUS	DAMAGING	DELETERIOUS
rs375382055	R211W	DELETERIOUS	DAMAGING	DELETERIOUS	DAMAGING	DELETERIOUS
rs375637720	E598K	DELETERIOUS	DAMAGING	DELETERIOUS	DAMAGING	DELETERIOUS

### **6.13 Determination of protein structural stability**

2 out of the 5nsSNPs were shown to be stabilising the protein, resulting in protein dysfunction. These nsSNPs were further taken for analysis.

**Table 10:** CUPSAT predictions on the effect of nsSNPs on protein stability

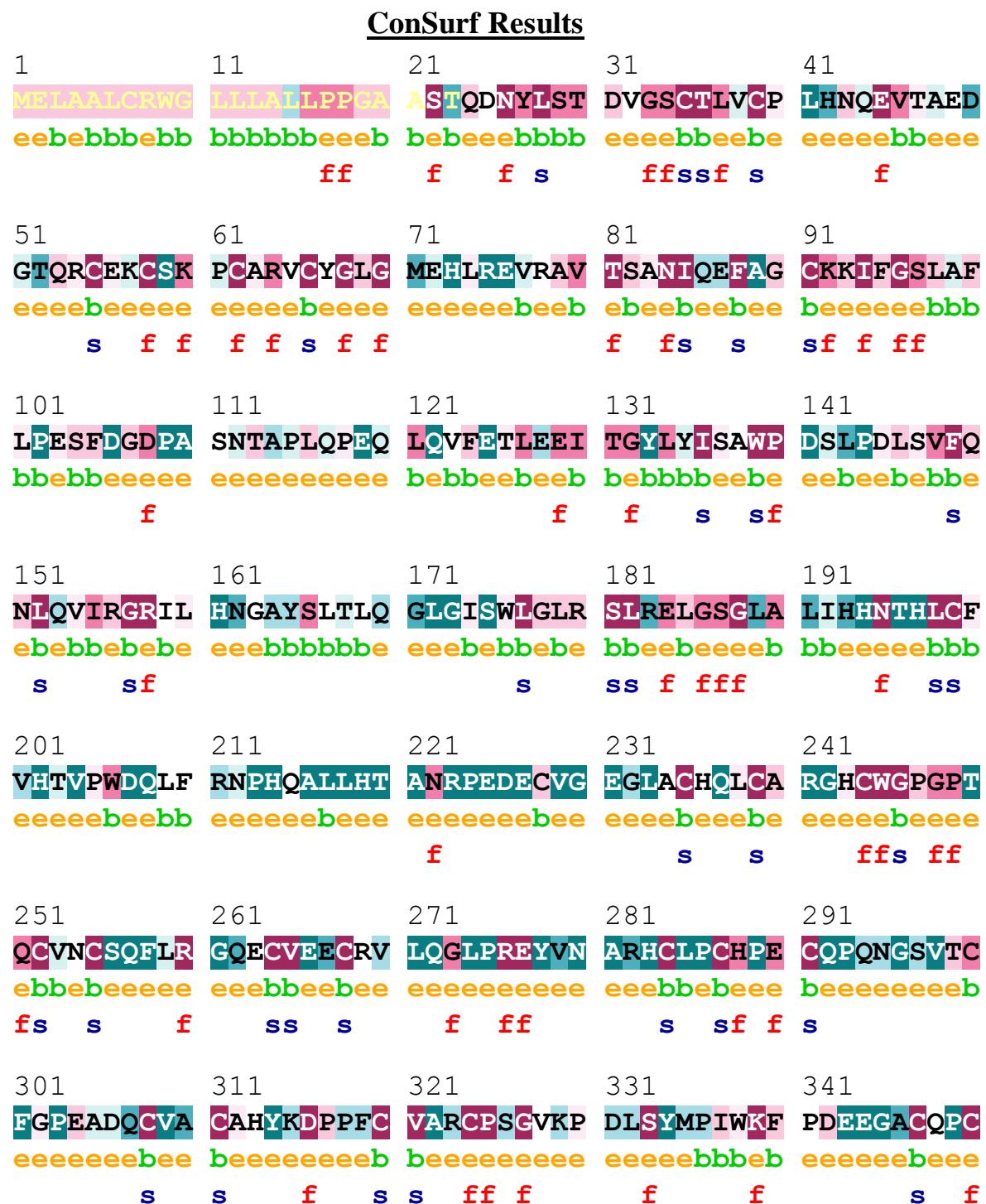
Amino acid change	Stabilising or destabilising	Predicted ( $\Delta\Delta G$ )
L479S	Destabilising	-9.7
D493H	Stabilising	0.59
E638K	Stabilising	0.77
E598K	Destabilising	-0.92
R621W	Destabilising	-2.91

