T06307

IN SILICO MUTATION ANALYSIS OF LAH2 & LAH3

Researcher:
Amara Jabeen
Reg # 02/FBAS/MSBI/F07

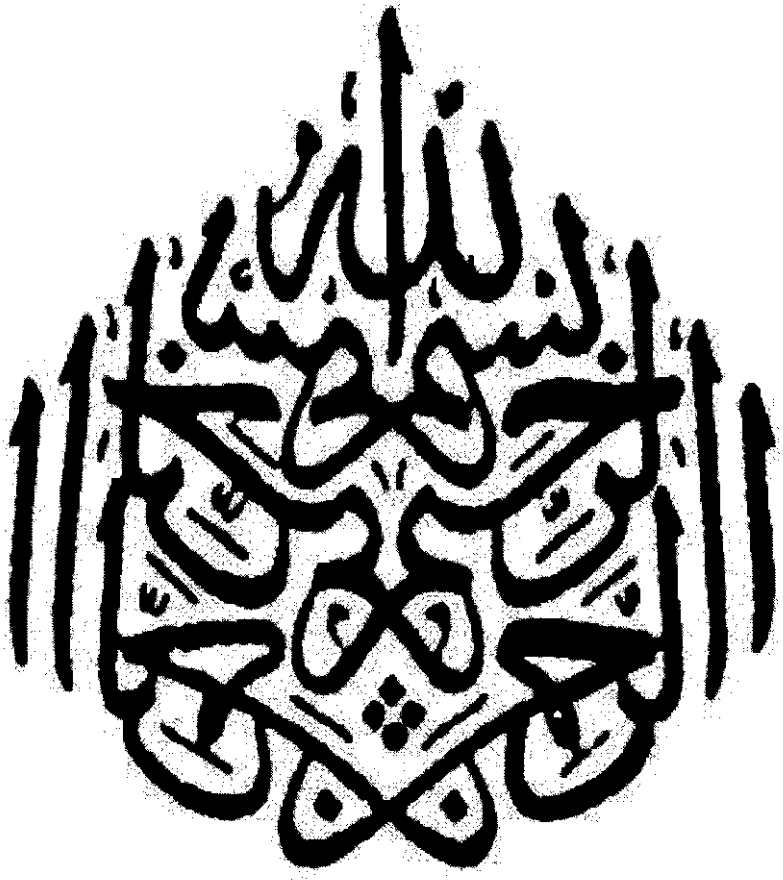
Supervisor:
Dr. Naveeda Riaz

Department of Bioinformatics &Biotechnology
Faculty of Basic and Applied Sciences
International Islamic university, Islamabad

Molecular biology
cytology
Genetics
Skeletal muscle, Sarcoglycanopathies

IN SILICO MUTATION ANALYSIS OF LAH2 & LAH3

Amara Jabeen
Reg #02/FBAS/MSBI/F07



DEDICATION

To Almighty ALLAH and the Holy Prophet

Muhammad (P.B.U.H)

&

My Family

CERTIFICATE

Title of Thesis: In Silico Mutation Analysis of LAH2 and LAH3

Name of Student: Amara Jabeen

Registration No: 02/FBAS/MSBI/F07

Accepted by the Department of Bioinformatics, Faculty of Basic and Applied Sciences, International Islamic University Islamabad, in partial fulfillment of the requirements for the Master of Sciences in Bioinformatics.

Viva Voce Committee

Chairman/Director/Head **(Dr. Naveeda Riaz)**

Supervisor **(Dr. Naveeda Riaz)**

Internal Examiner **(Dr. Asma Gul)**

External Examiner **(Dr. Asif Mir)**

(27th August, 2009)

ABSTRACT

Autosomal recessive hypotrichosis is a genetic hair disorder that is though not life threatening but it can lead to abhorrent effect on person's psyche. In order to develop a drug for autosomal recessive hypotrichosis, mechanism underlying this disease must be studied. There are total 3 forms of Autosomal recessive hypotrichosis that are LAH1, LAH2 and LAH3. Forms pertaining in Pakistan are LAH3 and LAH2. LAH3 is caused by Mutations in P2RY5 and LAH2 is caused by mutations in *LIPH*. Product of *LIPH* is LPA that binds with *P2RY5* to control hair growth. Current study encompasses broad way analysis of alterations brings by mutations in these genes at molecular level through bioinformatics tools. Bioinformatics tools are the most attractive option for a developing country like Pakistan that is full of genetic resources and offer a plenty of room for biological research that is already being restricted due to lack of financial resources. Analysis of effect of reported mutations on domains, motifs, pots-translational modifications, secondary structure, tertiary structure and physiochemical properties revealed that missense mutations effect motif and domains of these proteins that leads to change in functionality. Due to change in composition of protein, secondary structure that is formed got changed and leads to formation of altered tertiary structure. Domains that are added as a result of mutations, adopts independent 3D structure conformation that interrupts normal functioning. As a result of mutation an extra site of G_PROTEIN_RECEP_F1_1 receptor, signal peptide and transmembrane regions are added in *P2RY5*. Mutations in *LIPH* results in deletion of a motif also. Kin bind and Ykin motifs are also added in *P2RY5* as a result of mutations. Mutations in *P2RY5* and *LIPH* also lead to altered pattern of glycosylation and phosphorylation. Number of alpha helices and beta sheets also changes in *LIPH* and *P2RY5* as a consequence of mutations that result in altered secondary structures. During current study tertiary structure of proteins were predicted through homology modeling and threading approach. Mutated structures revealed the difference that is a consequence of mutations. Interactions pathway of both proteins is also predicted during current study by utilizing bioinformatics tools and databases. Valuable information that is stored in bulk amount in databases must be utilized instead of re inventing the wheel to accelerate biological research that has direct impact on betterment of mankind. In conclusion missense mutations in *LIPH*

and *P2RY5* brought drastic changings at molecular level that leads to imbalanced function as a consequence of which autosomal recessive hypotrichosis is caused and all the changings can be studied through bioinformatics without going for expensive laborious and time consuming experimental techniques. Target for LAH2 is *LIPH*. It can be targeted easily as knowledge of proteins interacting with it is also available and there is knowledge of consequences of mutations as *LIPH* is fully modeled in this study.

ACKNOWLEDGEMENT

First of all I am heartily thankful to ALLAH Almighty Who has given me the ability to apply the learned knowledge practically. During my thesis I interacted with many people whose support in assorted ways to my thesis deserved special mention. It is a pleasure to convey my gratitude also to them in my humble acknowledgment

First person to whom I would like to offer my regard is my supervisor Dr. Naveeda Riaz. I am very thankful to her for the supervision, advice, and guidance. Above all and the most needed, she provided me unflinching encouragement and support in various ways.

It would be an honor for me to record my gratitude to my parents for their myriad support, encouragement and love. Without their support it would be impossible for me to finish this thesis. I am indebted to them more than they know.

I am also grateful to our head of department Dr. Irfan for providing assistance. I am thankful to Ms Sana Aizad, Ms Samina Bilal and sir Ismail for their encouragement and advice at critical moments.

Collective and individual acknowledgments are also owed to my all friends especially Ms. Sumaira Nishat, Ms. Iffat Farzana, Ms. Harmain Rukh, Ms. shafaq Ehsan , Ms. Awaisa Ghazal, Ms. Mamoona mustaq, Ms. Shagufta Kanwal, Ms Tahira Noor and Ms. Saima Zubair for their countless support, motivation and love.

Lastly, I offer my regards and blessings to all of those who supported me in any respect during the completion of this thesis.

Amara Jabeen

Table of Contents

List of Tables

List of Figures

LIST OF ACRONYMS

DSG4	Desmoglein 4
LIPH	Lipase H
GPCR	Guanine Protein Coupled Receptors
LPA	Lysophosphatidic Acid
3DID	3D Interacting Domains
SMART	Simple Modular Architecture Research Tool
ELM	Eukaryotic Linear Motif
HNN	Hierarchical Neural Networks
BLAST	Basic Local Alignment Tool
FASTA	Fast All
KEGG	Kyoto Encyclopedia of Genes and Genomes
7tm_1	7 transmembrane
cAMP	Cyclic Adenosine Monophosphate
cGMP	Cyclic Guanosine Monophosphate
PKC	Protein Kinase C
PDK1	Pyruvate Dehydrogenase Kinase
EGFR	Epidermal Growth Factor Receptor
MAPK1	Mitogen Activated Protein Kinase 1
Hh	Alpha Helix

Ee	Extended Strand
Cc	Random Coil
PLA-1	POU domain, class 2, transcription factor 3 (Octamer-binding transcription factor 11)
EDG7	Lysophosphatidic acid receptor Edg-7 (LPA receptor 3)
PDNP2	Ectonucleotide pyrophosphatase/phosphodiesterase family member 2 precursor
STCH	Stress 70 protein chaperone microsome-associated 60 kDa protein precursor
LCAT	Solute carrier family 12 member 4
GP1B	Platelet glycoprotein Ib alpha chain precursor
PNPLA2	patatin-like phospholipase domain containing 2
SOAT1	Sterol O-acyltransferase 1
LIPF	Gastric triacylglycerol lipase precursor
CEL	Bile salt-activated lipase precursor
RB1	Retinoblastoma-associated protein
P2X4	P2X purinoceptor 4 (ATP receptor)
P2RX5	Tax1-binding protein 3 (Tax interaction protein 1)
P2RX1	P2X purinoceptor 1 (ATP receptor)
P2RX7	P2X purinoceptor 7 (ATP receptor)
RCBTB2	RCC1 and BTB domain-containing protein 2
P2XM	P2X purinoceptor 6 (ATP receptor)
HSD3B1	3 beta-hydroxysteroid dehydrogenase/Delta 5-->4-isomerase type I

ACO2	Aconitate hydratase, mitochondrial precursor
ENTPD1	Ectonucleoside triphosphate diphosphohydrolase 1
Pdb	Protein Data Bank
pI	Isoelectric Point
PSI-BLAST	Position specific Iterations - Basic Local Alignment Tool
ECL	Extra Cellular Loop

1. INTRODUCTION

Completion of human genome project has empowered us by genomic information; it has now become more convenient and effective to identify genetic factors involved in a disease, a state that is characterized by impairing of a normal function of an organism or a body. Diseases can be classified into infectious diseases and genetic diseases. Both environmental and genetic factors play role in causing any disease. Genetic disorder is caused by abnormalities that are present in genome of an individual.

Genetic disorders can be:

1. Monogenic
2. Complex
3. Chromosomal
4. Mitochondrial

Monogenic disorders are basically caused by defects in a single gene and can be classified into Autosomal recessive, Autosomal dominant and X-linked.

Missense mutation leads to single nucleotide change in a gene which is then translated and results in the formation of mutated protein. Protein structure is very crucial as it determines the function that the protein has to perform. (Duan *et al.*, 2001). Outcome of a mutated gene is abnormal protein that is incapable of carrying out normal functioning and as a consequence abnormal phenotype is encountered. Outcome of rapid advances of technology is accumulation of large amount of biological data. Large amount of proteins have been sequenced and are stored in databases like swissprot. This data is useless until or unless it is retrieved and exploited in research work. Data retrieval from a huge dataset is really a hampering task that is solved by bioinformatics tools. Data from different databases can be retrieved by bioinformatics tools and can be used to derive appropriate information. Determining effect of mutations on protein through experimental methods is very expensive, laborious and time consuming. Through bioinformatics tools useful information can be extracted from various databases. Through similarity searching homologs can be identified that can assist in function determination. Functional sites can be predicted through various tools. Interaction of particular protein can be identified. This retrieved information can then be used in

designing therapeutic targets for particular disease. Genetic diseases can affect phenotype of an organism by effecting hair, blood, skeletal muscles etc. Hypotrichosis is one of the genetic hair disorder which is contemplated to be modeled in the current study. Genetic conditions that leads to hair disorders can be due to some other disorder or it can be isolated. Hypotrichosis is one of the isolated from of genetic disorders. (Aslam *et al.*, 2004).

1.1 Hypotrichosis:

Hair loss due to genetic basis can be grouped into three categories these are:

1. Complete or partial absence of scalp hair in early childhood.
2. Diffused hypotrichosis that may possibly or not possibly deteriorate during lifetime.
3. Lack of hair in specifically segregated streaks.

Hypotrichosis is an inherited hair ailment that is characterized by a condition of having no hair growth and total baldness of the affected area that remains unchanged throughout the life of an affected individual. It affects an individual right from birth and it usually stays with them throughout their lives. . Despite of the fact that hairs are not necessary part of life, but hair loss can affect the feelings. When hair is thinning and Balding, It can lead to shame and depression.

1.2 Causes of Hypotrichosis:

Genetic eccentricities and are/or defects in embryonic development are major causes of Hypotrichosis. There are three stages of human hair growth cycle:

1. Anagen i.e. growth phase
2. Catagen i.e. degradation phase
3. Telogen i.e. resting phase

Growth periods of hairs between two and eight years are followed by a short period of two to four weeks which is characterized by degradation of hair. Resting phase then instigates and lasts for two to four months. Hair shedding takes place only after the next growth cycle initiates and a new hair shaft begins to emerge. Approximately 50-100 telogen hairs are fell off every day. This is considered as normal hair loss and accounts

for the hair loss encountered every day in the shower and with hair combing. These hairs usually re-grow. Ten percent of the hair follicles are in the resting phase at any time. Development of hair follicle takes place as a result of series transmission of reciprocal epithelial mesenchymal signals between the dermal papilla (DP) and the overlying epithelium during morphogenesis (Hisham *et al.*, 2003). Hair formation and growth takes place in hair follicle so all of the hair diseases are related to defects in the hair follicle (Wali *et al.*, 2007). Lysophosphatidic acid (LPA, 1- or 2-acyl-sn-glycerol 3-phosphate) which is a simple phospholipid interacts with G-protein-coupled seven transmembrane receptors (GPCRs) and nuclear hormone receptors. (Kano *et al.*, 2008). LPA is a product of Lipase H (*LIPH*). LPA controls the growth of the hair by interacting with *P2Y5* (Azeem, *et al.*, 2008*)*. *P2RY5* is a member of purine and pyrimidine nucleotide receptors family that are coupled to G proteins. Cell-surface G protein-coupled receptors mediate the cellular effects of LPA. Open reading frame for *P2RY5* is present in 17th intron (largest intron in RB1) of Retinoblastoma (RB1). (Song *et al.*, 2006). Retinoblastoma susceptibility gene resides on chromosome 13q14.12-13q14.2. It consists of 27 exons extended over 180 Kb region and 26 introns. It is a member of tumor suppressor family (Herbert *et al.*, 1996). *LIPH* is present on 3rd chromosome between region 3q26.33–q27.3 and is flanked by D3S2314 and D3S1602 markers (Aslam *et al.*, 2004). *LIPH* is a secreted protein that is expressed in intestine, lungs and pancreas. (Weijun *et al.*, 2002). Hair loss is a result of disturbances in hair growth life cycle, signaling molecules and pathways involved in hair follicle formation (Ghazanfar *et al.*, 2007) for instance mutations in *P2RY5* results in disruption of inner root sheath of hair follicles (Figure 1.1) (Azeem *et al.*, 2008*)*.

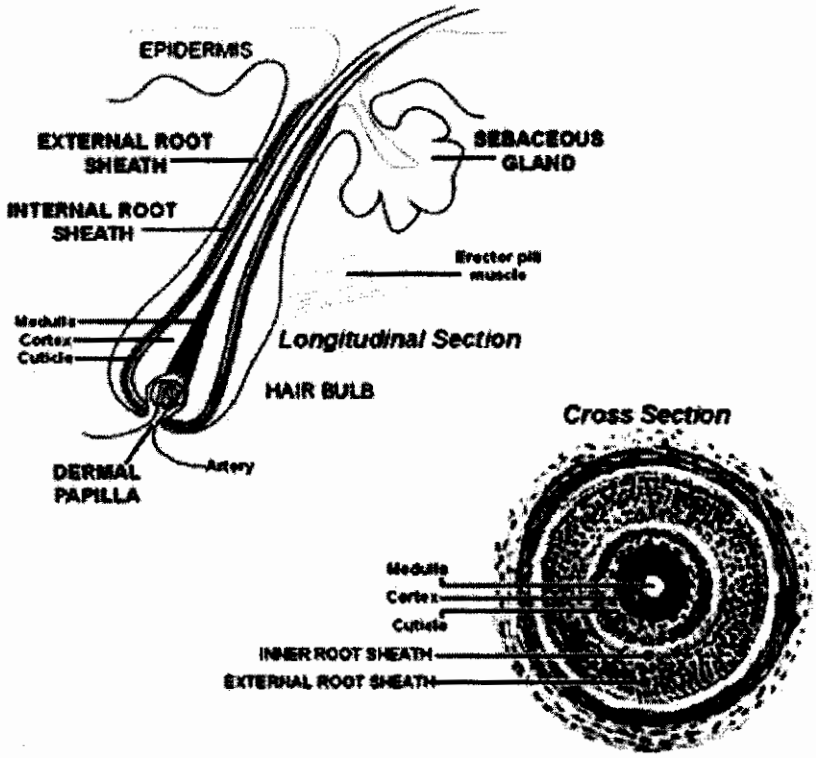


Figure 1.1: Structure of hair follicle

1.3 Forms of Autosomal Recessive Hypotrichosis:

Hypotrichosis can be classified into many types e.g. hypotrichosis, marie unna type, hypotrichosis simplex of scalp, hypotrichosis congenital with juvenile macular dystrophy (HJMD), hypotrichosis progressive patterned scalp with wiry hair, onycholysis, and cleft lip palate and autosomal recessive hypotrichosis. There are three genetically distinct forms of autosomal recessive hypotrichosis whose patients show similar type of clinical features these forms are:

1. Localized hypotrichosis or LAH that results from mutations in *desmoglein 4 (DSG4)*, located at chromosome 18q12.1, it is now renamed as LAH1.
2. Autosomal recessive hereditary hypotrichosis or AH that results due to mutations in *Lipase H (LIPH)* gene which is located at chromosome 3q27.2, it is now renamed as LAH2 (Wali *et al.*, 2007). Allelic variants of *LIPH* gene are shown in Table 1.2
3. Autosomal recessive hereditary hypotrichosis simplex (LAH3) that results from mutation in *P2RY5 l*ocated at chromosome 13q14.11–q21.32. The form of hypotrichosis that is most common in Pakistan is Autosomal recessive hereditary hypotrichosis simplex (LAH3).

1.4 Clinical features of Autosomal recessive hypotrichosis:

Autosomal recessive hypotrichosis is characterized by following clinical features:

1. Diffuse and progressive hair loss, usually begins in early childhood. (Pasternack *et al.*, 2008, Aboud *et al.*, 2008).
2. Hair on eyebrows, scalp, body and eyelashes are sparse or absent at all.
3. Absence of axillary hair.
4. Affected Males usually have sparse beards (Wali *et al.*, 2007, Aboud *et al.*, 2008). While in LAH 2 normal beard is present in affected males. (Aslam *et al.*, 2004)
5. In some of the cases disease is characterized by lack of normal hair follicle structures and comedo-like leftovers of the hair follicle. The leftovers of the hair follicle infundibulum shows hyperkeratinization. Morphogically Sebaceous

glands appear to be normal but lost acquaintances to the leftovers of the hair follicle infundibulum (Azeem *et al.*, 2008).

6. Hair on legs and arms are absent in males affected by LAH 2. (Aslam *et al.*, 2004)

1.5 Association of hypotrichosis with other diseases:

Beyond lack of hair individuals affected with hypotrichosis have many other physical or mental .Conditions like Graham-Little syndrome, Ofuji syndrome, cartilage-hair hypoplasia, Jeanselme and Rime hypotrichosis are some of the diseases where hypotrichosis is considered as a symptom. Hypotrichosis is also present in patients of Nicolaides-Baraitser syndrome (NBS) as a symptom.

(Castori *et al.*, 2008)

1.6 Genetics of the disease:

Mutations in *P2RY5* results in Autosomal recessive hypotrichosis simplex (LAH3). Total twelve mutations have been reported so far (Table 1.1). 5-kb deletion has been reported within the *DSG4* gene that leads to autosomal recessive localized hypotrichosis (LAH1). The deletion began 35 base pair upstream of exon 5 and ends at 289 base pair downstream of exon 8 that results in missing amino acids from 125 to 335. Two missense mutations have been identified in *LIPH* that leads to autosomal recessive hypotrichosis (LAH2) (Table 1.3)

1.7 Problem Definition:

Missense mutations do affect primary structure of protein as a consequence 3D structure changes. It is indispensable to predict effect of mutations on physio-chemical properties, post-translational modifications, domains, secondary structure and Tertiary structure in order to have better therapeutic strategies for a particular genetic disease.