### **6.12 Identification of nsSNPs on the domains of erbB2**

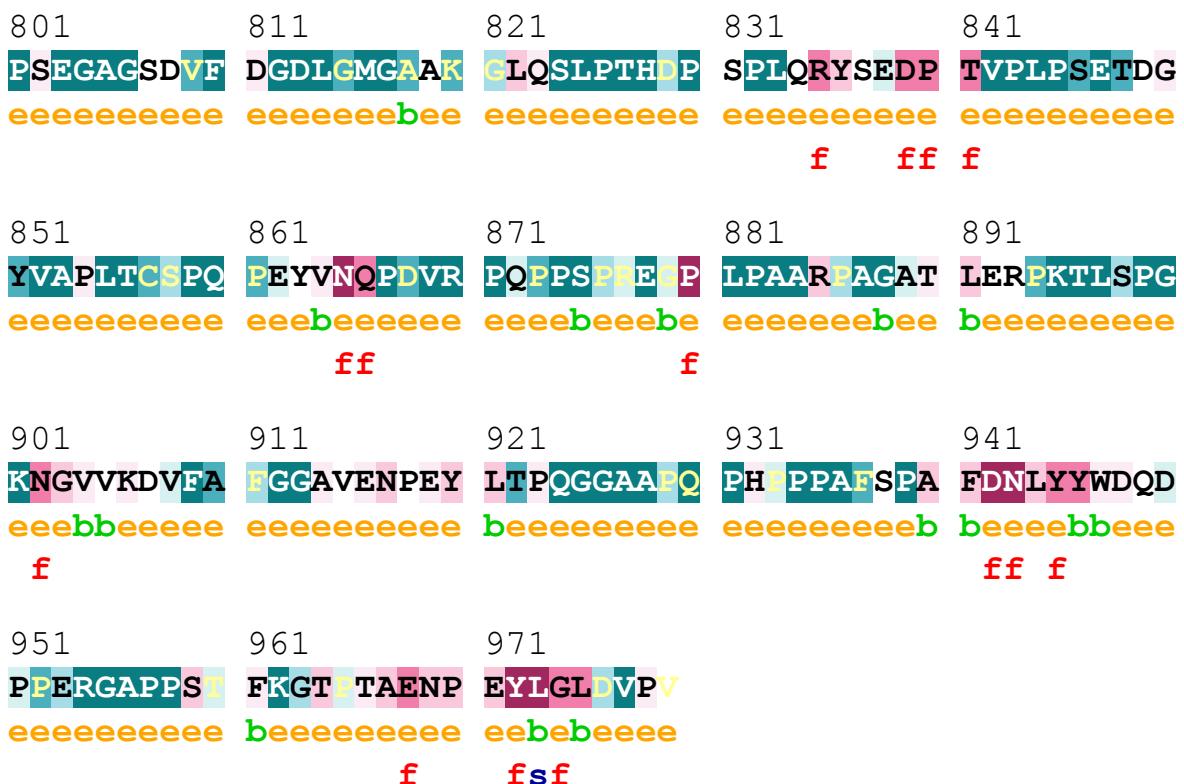
It found three functional domains of erbB2, which are Rcpt\_L-dom domain (90-208), GF\_recep\_IV domain (235-366) and Prot\_kinase\_dom domain (444-711) and demonstrated that all 6 Of 10 nsSNPs are positioned on these domains.

### 6.13 Evolutionary conservation analysis

Larger C324R, E330K, G352E, C363R, L364W, I383K, G491V, G510V, L533R, and G386D residues are more water repellent than wild type residues, according to the Project HOPE website, and these changes in size and hydrophobicity can disrupt H-bond interactions with neighbouring molecules.



351	361	371	381	391
<b>PINCTHSCVD</b>	<b>LDDKGCPAEQ</b>	<b>RASPLTSIIS</b>	<b>AVVGILLVVV</b>	<b>LGVVFGILIK</b>
eeeebeeeeeee	eeeeebeeeeee	eeeebbbbbb	bbbbbbbbbb	bbbbbbbbbb
<b>fsf f</b>				
401	411	421	431	441
<b>RRQQKIRKYT</b>	<b>MRRLLQETEL</b>	<b>VEPLTPSGAM</b>	<b>PNQAQMRLIK</b>	<b>ETELRKVKVL</b>
eeeeeeeeeb	beebbeeeee	beeeeeeeee	eeebbbebbe	eeeeeebebb
<b>ff f ff</b>	<b>ff ff f</b>	<b>ffffffff</b>	<b>fffs f sf</b>	<b>ff f f s</b>
451	461	471	481	491
<b>GSGAEGTVYK</b>	<b>GIWIPDGENV</b>	<b>KIPVAIKVLR</b>	<b>ENTSPKANKE</b>	<b>ILDEAYVMAG</b>
bbbbbbbbb	bbbbeeeeeb	ebebbbebbe	eeeeeebeee	bbeebbbb
<b>s sss s f s s ffff</b>	<b>fsfsssfss</b>	<b>f ff f f</b>		<b>ffs ss</b>
501	511	521	531	541
<b>VGSPYVSRLL</b>	<b>GICLTSTVQL</b>	<b>VTQLMPYGCL</b>	<b>LDHVRENRGR</b>	<b>LGSQDLLNWC</b>
bebabbbebb	bbbbeeeeb	bbebbbebbb	bebbeeeee	eebebbbebb
<b>fss sss f fs s sf s sss s f</b>				<b>sf ssfss</b>
551	561	571	581	591
<b>MQIAKGMSYL</b>	<b>EDVRLVHRDL</b>	<b>AARNVLVKSP</b>	<b>NHVKITDFGL</b>	<b>ARLLDIDETE</b>
bebabbbebb	eebebbbeb	ebeebbeeee	eebebeebbb	bebbeeeee
<b>sfsssfss ss f fssfs fsff sff</b>			<b>fsffss</b>	<b>sfss fff</b>
601	611	621	631	641
<b>YHADGGKVPI</b>	<b>KWMALESILR</b>	<b>RRFTHQSDVW</b>	<b>SYGVTVWELM</b>	<b>TFGAKPYDGI</b>
eeeeeeeb	ebbbbbb	eebeeeeebb	ebbbbbb	eebeeeeeb
<b>ff ff fsfs fsssssss</b>		<b>f fffffs</b>	<b>f sss sf s f f fff</b>	
651	661	671	681	691
<b>PAREIPDLLE</b>	<b>KGERLPQPPI</b>	<b>CTIDVYMIMV</b>	<b>KCWMIDSECR</b>	<b>PRFRELVSEF</b>
eebebebbe	eeeeeeeeeb	bbbbbbbbbb	ebbbbeeebe	eebebbbeeb
<b>sf sf ffffff s sss ssssss</b>			<b>fssssff f f s s fs</b>	
701	711	721	731	741
<b>SRMARDPQRF</b>	<b>VVIQNEDLGP</b>	<b>ASPLDSTFYR</b>	<b>SLLEDDDMGD</b>	<b>LVDAEYLVLP</b>
eebbeebeeb	bbbeeeeeeee	eeeeeeebbe	ebbeeeebee	bebebebebe
<b>ssffff f s f</b>				<b>f f f</b>
751	761	771	781	791
<b>QOGFFCPDPA</b>	<b>PGAGGMVHER</b>	<b>HRSSSTRSGG</b>	<b>CDDTLGLEPS</b>	<b>EEEAPRSPLA</b>
eeeebeeeee	eeeeeeeeee	eeeeeeeeee	eebebeeeee	eebebeeeee
			<b>f</b>	