Physiochemical properties of a protein are determined by the properties of amino acids that it contains. Physiochemical properties are really crucial in determining ligand that will bind to that protein e.g. if a protein has electrostatic potential it will act as a driven force for the ligand to be drawn from solvent and fit into the binding

site of the protein. (Kahraman, 2007). Isoelectric point is cruicial in understanding enzyme substrates interactions. Some physiochemical properties like hydrophobicity and charge are conserved during evolution. Isoelectric point can be determined through SDS-PAGE. Molecular weight can be determined through SDS PAGE, precipitation chromatography, elution sedimentation velocity, and gel permeation chromatography methods (Yamamoto, 2006). Domains are functional part of a polypeptide and it folds independently into a stable tertiary structure. (Chothia, 1992). Domain can be thought of as functional unit of a protein. (Bork, 1991). Secondary structure is important as it determines tertiary structure of a protein. Alpha helices and beta sheets take part in secondary structure formation. Some amino acids have greater propensity to be a part of alpha helices and some have greater propensity of forming beta sheets. Propensity of different residues is shown in Table 1.4. Strategies that have been currently used to predict 3D structures are X-ray Crystallography and Nuclear magnetic resonance spectroscopy but these methods are costly, protracted, time taking and have certain protein size constraints. Due to these reasons proteins structure information is still limited. Bioinformatics computational methods and molecular dynamic simulations are the solution to this problem and serve as alternative tool for protein structure prediction. (Liang *et al.*, 2005). To understand alterations brought out by mutations, effect of mutations at molecular level has to be highlighted. In order to have a therapy for autosomal recessive hyptrichosis, effect of mutations on physiochemical properties, domains, post-translational modifications 2D and 3D structures of *P2RY5*, *LIPH* and *DSG4* must be predicted.

1.9 Methodology adopted:

In order to design an effective agonist or antagonist, protein must be fully modeled in a way that its all functional sites must be identified. Bioinformatics offers a quick solution for identification of functional sites. Variety of tools is available that search against different databases to identify functional sites so the strategy that was adopted in current study is as follows:

1. Retrieving reported mutations from literature.

2. Prediction of functional sites by using different tools of motifs, domains, patterns, and post-translational modifications predictions.
3. Finding alignment of protein sequences with stored sequences using similarity search tools.
4. Prediction of normal and mutated proteins 3D structure.
5. Analysis of mutations affect at molecular level.
6. Screening a ligand molecule that can best bind with mutated structure of *LIPH* to prevent its malfunction.

1.10 Objectives of the study:

The objectives underneath this study are to:

1. Predict how mutations affect domains, post-translational modifications, secondary structure and physio-chemical properties of *P2RY5* and *LIPH.*
2. Predict 3D structure of *P2RY5* and *LIPH.*
3. Predict mutated tertiary structure and compare them with normal structures.

Table 1.1: Mutations reported in *P2RY5*

Serial No.	Mutation	Type	Alteration
1	c.436G>A	Missense	Glysine to Arginine (Azeem *et al.*, 2008)
2	c.36insA	Frameshift	stop at codon 16 (Azeem *et al.*, 2008)
3	c.160insA	Frameshift	stop at Codon 58 (Azeem *et al.*, 2008)
4	c.8G> C	Missense	Serine to Therionine (Azeem *et al.*, 2008)
5	c.565G	Missense	Glutamic acid to lysine (Azeem *et al.*, 2008)
6	c.69insCATG	Frameshif	and stop at codon 52 (Azeem *et al.*, 2008, Shimomura *et al.*,2008)
7	c.188A>T	Missense	Aspartic acid to valine (Azeem *et al.*, 2008), (Shimomura *et al.*,2008)
8	c.463C>T	Missense	Premature termination of translation (Pasternak *et al.*, 2008)
9	373_374delAA	Frameshift	premature termination of translation (Pasternack *et al.* , 2008)
10	c.742A>T	Missense	Aspargine to Tyrosine (Tariq et al., 2009)
11	c. .830C>T	Missense	Leucine to Proline (Tariq et al., 2009)

Table 1.2: Allelic variants of *LIPH*

Serial No.	Variant	Effect
1	Homozygous Exon 4 deletion	autosomal recessive hypotrichosis (AH) (Kazantseva *et al.*,2006)
2	c.346-350delATATA in exon 2	autosomal recessive hypotrichosis (Ali *et al.*, 2007)
3	c.659-660delTA in exon 5	autosomal recessive hypotrichosis (Jelani *et al.*, 2008)

Table 1.3 Mutations Reported in *LIPH*

Serial No.	Mutation	Type	Alteration
1	c.322T > C	Missense	p.W108R (Naz *et al.*, 2009)
2	c.2T > C	Missense	p.M1T (Naz *et al.*, 2009)

Table 1.4: Propensity of each amino acid to form secondary structure

Amino Acid	Alpha- helix Propensity	Beta sheet Propensity
Alanine	1.41	0.72
Arginine	1.21	0.84
Aspartic acid	0.99	0.39
Cystein	0.66	1.40
Glutamic acid	1.59	0.52
Glycine	0.43	0.58
Histadine	1.05	0.80
Isoleucine	1.09	1.67
Leucine	1.34	1.22
Lysine	1.23	0.69
Phenyl alanine	1.16	1.33
Methionine	1.30	1.14
Proline	0.34	0.31
Serine	0.57	0.96
Therionine	0.76	1.17
Tryptophan	1.02	1.35
Tyrosine	0.74	1.45
Valine	0.90	1.87
Asparagine	0.76	0.48
Glutamine	1.27	0.98

2. Materials and Methods

The current study i.e. "In Silico mutation analysis LAH2 & LAH3" is divided into four modules.

2.1 In silico modeling of *P2RY5* and *LIPH*.

2.2 Searching reported mutations in *P2RY5* and *LIPH* proteins through literature.

2.3 Analyzing effect of mutations on functional sites, physio-chemical properties secondary structure and tertiary structure.

2.4 Ligand Screening for *LIPH*.

2.1 In Silico modeling of *P2RY5* and *LIPH*:

P2RY5 and *LIPH* proteins must be modeled so that it can be investigated that whether missense mutations are affecting functional regions, 2D, 3D and physiochemical properties of these proteins or not. Biological data is available in bulk over the internet. The most practical way to model a protein is to use bioinformatics tools that are used to access data that is present in different databases. To search functional sites of *P2RY5* and *LIPH* following tools are used in this study:

2.1.1 Inter proscan

2.1.2 3DID

2.1.3 SMART

2.1.4 Scan Prosite

2.1.5 ELM

2.1.6 Motif scanner

2.1.7 NetNGlyc

2.1.8 kinasepho 2.0

2.1.9 NetCGlyc

2.1.10 NetPhos 2.0

2.1.11 2ZIP

2.1.1 Inter proscan:

Interpro scan is an integrative database for protein analysis. It integrates the functional sites and families predicted by various databases such as Gene3D, PANTHER, Pfam, PIRSF, PRINTS, ProDom, PROSITE, SMART, SUPERFAMILY and TIGRFAMs. This integration is a manual operation. Half of the total (Approximately 58,000) signatures available in the source databases are members of InterPro. Each member database's signatures are assembled via different methodology. When different signatures match the same composite of proteins at the same expanse on the sequence, they are placed into a single InterPro entry by a curator. (Hunter *et al.*, 2009)

2.1.2 3DID:

3DID is database for interacting domains i.e. it predicts interaction between domains of proteins. Interactions of proteins that have known 3D structure are explored and two types of information are derived that are domain to domain interactions and domain to peptide interactions. Domain–domain interactions are characterized by the binding of two globular domains and usually occur in multimeric enzymes and large multiprotein complexes. In domain-peptide interactions a globular domain binds and interacts with motifs of some other proteins. (Stein *et al.*, 2008)

2.1.3 Simple Modular Architecture Research Tool (SMART):

It is protein domain identification tool. Currently 784 protein domains manually curated models are contained in SMART. It contains manually curated Hidden Markov Models for many domains. SMART uses uniProt as a source of protein sequences. It also integrates different metabolic pathways information. It loads extended information on all detected SMART domains upon user request. (Ivica *et al.*, 2009)

2.1.4 Scan Prosite:

ScanProsite is a secondary database that identifies protein matches against PROSITE database's signatures. It can be thought of as enhanced version of PROSITE database. It detects structural and functional intra domain residues for PROSITE profiles. It matches both UniProt Knowledgebase sequences and user defined sequences against PROSITE

profiles. Output can be shown in simple text form and/ or graphical views can be generated. (Virginie *et al.*, 2006)

2.1.5 Eukaryotic Linear Motif server (ELM):

Certain non-globular large segments of multi-domain proteins subsist that are unable to fold into a proper tertiary structure but they can be of significant importance. ELM is used to detect these linear motifs in eukaryotic organisms. These linear motifs are not detected by other pattern and profile search databases. Motifs predicted by ELM may be a result of false sequence matches i.e. sequence matched to a motif is not a functional site in real sense. Context filters are developed to eliminate these false positives. Cell compartment filtering, Globular domain filtering and Taxonomic filtering is used. (Rune *et al.*, 2003)

2.1.6 Motif Scanner:

Motif scanner scan motifs in a given protein sequence. Oriented peptide library experiments are used to determine motifs. Position specific scoring matrix (PSSM) is used to represent each sequence motif.). Motif scanner computationally predict all the motifs in a query sequence and database search finds all proteins in a protein database, such as SWISS-PROT, TrEMBL, EMBL that matches a given motif. (John *et al.*, 2003)

2.1.7 NetNGlyc:

NetNGlyc is used to predict Glycosylation of a protein, which is a post translational modification that affects protein folding localization and trafficking. NetNGlyc uses artificial neural networks to predict glycosylation sites. (Gupta *et al.*, 2004)

2.1.8 Kinase Phos 2.0:

Kinase phos 2.0 uses concept of support vector machines (SVM) with the protein coupling pattern and protein sequence profile. Library, namely LIBSVM , is applied for training the predictive models.(Wong *et al.*, 2007)

2.1.9 NetCGlyc:

NetCGlyc uses artificial neural networks to predict C-mannosylation sites in mammals, which is a post-translational modification. (Julenius, 2007)

2.1.10 NetPhos 2.0:

Netphos 2.0 server uses artificial neural networks (ANN) to predict phosphorylation sites in provided sequence. Range of sensitivity of this method is 69% to 96 %.
(Blom *et al.*, 1999)

2.1.11 2ZIP – Server:

2ZIP server predicts leucine zippers from sequence alone and requires no information from homologue. It integrates coiled coil prediction algorithm with an estimated search for the characteristic leucine repeat to predict leucine zippers in a supplied sequence. The advantage of this server over other servers is that it avoids false positives to a greater extent. (Rivals *et al.*, 1998).

After prediction of functional sites, physiochemical properties of *P2RY5* and *LIPH* were predicted by using protparam.

2.1.12 ProtParam:

ProtParam computes molecular weight, amino acid composition, theoretical isoelectric point, atomic composition, instability index, extinction coefficient, estimated half-life, aliphatic index and grand average of hydropathicity (GRAVY) of a protein from its sequence. The only input is protein sequence or its SwissProt accession number without requiring any additional parameters.
http://www.expasy.ch/tools/protparam-doc.html
To predict secondary structure following tools were used.
2.1.13 HMM top
2.1.14 Hierarchical Neural Network (HNN)

2.1.13 HMM top:

The HMMTOP is a transmembrane topology prediction server that not only predicts the topology of transmembrane proteins but it also predicts localization of helical transmembrane segments. The basis for HHMTOP is that topology of transmembranes

are determined by the maximum divergence of composition of sequence segment's amino acid. (Tusnady & Simon, 2007)

2.1.14 Hierarchical Neural Network (HNN):

HNN comprises two types of networks that are sequence to structure network and structure to structure network. It also incorporates data on physiochemical properties that is utilized usually by structure to structure network.

(http://npsa-pbil.ibcp.fr/NPSA/npsa_references.html#hnn)

There are three methods for prediction of tertiary structures and these are Ab initio, threading and homology modeling. Threading is used to judge the fold of the protein. The accuracy and applicability of models produced by Ab initio methods are very low. Homology modeling gives the most accurate results if a known protein structure that is sufficiently similar to the modeled sequence is available so homology modeling and threading was used in this study. For homology modeling a template has to be searched by aligning target sequences with sequences that are stored in databases. For similarity searching following tools were used in current study:

2.1.15 BLAST

2.1.16 FASTA

2.1.15 BLAST:

BLAST is a tool that is used for similarity searching of protein and DNA sequences. It find matching sequences in an input sequence by searching a huge sequence database like Genbank, pdb, etc. BLAST looks for local alignments PSI-BLAST is a type of BLAST that is used to uncover several new and interesting protein family members.

(Altschul *et al.*, 1997)

2.1.16 FASTA:

FASTA is a program for high speed alignment of pairs of protein and DNA sequences. It is a sensitive imitative of the FASTP program. It locally align the sequences by matching k-tuples (words). It allows DNA or protein sequences comparison on the basis of variety of alternative scoring matrices.(Pearson & Lipman , 1988)

Modeller was used for comparative modeling.

2.1.17Modeller:

Modeller is a package for comparative modeling. It builds 3D models of a target protein using templates that are retrieved from macromolecule structure databases such as Protein data bank (PDB). The template should have homology with target sequence. Through modller, model of target is built in three steps target-template alignment, model building and model evaluation. (Eswar *et al.,* 2007)

2.1.18 Threading:

In threading amino acid sequence is evaluated against one of the already known 3D structure of a protein. The sequence-structure fit quality is evaluated using inter residue potentials of mean force or some other statistical parameters (Rost *et al*, 1997).

Fold prediction was done by using SAMT02 server.

2.1.18.1 SAM T02:

It searches similar protein sequences in NR and then aligns them to provide sequence logos with relative conservation of different sites Full three dimensional protein models are constructed by local structure predictions that are done with neural networks for numerous different local structure alphabets and hidden Markov models are created then fold recognition and alignment to proteins in the Protein Data Bank are completed.

(http://compbio.soe.ucsc.edu/SAM_T08/T08-query.html)

To visualize the predicted 3D structure of *P2RY5* and *LIPH* following tools are used.

2.1.19 swiss PdbViewer

2.1.20 Rasmol

2.1.19 Swiss PDB Viewer:

Swiss PDB viewer provides an environment to visualize PDB files i.e. it displays 3D structure of a protein. Moreover Swiss PDB viewer is used as a tool for comparative modeling and it is also used to insert mutations in a 3D structure and to visualize the changed residue. (Guex & Peitsch , 1997)

2.1.20 Rasmol

RasMol is a molecular graphics tool anticipated for the visualizing proteins, nucleic acids and other small molecules. The program is aimed at display and generation of images that are of publication quality. The program reads in coordinate file of a molecule and interactively reveals the molecule on the screen in a wide variety of color schemes and molecule representations.

(http://pps00.cryst.bbk.ac.uk/tech/rasmol/rasmol-doc.html)

Predicted structure must be evaluated in order to gain confidence. For evaluation of predicted tools following tools were used:

2.1.21 Procheck

2.1.22 What-If

2.1.23 Protein structure analysis (Prosa)

2.1.21 Procheck:

Procheck is used to assess stereo chemical quality of a predicted protein structure at both global and local level. Distribution of phi, psi and chi torsion angles and hydrogen bond energies are considered as global parameters. Parameters are calculated from structural coordinates. (Morris *et al.*, 1992)

2.1.22 What-IF:

WHAT IF is an integrated tool that is used to display, control, and investigate proteins, nucleic acids, small molecules, and their interactions. It can be used for inserting mutations in PDB files. Its menu driven function, in concert with the use of default values wherever input is required by the user, make its use easier.

(Vriend, 1990).

2.1.23 Protein structure analysis (Prosa):

Prosa is a tool that is used for the refinement and evaluation of experimental protein structure, structure prediction and modeling. It highlights problems in protein structure by displaying energy plots and scores. Problematic parts of a structure are shown and highlighted in a 3D molecule viewer. (Wiederstein & Sippl, 2007)

Patterns that were predicted thorough Scan prosite was visualized in 3D structure by using CN3D that also showed domains in 3D structure of *LIPH.*

2.2 Searching reported misense mutations in *P2RY5* and *LIPH.*

Research papers in which mutations in *P2RY5* and *LIPH* are reported were retrieved from Pubmed and then mutations were inserted into 3D structures using WHATIF and Swiss PDB viewer tools.

2.3 Analyzing affect of mutations on functional sites, physio-chemical properties secondary structure and tertiary structure:

Mutations were checked for their presence in functional sites, secondary structure elements and 3D structures. Comparison of mutated and normal 3Dstructure is done to reveal effect of missense mutation on 3D structure. Effect of mutations on physiochemical properties, functional sites secondary structure and tertiary structure were analyzed to understand molecular cause of autosomal recessive hypotrichosis.

2.4 Ligand finding for LIPH:

For searching appropriate ligand, following strategy was adopted

2.4.1 Active site prediction

2.4.2 Ligand screening

2.4.1 Active site prediction:

For Active site prediction Par3d tool was used.

2.4.1.1PAR3D:

PAR3d utilizes training set that is a set of formerly calculated and stored values of geometrical parameters of a set of known proteins for prediction of active sites of a query protein. Input of this server is a pdb file of query protein. Active site prediction is done in three steps that are extraction of possible sites, spatial arrangement of these extracted sites and comparison with stored geometries. (Goyal *et al.*, 2007)

2.4.2 Ligand screening:

Ligand screening was done by KEGG Ligand database.

2.4.2.1 KEGG:

The Kyoto Encyclopedia of Genes and Genomes (KEGG) is the primary database resource for indulgent of higher order functional denotation and utilities of the cell or the organism from its genomic information. It consists of the PATHWAY database that is meant for the automated information of molecular interaction networks like pathways and complexes, GENES database meant for the information about genes and proteins produced by genome sequencing projects and LIGAND database meant for the information concerning chemical compounds and chemical reactions relevant to cellular processes.

(Kanehisa *et al.*, 2002).

Table 2.1 Summary of tools used during current study

S.No.	Tool	Output
1	Inter proscan	Domains
2	3DID	Domain interactions
3	SMART	Protein interactions
4	Scan Prosite	patterns
5	ELM	Motifs
6	Motif scanner	Motifs
7	NetNGlyc	N-glycosylation sites
8	kinasepho 2.0	Kinase phosphorylation sites
9	NetCGlyc	Mannosylation sites
10	2ZIP	Leucine zippers
11	ProtParam	Physiochemical Properties
12	HMM top	Topology
13	HNN	Secondary structure
14	BLAST	Alignment
15	FASTA	Alignment
16	Modeller	3D model
17	SAM T02	3D model
18	swiss PdbViewer	Visualizer
19	Rasmol	Visualizer
20	CN3D	Visualizer
21	Procheck	Structure Evaluation
22	What If	Structure Evaluation
23	Prosa	Structure Evaluation
24	PAR3D	Active sites
25	KEGG Ligand	Ligands

3. RESULTS

3.1 Functional Sites predictions:

During evolution some regions of proteins are more conserved than others and these regions are basically responsible for functioning of a protein. These regions are called domains. Within domains, motifs are present that play a momentous role in determining function of a protein. Mutations in these regions can affect normal function of the protein.

3.1.1 Domains of *P2RY5* and LIPH

3.1.1.1 *P2RY5*:

InterproScan was used to predict domains shared by *P2Y5* with other proteins. *P2Y5* share domains with 7tm, GPCR rhodopsin-like family (Table 3.1(a)), P2Y5 purinoceptor family (Table 3. 1(b)) and GPCR rhodopsin like superfamily (Table 3. 1(c)). Interpro also predicted some domains that are not assigned to any family yet (Table 3.1(d)).