**Figure 26:** CONSURF results

**Table 11:** Consurf investigated the evolutionary conservation of erbB2 amino acids.

SNP Id	Residue and Position	Conservation score	Prediction
rs28933368	E638K	9	Highly conserved and exposed(F)
rs121913468	D493H	8	Highly conserved and exposed(F)

#### 6.14 Structure analysis of wild type and mutant models

The normal and 10 mutations of the erbB2 protein were modelled in 3D using the Phyre2 tool. For the Rcept L-dom domain, we used the 1n8z template, and for the Prot kinase dom domain, we used the 5wn0 template. The two harmful nsSNPs were each modified in the native sequence of the templates, and 3D models for all of the mutants were anticipated. Using TM-scores and RMSD scores, Phyre2 was utilised to compare structural similarities between native and mutant models. In the mutant models, all of the nsSNPs in the Protein Kinase Domain exhibited a high RMSD value. A bigger RMSD value indicates greater structural dissimilarity between wild type and mutant models. SWISS MODEL was also utilised to investigate the 3D structures of two nsSNPs discovered on erbB2 protein domains.

**Table 12:** TM score and RMSD predictions for nsSNPs in erbB2

Serial No.	AA substitution	Domain	Tm Score	RMSD
1.	D493H	Prot_kinase_dom domain	0.92	1.722
2.	E638K		0.90	1.578

Note: We considered only those SNPs whose RMSD value is greater than 1.

**Table 13:** Various parameters of 2 nsSNPs investigated by Swiss Model

SNPs	QMEN	Cb	All atom	Solvation	Torsion	Template
Wild Type	-1.97	-2.10	-2.20	0.09	-1.64	c5wnoA
D493H	-1.95	-1.84	-2.08	-0.00	-1.64	c5wnoA
E638K	-1.97	-1.78	-2.16	-0.05	-1.66	c5wnoA

### 6.15 Structural effect of point mutation on erbB2 protein

D493H and E638K with larger sizes are more water-repellent than wild type residues, according to the Project HOPE server, and these differences in size and hydrophobicity can disrupt H-bonds within the structure.

**Table 14:** Structural effect of 5 nsSNPs over erbB2 protein using Project Hope

Residue	Structure	Properties
D493H		<p>The mutant residue is nearly twice as large as the wild-type residue.</p> <p>The wild-type residue had a NEGATIVE charge, but the mutant residue had a NEUTRAL charge.</p>
E638K		<p>The mutant residue is more than double the size of the wild-type residue.</p> <p>The charge of the wild-type residue was NEGATIVE, while the charge of the mutant residue is NEUTRAL.</p>