Table 3.1(a) 7tm, GPCR, rhodopsin like

S No.	Accession No.	Domain name	Domain	Location in the protein
1	PR00237	GPCRRHODOPSN		19-43, 52-73, 96-118,132-153, 185-208, 227-251, 273-299
2	PF00001	7tm_1		34-291

PRINTS retrieved GPCRRHODOPSN domain that is predicted to be shared by P2Y5. This domain has 7 motifs that are shared by most of the family members of 7tm GPCR,

rhodopsin like family (Attwood & Findlay, 1994). These motifs are shared by 686 receptors according to result of OWL database. These 7 motifs encode hydrophobic membrane-spanning regions G protein-coupled receptors are of five types that are "rhodopsin-like GPCRs, the secretin-like GPCRs, the cAMP receptors, the fungal mating pheromone receptors, and the metabotropic glutamate receptor family". Rhodopsin-like GPCRs family contains hormone, neurotransmitter and light receptors.7 tm_ 1 has signal transducer activity. Interaction pathway of 7tm_1 domain was predicted by using 3DID (Figure 3.1).

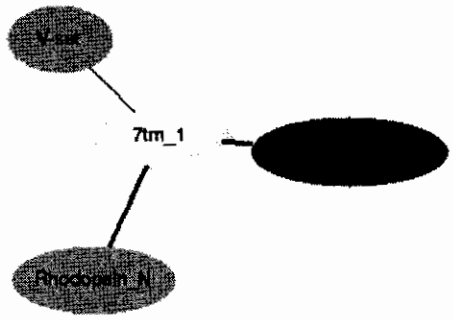


Figure 3.1: Interaction pathway of 7 tm _1 domain

Interaction pathway of 7 tm_1 domain shows that it interacts with V-Set, rhodopsin N and phage lysozyme domains. V-set has role in protein binding. Phage lysozyme has catalytic activity. Rhodopsin N is G Protein coupled receptor that utilizes heterotrimeric GTP binding proteins for the transduction of extracellular signals to intracellular events. So these 7 transmembrane regions of *P2RY5* are responsible for transducing extra cellular signals by interacting with Rhodpsin N and by facilitation of phage lysozyme.

Table 3.1(b): P2Y5 Purinoceptor Family

Accession No.	Domain name	Domain	Location
PR01067	P2Y5ORPHANR		9-22, 40-55, 169-179, 249-262, 264,279, 293-303

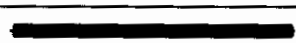
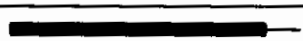
P2Y5ORPHANR is a 6-element. Motif 1 is present at the N-terminus, motif 2 covers the first cytoplasmic loop, 3rd motif is present in the second external loop, motif 4 spans the C-terminus of domain 6and leads to the third external loop, third external loop contains fifth motif and sixth motif inhabits the C-terminus.

Table 3.1(c): *P2RY5* Purinoceptor domain

Accession No.	Domain name	domain	Location
PS50262	G_PROTEIN_RECEP_F1_2		34-291

P2Y5 also share domain with GPCR rhodopsin like superfamily. G_PROTEIN_RECEP_F1_2 is a receptor that belongs to GPCR rhodopsin like superfamily. It has seven hydrophobic regions. N terminus of this receptor lies on the extracellular part of the membrane and is glycosylated and C terminus are on cytosol and is phosphorylated. To link seven transmembrane rehgions three intracellular loops alternate with three extracellular loops.

Table 3.1(d) Unintegrated domains

Sreial No	Accession No.	Domain name	domain
1	G3DSA:1.20.1070.10	G3DSA:1.20.1070.10	
2	PTHR19264	PTHR19264	
3	PTHR19264:SF280	PTHR19264:SF280	
4	SSF81321	SSF81321	

3.1.1.1.1 Mutation analysis:

After inserting reported mutations three domains are added (Table 3.2 (a) and Table 3. 2 (b)). Each mutation resulted in same change so, it can be concluded that wherever

mutation occurs in *P2Y5*, structural change that each mutation brought is same. G_PROTEIN_RECEP_F1_1 domain is added, in normal protein G_PROTEIN_RECEP_F1_2 domain is present, this domain is still present in mutated structure but site for G_PROTEIN_RECEP_F1_1 domain is also added. This receptor site might compete with normal site and may inhibit or reduce binding of ligand that is LPA. LPA binds with *P2Y5* to control hair growth if LPA would not bind with *P2Y5* properly then hair defect will occur. Due to these mutations a signal peptide, which was absent in normal *P2Y5*, is added. 6 transmembrane regions are also added. Signal peptide localizes the protein to specific regions within a cell. Due to addition of signal peptide in mutated structures protein might be localized to a certain location and unable to reach its specific location where it performs required function.

Table 3.2(a) 7tm, GPCR, rhodopsin like

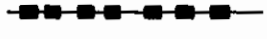
Serial No.	Accession No.	Domain name	Domain
1	PR00237	GPCRRHODOPSN	
2	PF00001	7tm_1	
3	PS00237	G_PROTEIN_RECEP_F1_1	

Table 3.2(b) Unintegrated domains

Serial No	Accession No.	Domain name	Domain
1	G3DSA:1.20.1070.10	G3DSA:1.20.1070.10	
2	PTHR19264	PTHR19264	
3	PTHR19264:SF280	PTHR19264:SF280	
4	signalP	Signal peptide	
5	tmHmm	Transmembrane regions	
6	SSF81321	SSF81321	

3.1.1.2 *LIPH* Domains:

Interpro Scan results for *LIPH* showed that it shares domains with lipase (Table 3.3 (a)), lipase N-terminal (Table 3.3 (b)) and lipoprotein lipase (Table 3.3 (c)) families. There are also some unintegrated domains that are shown in Table 3.3 (d)

Table 3.3(a): Lipase domains in *LIPH*

S.No.	Accession no	Domain name	Domain	Location
1.	PR00821	TAGLIPASE		59-78, 81-95, 104-119, 147-165, 246-261,325-240
2.	PTHR11610	LIPASE		3-450

The two domains retrieved by Prints and panther are taglipase and lipase respectively. TAGLIPASE belongs to triacylglycerol lipase family. Lipases are basically lipolytic enzymes that are involved in catalyzing hydrolysis of ester linkages of triglycerides. . Interaction pathway of Lipase domain is shown in Figure 3.2. It interacts with colipase and PLAT domain.

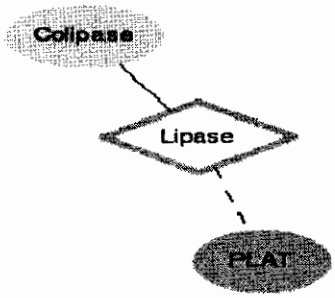


Figure 3.2: Interaction pathway of lipase domain

Table 3.3(b): Lipase, N-Terminal domain in *LIPH*

Accession no	Domain name	Domain	Location
PF00151	Lipase		39-326

PFAM also retrieved Lipase domain from Lipase N-terminal family.

Table 3.3(c): Lipoprotein Lipase, LIPH domain in *LIPH*

Accession no	Domain name	Domain	Location
PIRSF000865	Lipoprotein Lipase, LIPH type		1-450

Table 3.3(d): Unintegrated domains in *LIPH*

Serial no.	Accession no	Domain name	Domain	Location
1.	G3DSA:3.40.50.1820			38-328
2	PTHR11610:SF12	MEMBRANE-BOUND PHOSPHATIDIC ACID SPECIFIC PHOSPHOLIPASE A1		3-450
3	SignalP	Signal peptide		1-18
4	SF53474	alpha/beta-Hydrolases		5-237

3.1.1.2.1(b) p. M1T:

No effect of this mutation is predicted on domains of LIPH.

3.1.2 Patterns:

Patterns can be thought of as textual representation of motifs.

3.1.2.1 Patterns for *P2RY5*:

Total five patterns were found in *P2RY5* (Table 3.5)

Table 3.5: Patterns found in *P2RY5*

Predicted Site	Protein access No.	Found Pattern
N-glycosylation site.	PS00001	N-{P}-[ST]-{P}.
Protein kinase C phosphorylation site	PS00005	[ST]-x-[RK].
Tyrosine kinase phosphorylation site.	PS00007	[RK]-x(2,3)-[DE]-x(2,3)-Y
N-myristoylation site.	PS00008	G-{EDRKHPFYW}-x(2)-[STAGCN]-{P}
G-protein coupled receptors family 1 signature	PS00237	[GSTALIVMFYWC]-[GSTANCPDE]-{EDPKRH}-x-{PQ}-[LIVMNQGA]-{RK}-{RK}-[LIVMFT]-[GSTANC]-[LIVMFYWSTAC]-[DENH]-R-[FYWCSH]-{PE}-x-[LIVM].

3.1.2.2 Predicted patterns in LIPH:

7 patterns were found in *LIPH.* (Table 3.6)

Table 3.6: Patterns found in *LIPH*

Predicted Site	Protein access No.	Found Pattern
N-glycosylation site	PS00001	N-{P}-[ST]-{P}.
cAMP- and cGMP-dependent protein kinase phosphorylation site.	PS00004	[RK](2)-x-[ST].
Protein kinase C phosphorylation site.	PS00005	[ST]-x-[RK].
Casein kinase II phosphorylation site.	PS00006	ST]-x(2)-[DE].
N-myristoylation site.	PS00008	G-{EDRKHPFYW}-x(2)-[STAGCN]-{P}.
Cell attachment sequence.	PS00016	R-G-D.
Leucine zipper pattern.	PS00029	L-x(6)-L-x(6)-L-x(6)-L.

3.1.3 Motifs:

Motifs in LIPH and *P2RY5* were predicted through:

3.1.3.1 Motif Scanner

3.1.3.2 ELM

3.1.3.1 Motif Scanner:

3.1.3.1 .1 *P2RY5*:

Motifs predicted by motif scanner in *P2RY5* and *LIPH* are given in Table 3.7 and 3.9. Their diagrammatical representations are shown in Figure 3.3 and Figure 3.6 respectively.

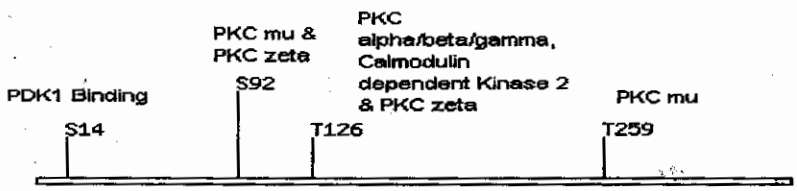


Figure 3.3: Predicted Motifs in *P2RY5*

Tale 3.7: Predicted motifs of *P2RY5*

Motif	Site	Score	Percentile	Sequence	SA
PKC mu	S92	0.3202	0.007 %	GDLLCKISVMLFYTN	0.261
PKC alpha/beta/gamma	T126	0.4187	0.289 %	VYPFKSKTLRTKRNA	2.066
PKC mu	T259	0.4530	0.407 %	YSLVRTQTFVNCSVV	0.887
Calmodulin dependent Kinase 2	T126	0.5136	0.813 %	VYPFKSKTLRTKRNA	2.066
PKC zeta	S92	0.5315	0.813 %	GDLLCKISVMLFYTN	0.261
PKC zeta	T126	0.5408	0.978 %	VYPFKSKTLRTKRNA	2.066
PDK1 Binding	S14	0.5910	0.568 %	SHCFYNDSFKYTLYG	2.324

Motif scan retrieved 7 motifs for *P2RY5*. PKC mu belongs to Protein kinases C family and takes part in various receptor mediated extracellular signal transduction pathways. PKC alpha/beta/gamma plays a vital role in cell transformation, cell adhesion, cell volume control and cell cycle checkpoint. According to some studies this gene might be an elementary controller of cardiac contractility and Ca(2+) managing in myocytes. Calmodulin dependent Kinase 2 belongs to Ca(2+)/calmodulin-dependent protein kinase subfamily. In mammalian

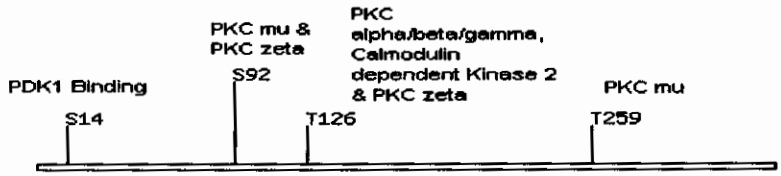


Figure 3.3: Predicted Motifs in *P2RY5*

Tale 3.7: Predicted motifs of *P2RY5*

Motif	Site	Score	Percentile	Sequence	SA
PKC mu	S92	0.3202	0.007 %	GDLLCKISVMLFYTN	0.261
PKC alpha/beta/gamma	T126	0.4187	0.289 %	VYPFKSKTLRTKRNA	2.066
PKC mu	T259	0.4530	0.407 %	YSLVRTQTFVNCSVV	0.887
Calmodulin dependent Kinase 2	T126	0.5136	0.813 %	VYPFKSKTLRTKRNA	2.066
PKC zeta	S92	0.5315	0.813 %	GDLLCKISVMLFYTN	0.261
PKC zeta	T126	0.5408	0.978 %	VYPFKSKTLRTKRNA	2.066
PDK1 Binding	S14	0.5910	0.568 %	SHCFYNDSFKYTLYG	2.324

Motif scan retrieved 7 motifs for *P2RY5*. PKC mu belongs to Protein kinases C family and takes part in various receptor mediated extracellular signal transduction pathways. PKC alpha/beta/gamma plays a vital role in cell transformation, cell adhesion, cell volume control and cell cycle checkpoint. According to some studies this gene might be an elementary controller of cardiac contractility and Ca(2+) managing in myocytes. Calmodulin dependent Kinase 2 belongs to Ca(2+)/calmodulin-dependent protein kinase subfamily. In mammalian cells four different chains that are alpha, beta, gamma, and

delta constitute this enzyme. PKC Zeta has kinase activity independent of calcium and diacylglycerol

3.1.3.1 .1 .1 Mutation Analysis:

After inserting mutations ErkD-domain and EGFR Kinase were added due to mutation p.G146R and p.N248Y respectively (Table 3.8 (a) and 3.8 (b)), (Figure 3.4 and 3.5).

ErkD-domain is present in mitogen-activated protein kinase1. MAPK 1 serve as an integral position for various biochemical signals and are responsible for various cellular processes such as transcription regulation, proliferation, differentiation, and development.

Table 3.8(a): Additional motif in mutated *P2RY5*

Motif	Site	Score	Percentile	Sequence	SA
ErkD-domain	V152	0.6654	0.729 %	IRGSAPAVFVQSTHS	0.296

Figure 3.4: Site of Additional motif in mutated *P2RY5*

Table 3.8(b): Additional motif in mutated *P2RY5*

Motif	Site	Score	Percentile	Sequence	SA
EGFR Kinase	Y248	0.4382	0.979 %	CFVPYNIYLILYSLV	0.200

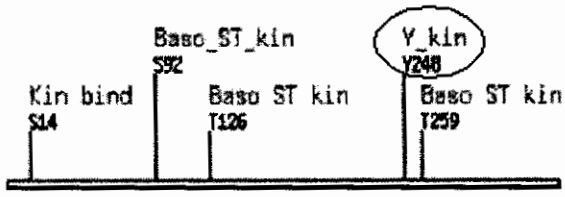


Figure 3.5: Site of Additional motif in mutated *P2RY5*

3.1.3.1 .2 *LIPH*:

Table 3.9: Predicted Motifs in *LIPH*

Motif	Site	Score	Percentile	Sequence	SA
Itk Kinase	Y149	0.4327	0.690 %	GASLDDIYMIGVSLG	0.289
PLCg N-terminal SH2	Y253	0.4113	0.994 %	CDHQRSVYLYLSSLR	0.381
Src SH2	Y253	0.4197	0.721 %	CDHQRSVYLYLSSLR	0.381
Nck SH2	Y213	0.4951	0.881 %	SDTDALGYKEPLGNI	1.569
Nck 2nd SH3	P179	0.4999	0.522 %	GRITGLDPAGPLFNG	0.754
PKC alpha/beta/gamma	T348	0.4668	0.822 %	NVRRGDITIKLRDKA	0.448
PKC zeta	S377	0.5092	0.496 %	FQKYHQVSLLARFNQ	0.271
Casein Kinase 1	S400	0.3554	0.103 %	SLMFSTGSLIGPRYK	0.251
PDK1 Binding	S29	0.5792	0.447 %	CPSFTRLSFHSAVVG	0.783
PDK1 Binding	S272	0.6134	0.843 %	ITAYPCDSYQDYRNG	1.246
Erk D-domain	I349	0.6051	0.280 %	VRRGDITIKLRDKAG	0.525

Table 3.10: ELM results description for *P2RY5*

Name	Description	Cellular Location
CLV_NDR_NDR_1	N-Arg dibasic convertase (nardilysine) cleavage site (Xaa-|-Arg-Lys or Arg-|-Arg-Xaa)	extracellular, Golgi apparatus, cell surface
CLV_PCSK_FUR_1	Furin (PACE) cleavage site (Arg-Xaa-[Arg/Lys]-Arg-|-Xaa)	Golgi membrane, extracellular, Golgi apparatus
CLV_PCSK_PC1ET2_1	NEC1/NEC2 cleavage site (Lys-Arg-|-Xaa)	Golgi membrane, extracellular, Golgi apparatus
CLV_PCSK_SKI1_1	Subtilisin/kexin isozyme-1 (SKI1) cleavage site ([RK]-X-[hydrophobic]-[LTKF]-|-X)	endoplasmic reticulum, endoplasmic reticulum lumen, Golgi apparatus
LIG_14-3-3_3	Consensus derived from natural interactors which do not exactly match the mode1 and mode2 ligands	nucleus, cytosol, internal side of plasma membrane
LIG_BRCT_BRCA1_1	Phosphopeptide motif which directly interacts with the BRCT (carboxy-terminal) domain of the Breast Cancer Gene BRCA1 with low affinity	nucleus, BRCA1-BARD1 complex
LIG_EH1_1	The engrailed homology domain 1 motif is found in homeodomain containing active repressors and other transcription families, and allows for the recruitment of	Nucleus

	Groucho/TLE corepressors.	
LIG_FHA_1	Phosphothreonine motif binding a subset of FHA domains that show a preference for a large aliphatic amino acid at the pT+3 position.	Nucleus
LIG_MAPK_1	MAPK interacting molecules (e.g. MAPKKs, substrates, phosphatases) carry docking motif that help to regulate specific interaction in the MAPK cascade. The classic motif approximates (R/K)xxxx#x# where # is a hydrophobic residue.	nucleus, cytosol
LIG_PDZ_3	Class III PDZ domains binding motif	cytosol, plasma membrane, membrane

3.1.3.2 .2Results of ELM for *LIPH*:

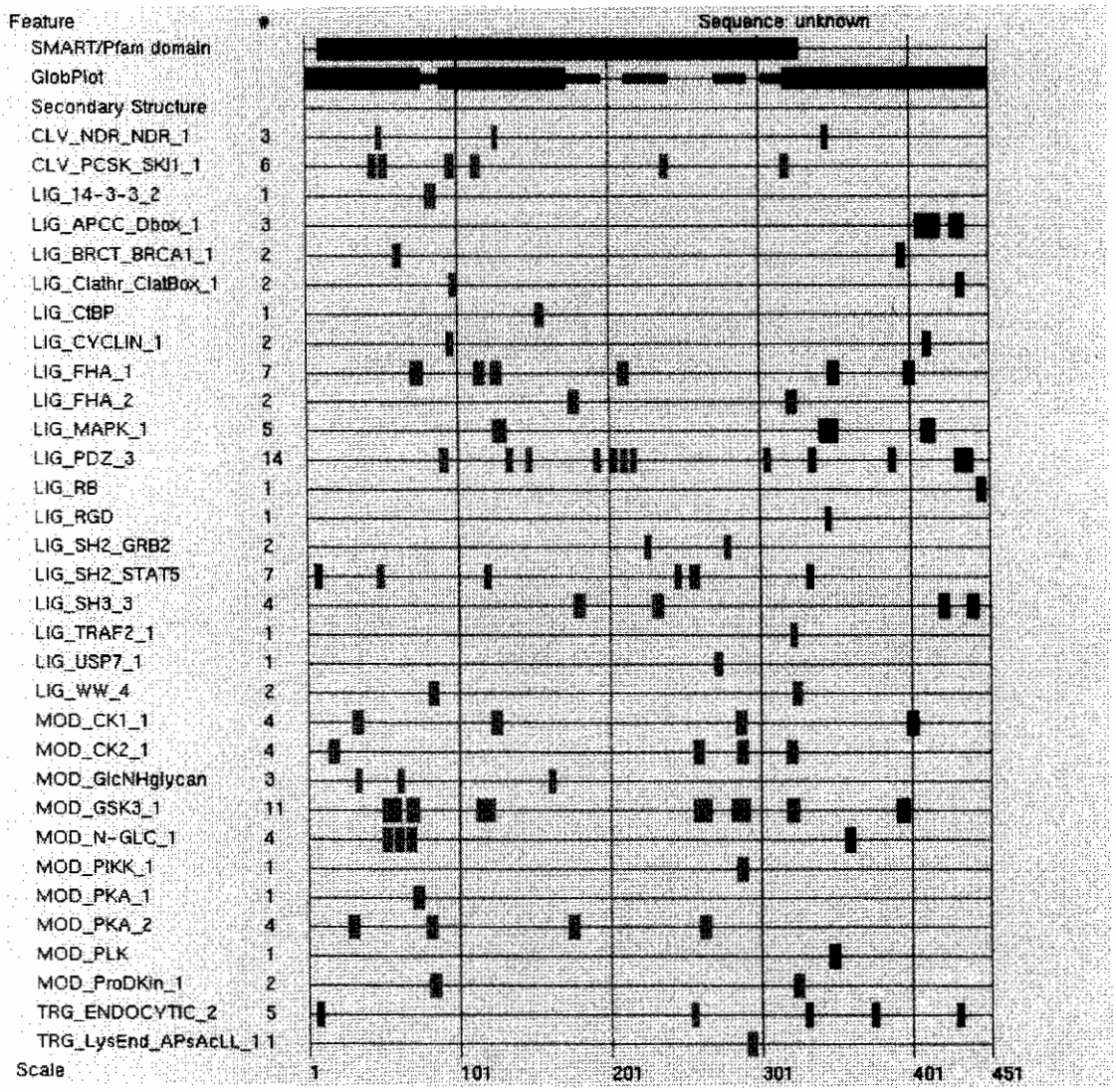


Figure 3.8: ELM result of *LIPH*

Table 3.11: ELM results description for *LIPH*

Name	Description	Cellular location
CLV_NDR_NDR_1	N-Arg dibasic convertase (nardilysine) cleavage site (Xaa-|-Arg-Lys or Arg-|-Arg-Xaa)	extracellular, Golgi apparatus, cell surface
LIG_APCC_Dbox_1	An RxxL-based motif that binds to the Cdh1 and Cdc20 components of APC/C thereby targeting the protein for destruction in a cell cycle dependent manner	nucleus, cytosol
LIG_BRCT_BRCA1_1	Phosphopeptide motif which directly interacts with the BRCT (carboxy-terminal) domain of the Breast Cancer Gene BRCA1 with low affinity	nucleus, BRCA1-BARD1 complex
LIG_Clathr_ClatBox_1	Clathrin box motif found on cargo adaptor proteins, it interacts with the beta propeller structure located at the N-terminus of Clathrin heavy chain.	Golgi apparatus, Golgi trans-face, cytosol, cytoskeleton, clathrin-coated endocytic vesicle
LIG_CYCLIN_1	Substrate recognition site that interacts with cyclin and thereby increases phosphorylation by cyclin/cdk complexes. Predicted protein should have the MOD_CDK site. Also used by cyclin inhibitors.	nucleus, cytosol

3.1.4 Post-Translational Modifications:

Polypeptide that is formed after translation might not be the final protein. At the termini of proteins numerous sequence motifs encode signals for posttranslational modifications (Frank *et al.,* 2003). There are certain posttranslational modifications that affect physical and chemical properties of proteins and their occurrence is crucial for proper functioning of a particular protein. Posttranslational modifications are accomplished by addition of some functional group, addition of some other peptides to protein, by changing chemical nature of amino acids and by bringing out structural changes.

3.1.4.1 Prediction of N-glycosylation sites:

Glycosylation is the procees of addition of glycosyl group to Aspragine, hydroxyline, serine or therionine. Glycosylation can affect protein folding, protein localization, protein trafficking, protein solubility, antigenicity, biological activity, half-life, and cell-cell interactions (Gupta *et al.,* 2004). Through pattern searching it was revealed that *P2RY5* and *LIPH* have N-glycosylation sites i.e. addition of glycosyl group to aspargine so this site was investigated through NetNGlyc 1.0 Server (Table 3.12 and Table 3.13) and (Figure 3.9 and 3.10).

Table 3.12: Predicted N-glycosylation sites in *P2RY5*

Position	Potential	Jury agreement	N-Glyc result
5 NSSH	0.5869	(8/9)	+
12 NDSF	0.4376	(7/9)	-
50 NETT	0.5957	(7/9)	+
163 NASE	0.4137	(7/9)	-
200 NVTC	0.7462	(9/9)	++
223 NKTK	0.6307	(8/9)	+
262 NCSV	0.5583	(7/9)	+
306 NWSV	0.3421	(9/9)	– –
340 NESA	0.4327	(6/9)	-

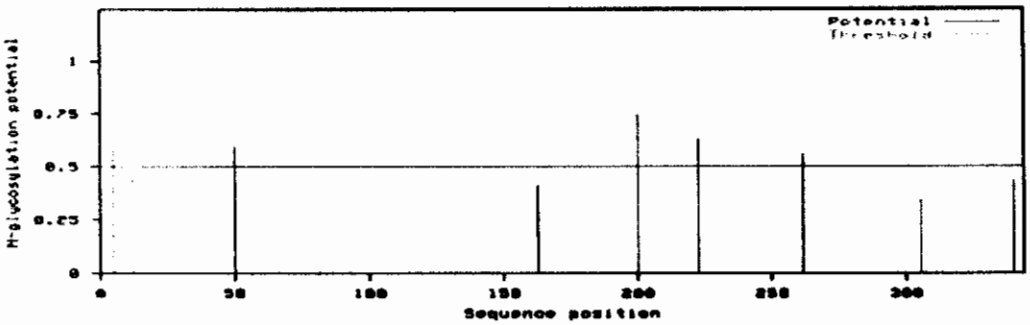


Figure 3.9: Predicted N-glycosylation sites in *P2RY5*

According to above graph there are total nine aspargines in *P2RY5* that can be gycosylated, four of which fall below threshold value and are predicted to be non-glycosylated (Table 3.9)

3.1.4.1.1. Mutation analysis for *P2RY5*:

Mutations that have been reported do not have any effect on N-glycosylation of *P2RY5*

Table 3.13: result of NetNGlyc prediction for *LIPH*:

Position	Potential	Jury agreement	N-Glyc result
50 NLTC	0.6992	(9/9)	++
58 NSSA	0.5706	(6/9)	+
66 NVTK	0.8354	(9/9)	+++
357 NTTE	0.6519	(9/9)	++

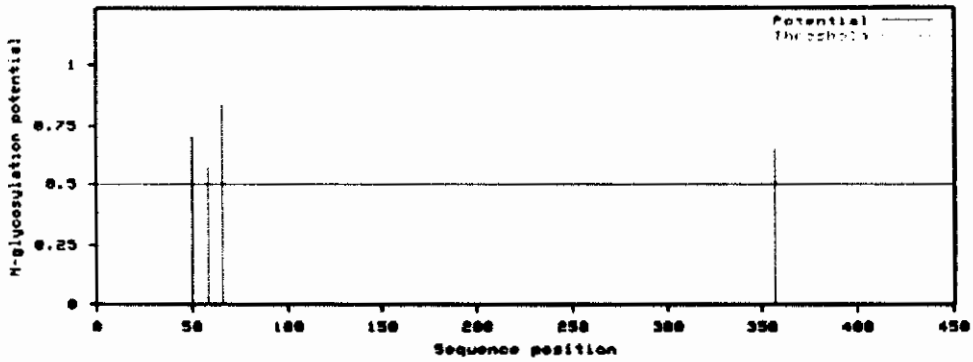


Figure 3.10: Predicted N-glycosylation sites in *LIPH*

There are total four aspargines in LIPH that are glycosylated.

3.1.4.1.2. Mutation Analysis for LIPH:

Mutations that have been reported do not have any effect on N-glycosylation of *LIPH*

3.1.4.2 Phosphorylation sites prediction:

Tyrosine kinase transfers phosphate group from ATP to tyrosine. Protein tyrosine kinase transmits signals for growth, differentiation, adhesion, motility, and death alteration in their activity can lead to diabetes, cancer and immune system malfunction. (Grossmann., 2004) . "Phosphorylation is important for cellular regulation, cellular signal pathways, metabolism, growth, differentiation and membrane transport" (Wong , 2007).

Prediction of phophorylation at serine, therinone and tyrosine site is important as it affects multitude of cellular signaling processes (Blom *et al.*, 1999) NetPhos 2.0 Server was used to predict phophorylation sites in *P2RY5* and *LIPH* Phophorylation sites for *P2RY5* are shown in Table 3.14 and Figure 3.11. Phophorylation sites for *LIPH* are shown in Table 3.15 and Figure 3.14.

Table 3.14: Predicted phophorylated serine, therionine and tyrosine in *P2RY5*

Position	Context	Score	Prediction
159	QSTHSQGNN	0.965	*S*
308	MKNWSVRRS	0.993	*S*
312	SVRRSDFRF	0.919	*S*
317	DFRFSEVHG	0.980	*S*
129	KTLRTKRNA	0.711	*T*
176	FPEATWKTY	0.622	*T*
259	VRTQTFVNC	0.797	*T*
332	HNLQTLKSK	0.769	*T*
17	DSFKYTLYG	0.698	*Y*
55	ETTTYMINL	0.907	*Y*
97	VMLFYTNMY	0.687	*Y*

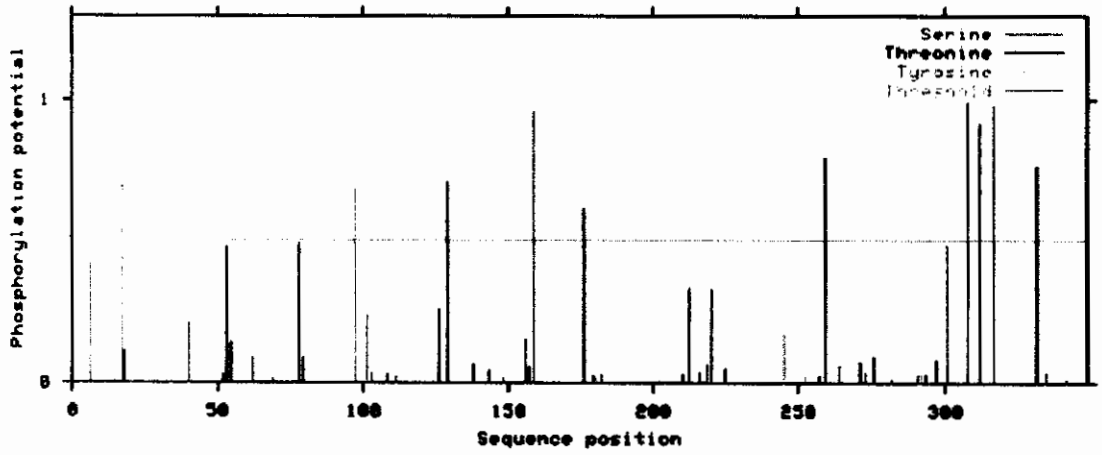


Figure 3.11: Graph showing predicted phophorylated serine, therionine and tyrosine in *P2RY5*

There are total 4 serines, therionines and 3 tyrosines in *P2RY5* that are phosphorylated.