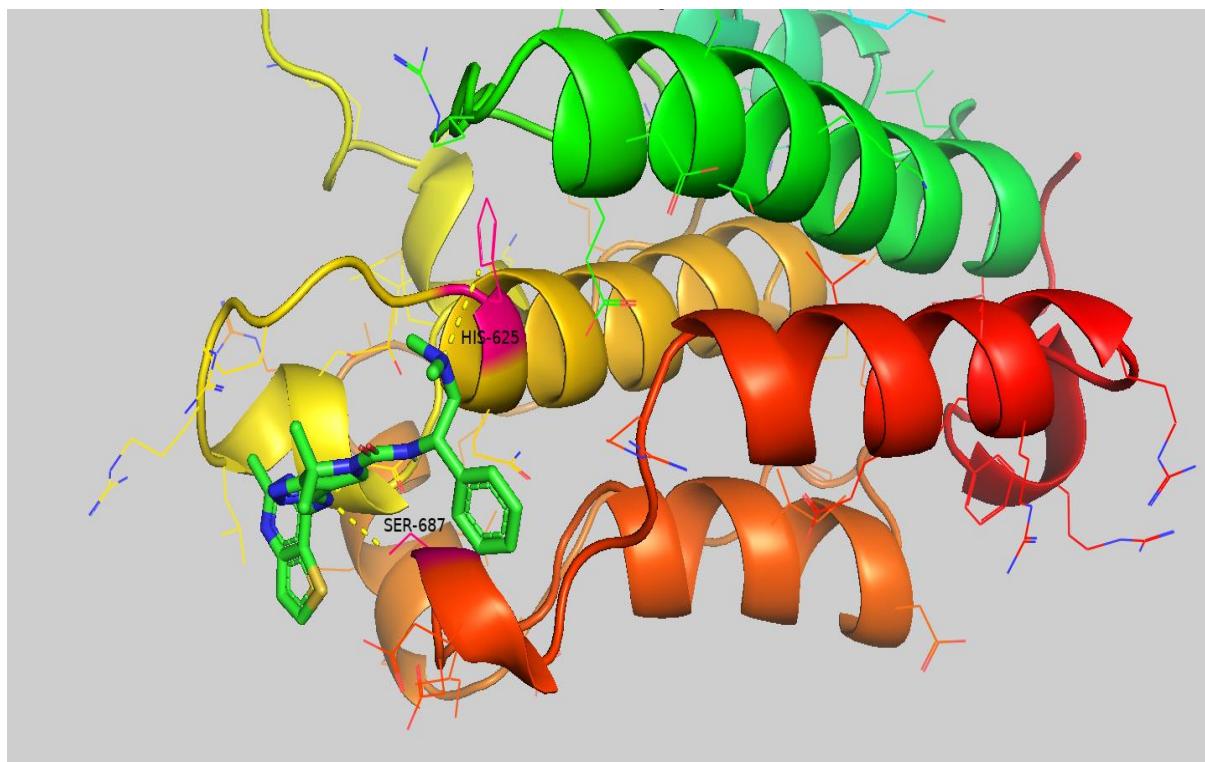
### 6.18 Molecular docking analysis

The binding affinity of the altered erbB2 with other proteins was represented by molecular docking studies(3E) -N-[(1S)-2-(dimethylamine)-1-phenylethyl] 6-[(2-

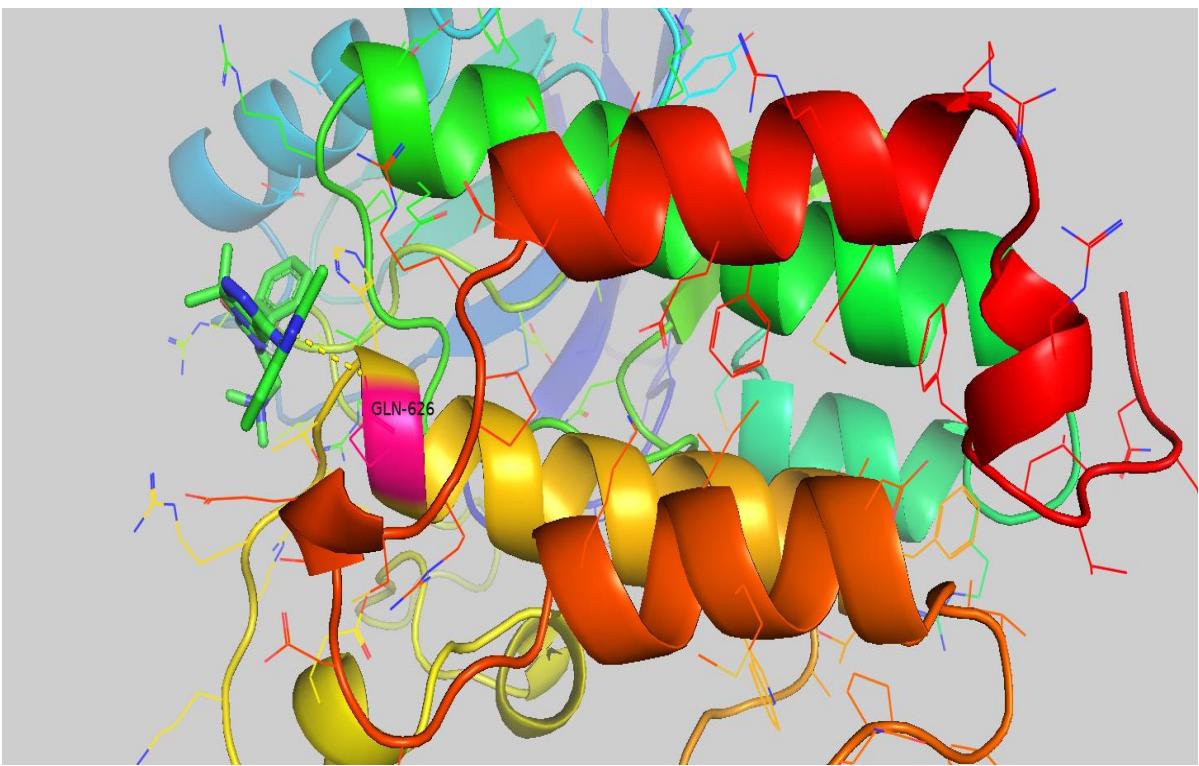
methylthieno[3,2-d] pyrimidin-4-yl) amino]-6,6-dimethyl-3-[(2-methylthieno[3,2-d] pyrimidin-4-yl)-1,4-dihdropyrrolo[3,4-c] pyrazole-5-carboxamide(X4Z). D493H and E638K were found to increase the binding affinity with X4Z. PYMOL software was used to investigate the binding affinity (kcal/mol) and bonding interaction patterns of these docked complexes. With X4Z, the E638K variation revealed the greatest increase in binding affinity. E638K establishes one hydrogen bond with X4Z, increasing the binding affinity to 7.1 kcal/mol. D493H enhanced the binding affinity to 6.7 kcal/mol by forming one hydrogen bond with the ligand.

**Table 15:** Docking with X4Z

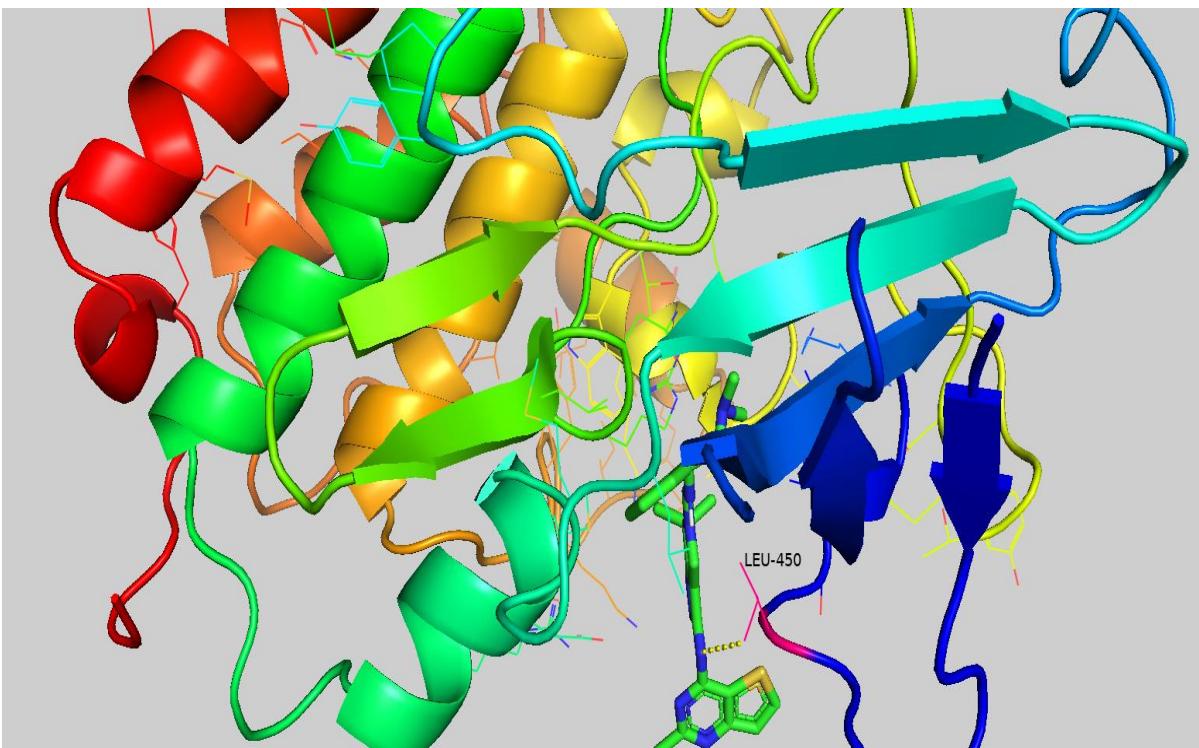
Amino Acid Change Mutant Model	Binding Affinity (Kcal/Mol)
WILD	-6.1
D493H	-6.7
E638K	-7.1



**Figure 27:** Docked wild erbB2 with X4Z



**Figure 28:** Docked D493H MUTANTerbB2 with X4Z



**Figure 29:** Docked E368K mutant erbB2 with X4Z

## **7. DISCUSSION**

SMAD4 acts as a tumour suppressor in a range of human malignant tumours (Liu et al., 2015). Colorectal cancer (Pai et al., 2013), head and neck cancer (Lin et al., 2019), and other disorders have been linked to SMAD4 suppression. During biomolecular interactions, differences in the structural conformation of the SMAD4 protein are crucial for its function; nevertheless, nsSNPs may create aberrant conformations, resulting in the loss of its tumour suppressive properties. (Papageorgis et al., 2012). As a result, it is critical to determine the consequences of SMAD4 detrimental nsSNPs and, as a result, their relationship with various disorders. We used computational tools to identify the most harmful nsSNPs and their consequences for the structure and function of the We used various tools to screen out 23 highly harmful nsSNPs from 477 nsSNPs reported in the NCBI database. The prediction scores provided by these six methods were used to choose these 23 harmful nsSNPs. The InterPro tool was then used to locate these nsSNPs on various SMAD4 domains. All of the nsSNPs were discovered to be on the MH2 domain, which induces cell death (Zhao et al., 2018). For structural and functional activity, protein stability is necessary (Deller et al., 2016). Using the I-Mutant software, we found harmful nsSNPs that could affect SMAD4 stability. As a result, we concentrated on the effects of the 23 nsSNPs on SMAD4 protein stability. The protein's stability was lowered by all nsSNPs. Protein stability determines the protein's conformational structure and hence its activity. Misfolding, disintegration, and aberrant protein aggregation can all be caused by changes in protein stability (Hossain et al., 2020). Furthermore, evaluating whether a mutation is deleterious to the host requires evolutionary conservation of the protein sequence. We discovered that highly harmful nsSNPs with high conservation scores are located in highly conserved areas, increasing the likelihood of cancer by inactivating SMAD4. The work will continue with the use of the Phyre2 homology modelling programme to examine the structural consequences of these detrimental nsSNPs. Using the templates, we constructed wild type and mutant SMAD4 protein models. The RMSD and TM score of these Phyre2-generated wildtype and mutant protein models were also predicted. A lower TM score and a higher RMSD imply that the mutant protein structure is different from the original (Hossain et al., 2020). Based on these criteria, we selected 23 harmful nsSNPs in the MH2 domain.

Swiss Model gave the QMEN, solvation, and torsion scores for our 23 nsSNPs. These 23 extremely harmful nsSNPs, according to Consurf, affect the structure of the protein, with 12 nsSNPs being structural and 11 nsSNPs being functional residues. The Missense 3D programme visualised the impacts of the 23 structural nsSNPs, indicating that ten of them have a negative influence on the SMAD4 protein's structural shape. Ten nsSNPs out of 23 were predicted to stabilise the protein. Six of the thirteen destabilising proteins displayed structural damage and were likewise destabilised by CUPSAT. Only five of the 10 nsSNPs were anticipated to cause structural damage. The ligand was docked with the 10 nsSNPs that were stabilised. Three variations (R361C, Y353S, I383K, W509R, W524L, L533R, and G386D) improved the binding affinity with Alpha-1-fucose when compared to wild type residues, according to docking research. The most notable reduction in binding affinity was identified in R361C, where the binding pocket showed a substantial loss of H-bond interactions. In summary, molecular docking research demonstrated that the aforementioned variations can have a considerable impact on SMAD4 protein functional activity. Only Y353S was predicted to be phosphorylated by NETPHOS 3.1, out of all nsSNPs. As a result, our research will establish a solid foundation for identifying functional SNPs.

The first persistent genetic mutation revealed in breast cancer was ERBB2 amplification (Hynes, 2016). Variations in structural conformation of the ERBB2 protein during biomolecular interactions are critical for its function to be carried out, however, nsSNPs may cause abnormal conformations, resulting in the loss of its tumour suppressive properties (Papageorgis et al., 2012). As a result, it is critical to determine the impact of detrimental ERBB2 nsSNPs and their association with a variety of illnesses. The most detrimental nsSNPs and their implications for the structure and function of the erbB2 protein were identified using computational analysis. We screened out 10 significantly harmful nsSNPs from 1104 nsSNPs detected in the NCBI database using five in silico SNP prediction tools. The prediction scores provided by these six methods were used to choose these 10 harmful nsSNPs. The position of these nsSNPs on different ERBB2 domains was also determined using the InterPro programme. It was discovered that six of them were located on apoptosis-related domains (Zhao et al., 2018). Protein stability is necessary for its structure and functional activity. (Deller et al., 2016). We discovered two harmful nsSNPs in the ERBB2 gene using the CUPSAT programme. As a result, we concentrated on the effects of the two erbB2 protein-stabilizing nsSNPs indicated above. Protein stability controls the protein's conformational structure and

hence its function. Misfolding, breakdown, or abnormal protein aggregation can all be caused by changes in protein stability (Hossain et al., 2020). Evolutionary conservation in the protein sequence is critical for determining if a mutation is harmful to the host. We discovered that extremely damaging nsSNPs with high conservation scores were found in highly conserved locations, indicating that inactivating ERBB2 increases the risk of cancer. The researchers will utilise the Phyre2 homology modelling programme to investigate the structural repercussions of these harmful nsSNPs. The templates were used to build wild type and mutant ERBB2 protein models. The RMSD and TM score of these Phyre2-generated wildtype and mutant protein models were also calculated. A greater RMSD and a lower TM score indicate that the mutant protein structure is more different from the natural one (Hossain et al., 2020). We chose two harmful nsSNPs in the Protein Kinase domain based on these criteria. For our targeted 2 nsSNPs, Swiss Model supplied QMEN, solvation, and torsion scores. These two particularly dangerous nsSNPs, according to Consurf, have a negative impact on protein structure. SNPs were docked with the ligand. When compared to the wild type residues, each of these 2nsSNPs significantly improved the binding affinity of X4Z, according to docking studies. The binding affinity was reduced the highest in E638K, where the binding pocket revealed a significant decrease of H-bond interactions. In conclusion, molecular docking research demonstrated that the aforementioned variations can have a considerable impact on erbB2 protein functional activity. As a result, this research will give a useful framework for identifying functional SNPs.

## **8.CONSLUSIONS**

The protein ERBB2 has anti-tumor properties. The MH2 domain is one of the protein's functional domains. As a result, the structural conformation of this domain is critical for it to perform its function. This in silico investigation of the protein's functional SNPs reveals the potential negative impact of nsSNPs on ERBB2. The consequences of nsSNPs on the structure and function of the ERBB2 protein have never been predicted before. Our findings demonstrate that stabilising mutations in the MH2 domain R361C, Y353S, I383K, W509R, W524L, L533R, and G386D increase binding affinity, with R361C showing the most significant drop. Using the missense 3D tool, only six of the 13 destabilising nsSNPs were found to cause structural damage. Because this protein has been linked to a variety of cancers, our findings will be useful in the development of prospective diagnostic and therapeutic approaches, which will include experimental validation of variants and large-scale clinical trials.

Amplification of the erbB2 protein results in a variety of malignancies. The Protein Kinase domain is one of the functional domains of this protein. As a result, the structural shape of this domain is important for it to fulfil its job. This in silico analysis of functional SNPs in erbB2 gives light on the protein's potential for damage. This is the first study to predict how nsSNPs affect erbB2 protein structure and function. According to our findings, three nsSNPs in the MH2 domain (E638K, D493H) diminish binding affinity, with E638K having the greatest effect. Because this protein has been linked to a variety of cancers, our findings will be useful in the development of prospective diagnostic and prognostic techniques that will include both experimental validation of variants and large-scale clinical trials.

## **9. FUTURE PROSPECTIVES**

SNP analysis in cancer research has a plethora of clinical, public health, and cancer biology consequences. New discoveries could lead to more focused cancer treatment and preventive options. The early exhilaration of using genetic diversity to treat cancer has given way to the realisation that the new era of genomic medicine will face a complicated set of problems that will test the scientific, clinical, and social adaptations. There are already examples in pharmacogenomics and cancer susceptibility that demonstrate the relevance of this discipline, but they are only the beginning. In the coming decade, genetic variation may be investigated and implemented in cancer diagnosis, prevention, and treatment.

SNP analysis holds the promise of providing insights into exposure and cancer, as well as laying the groundwork for lifestyle (i.e., dietary, exercise, and weight control) and chemoprevention interventions. Secondary prevention, on the other hand, will focus on screening and public health initiatives aimed at identifying and intervening in high-risk people. As information about pharmacogenomics, second malignancies, and long-term problems is integrated, the latter will become more problematic. This is complicated further by the fact that the effects of one group of SNPs may be detrimental to one outcome while being beneficial to another. The development of more advanced analytical tools to deal with the growing data will be a big problem in the future. We must address the complex issue of gene–gene interactions if SNP profiles are to be used in clinical settings. Finally, putting what we've learned from SNP studies into practise will almost definitely entail a significant shift in practise paradigm. Practitioners could make treatment decisions or offer advice based on a risk assessment. To begin addressing the key issues of SNPs, haplotypes, and cancer in different populations, a thorough understanding of the extent and structure of genetic variation, particularly SNPs, haplotypes, and linkage disequilibrium for cancer-related genes and pathways, is required.

## REFERENCES

- Adzhubei, I., Jordan, D.M. and Sunyaev, S.R., 2013. Predicting functional effect of human missense mutations using PolyPhen-2. *Current protocols in human genetics*, 76(1), pp.7-20.
- Ahmad, T., Valentovic, M.A. and Rankin, G.O., 2018. Effects of cytochrome P450 single nucleotide polymorphisms on methadone metabolism and pharmacodynamics. *Biochemical pharmacology*, 153, pp.196-204.
- Aretz, S., Stienen, D., Uhlhaas, S., Stolte, M., Entius, M.M., Loff, S., Back, W., Kaufmann, A., Keller, K.M., Blaas, S.H. and Siebert, R., 2007. High proportion of large genomic deletions and a genotype–phenotype update in 80 unrelated families with juvenile polyposis syndrome. *Journal of medical genetics*, 44(11), pp.702-709.
- Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T. and Ben-Tal, N., 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic acids research*, 44(W1), pp. W344-W350.
- Bendl, J., Stourac, J., Salanda, O., Pavelka, A., Wieben, E.D., Zendulka, J., Brezovsky, J. and Damborsky, J., 2014. PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS computational biology*, 10(1), p.e1003440.
- Capriotti, E., Calabrese, R., Fariselli, P., Martelli, P.L., Altman, R.B. and Casadio, R., 2013. WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC genomics*, 14(3), pp.1-7.
- Choi, Y. and Chan, A.P., 2015. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*, 31(16), pp.2745-2747.
- Choi, Y., Sims, G.E., Murphy, S., Miller, J.R. and Chan, A.P., 2012. Predicting the functional effect of amino acid substitutions and indels.
- Deller, M.C., Kong, L. and Rupp, B., 2016. Protein stability: a crystallographer's perspective. *Acta Crystallographica Section F: Structural Biology Communications*, 72(2), pp.72-95.
- Erichsen, H.C. and Chanock, S.J., 2004. SNPs in cancer research and treatment. *British journal of cancer*, 90(4), pp.747-751.

- Hunter, S., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L. and Finn, R.D., 2009. InterPro: the integrative protein signature database. *Nucleic acids research*, 37(suppl\_1), pp. D211-D215.
- Hynes, N.E., 2016. ErbB2: from an EGFR relative to a central target for cancer therapy. *Cancer research*, 76(13), pp.3659-3662.
- Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. and Sternberg, M.J., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols*, 10(6), pp.845-858.
- Kitts, A. and Sherry, S., 2011. The single nucleotide polymorphism database (dbSNP) of nucleotide sequence variation. The NCBI handbook. *Bethesda, MD: US National Center for Biotechnology Information*.
- Lin, L.H., Chang, K.W., Cheng, H.W. and Liu, C.J., 2019. ERBB2 somatic mutations in head and neck carcinoma are associated with tumor progression. *Frontiers in oncology*, p.1379.
- Liu, Y., Sheng, J., Dai, D., Liu, T. and Qi, F., 2015. ErbB2 acts as tumor suppressor by antagonizing lymphangiogenesis in colorectal cancer. *Pathology-Research and Practice*, 211(4), pp.286-292.
- Mafficini, A., Brosens, L.A., Piredda, M.L., Conti, C., Mattiolo, P., Turri, G., Mastrosimini, M.G., Cingarlini, S., Crinò, S.F., Fassan, M. and Piccoli, P., 2022. Juvenile polyposis diagnosed with an integrated histological, immunohistochemical and molecular approach identifying new ERBB2 pathogenic variants. *Familial Cancer*, pp.1-11.
- Matsuda, K., 2017. PCR-based detection methods for single-nucleotide polymorphism or mutation: real-time PCR and its substantial contribution toward technological refinement. *Advances in clinical chemistry*, 80, pp.45-72.
- Mi, H., Ebert, D., Muruganujan, A., Mills, C., Albou, L.P., Mushayamaha, T. and Thomas, P.D., 2021. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic acids research*, 49(D1), pp.D394-D403.
- Pai, J., Kozak, M.M., Limaye, M.R., Jayachandran, P., Anderson, E.M., Shaffer, J.L., Longacre, T.A., Pai, R.K., Koong, A.C. and Chang, D.T., 2013. ERBB2 Inactivation and Prognosis in Colorectal Cancer. *International Journal of Radiation Oncology, Biology, Physics*, 87(2), p.S342.

- Papageorgis, P., Cheng, K., Ozturk, S., Gong, Y., Lambert, A.W., Abdolmaleky, H.M., Zhou, J.R. and Thiagalingam, S., 2011. ErbB2 inactivation promotes malignancy and drug resistance of colon cancer. *Cancer research*, 71(3), pp.998-1008.
- Parthiban, V., Gromiha, M.M. and Schomburg, D., 2006. CUPSAT: prediction of protein stability upon point mutations. *Nucleic acids research*, 34(suppl\_2), pp.W239-W242.
- Schwede, T., Kopp, J., Guex, N. and Peitsch, M.C., 2003. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic acids research*, 31(13), pp.3381-3385.
- Srinivasan, S., Clements, J.A. and Batra, J., 2016. Single nucleotide polymorphisms in clinics: fantasy or reality for cancer? *Critical Reviews in Clinical Laboratory Sciences*, 53(1), pp.29-39.
- Trott, O. and Olson, A.J., 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31(2), pp.455-461.
- Vallejos-Vidal, E., Reyes-Cerpa, S., Rivas-Pardo, J.A., Maisey, K., Yáñez, J.M., Valenzuela, H., Cea, P.A., Castro-Fernandez, V., Tort, L., Sandino, A.M. and Imarai, M., 2020. Single-nucleotide polymorphisms (SNP) mining and their effect on the tridimensional protein structure prediction in a set of immunity-related expressed sequence tags (EST) in Atlantic salmon (*Salmo salar*). *Frontiers in Genetics*, 10, p.1406.
- Vaser, R., Adusumalli, S., Leng, S.N., Sikic, M. and Ng, P.C., 2016. SIFT missense predictions for genomes. *Nature protocols*, 11(1), pp.1-9.
- Venselaar, H., Te Beek, T.A., Kuipers, R.K., Hekkelman, M.L. and Vriend, G., 2010. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC bioinformatics*, 11(1), pp.1-10.
- Vignal, A., Milan, D., SanCristobal, M. and Eggen, A., 2002. A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution*, 34(3), pp.275-305.
- Witham, S., Takano, K., Schwartz, C. and Alexov, E., 2011. A missense mutation in CLIC2 associated with intellectual disability is predicted by in silico modeling to affect protein stability and dynamics. *Proteins: Structure, Function, and Bioinformatics*, 79(8), pp.2444-2454.

- Woodford-Richens, K.L., Rowan, A.J., Gorman, P., Halford, S., Bicknell, D.C., Wasan, H.S., Roylance, R.R., Bodmer, W.F. and Tomlinson, I.P.M., 2001. ERBB2 mutations in colorectal cancer probably occur before chromosomal instability, but after divergence of the microsatellite instability pathway. *Proceedings of the National Academy of Sciences*, 98(17), pp. 9719-9723.
- Yingling, J.M., Datto, M.B., Wong, C., Frederick, J.P., Liberati, N.T. and Wang, X.F., 1997. Tumor suppressor ErbB2 is a transforming growth factor beta-inducible DNA binding protein. *Molecular and cellular biology*, 17(12), pp.7019-7028.
- Yuan, S., Chan, H.S. and Hu, Z., 2017. Using PyMOL as a platform for computational drug design. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 7(2), p.e1298.
- Zhao, M., Mishra, L. and Deng, C.X., 2018. The role of TGF- $\beta$ /ERBB2 signaling in cancer. *International journal of biological sciences*, 14(2), p.111.