3.1.4.2.1 Mutation Analysis:

Mutation p.G146R results in phohorylation of serine at position number 148 result is shown in Figure 3.12. Mutation p.L277P results in phophorylation of therionine at position number 276 (Figure 3. 13)

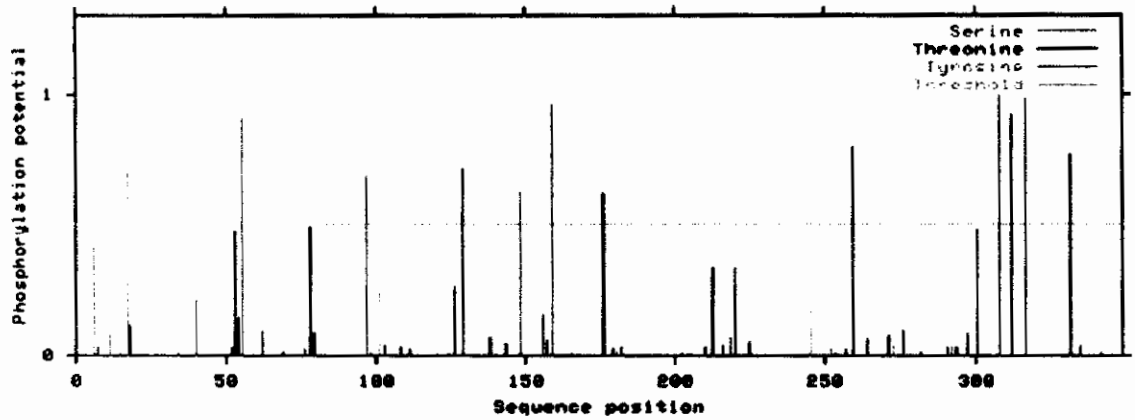


Figure 3.12: Depiction of additional Serine phosphorylation in mutated *P2RY5*

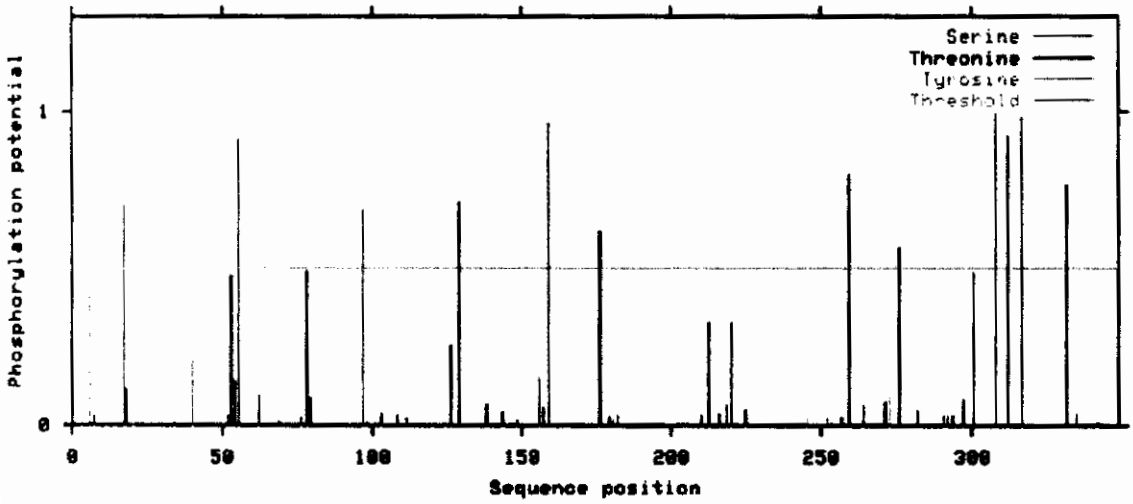


Figure 3.13: Depiction of additional Therionine phosphorylation in mutated *P2RY5*

Table 3.15: Predicted phophorylated serine, therionine and tyrosine in *LIPH*

Position	Context	Score	Prediction
14	LLCLSRSDA	0.987	*S*
29	FTRLSFHSA	0.959	*S*
97	KGLLSVEDM	0.987	*S*
121	YTHASSKTR	0.726	*S*
122	THASSKTRK	0.935	*S*
144	AEGASLDDI	0.729	*S*
206	DVIHSDTDA	0.979	*S*
262	SLRESCTIT	0.970	*S*
272	YPCDSYQDY	0.975	*S*
287	SCGTSQKES	0.996	*S*
324	TAEESPFCM	0.620	*S*
361	NTTESKINH	0.970	*S*
21	DAEETCPSF	0.564	*T*
81	GFRPTGSPP	0.650	*T*
124	ASSKTRKVA	0.608	*T*
320	AFFDTAEES	0.668	*T*
348	RGDITIKLR	0.729	*T*
149	LDDIYMIGV	0.597	*Y*
213	DALGYKEPL	0.690	*Y*
223	NIDFYPNGG	0.501	*Y*
273	PCDSYQDYR	0.965	*Y*
276	SYQDYRNGK	0.598	*Y*

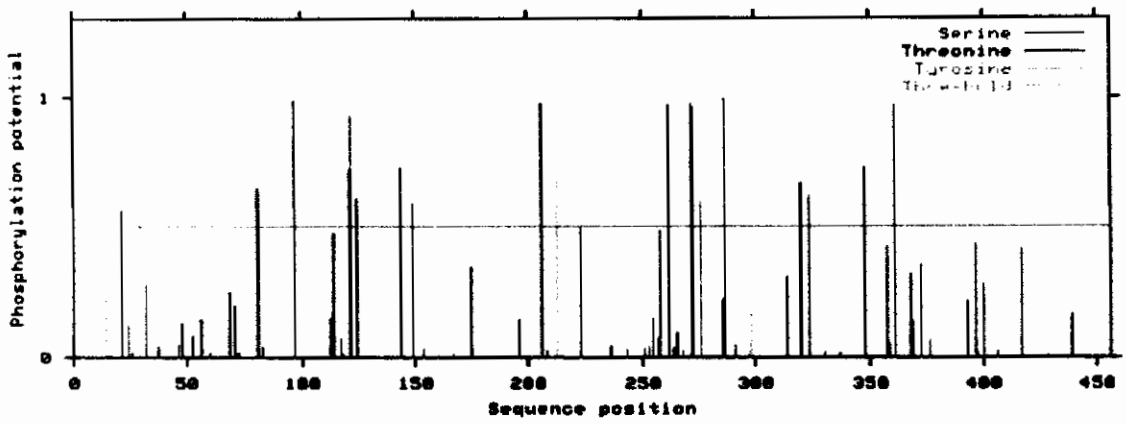


Figure 3.14: Graph showing predicted phophorylated serine, therionine and tyrosine in *LIPH*

3.1.4.2.2 Mutation Analysis:

No effect of mutations is predicted on phosphorylation of Serine, Therionine and Tyrosine in *LIPH.*

3.1.4.3 cGMP dependent protein kinases:

Cyclic GMP dependent protein kinases phosphorylation sites in *LIPH* were predicted by using kinasepho 2.0 server. Results are shown in Table 3.16.

Table 3.16: cGMPdependent protein kinases phosphorylation sites in *LIPH*

Serial No.	Residue No.	Sequence	Score	Kinase
1	9	YLFISLLCL	0.509131	PKG
2	14	LLCLSRSDA	0.505315	PKG
3	16	CLSRSDAEE	0.5	PKG
4	24	ETCPSFTRL	0.5	PKG
5	29	FTRLSFHSA	0.5	PKG
6	32	LSFHSAVVG	0.5	PKG
7	59	QTINSSAFG	0.505849	PKG
8	60	TINSSAFGN	0.506011	PKG
9	83	RPTGSPPVW	0.5	PKG
10	97	KGLLSVEDM	0.505948	PKG
11	121	YTHASSKTR	0.5	PKG
12	122	THASSKTRK	0.5	PKG
13	144	AEGASLDDI	0.507001	PKG
14	154	MIGVSLGAH	0.507669	PKG
15	160	GAHISGFVG	0.505055	PKG
16	196	RLDPSDAQF	0.506767	PKG
17	206	DVIHSDTDA	0.510418	PKG
18	251	DHQRSVYLY	0.506399	PKG
19	257	YLYLSSLRE	0.506622	PKG
20	258	LYLSSLRES	0.505535	PKG
21	262	SLRESCTIT	0.505339	PKG
22	272	YPCDSYQDY	0.508606	PKG
23	283	GKCVSCGTS	0.508778	PKG
24	287	SCGTSQKES	0.5	PKG
25	291	SQKESCPLL	0.5	PKG
26	324	TAEESPFCM	0.507842	PKG
27	361	NTTESKINH	0.508343	PKG
28	377	YHQVSLLAR	0.507638	PKG
29	393	VAAISLMFS	0.508694	PKG
30	397	SLMFSTGSL	0.507689	PKG
31	400	FSTGSLIGP	0.5	PKG
32	417	MKLRSLAHP	0.5	PKG

3.1.4.4 cAMP dependent protein kinases:

Cyclic AMP dependent protein kinases phosphorylation sites in *LIPH* were predicted by using kinasepho 2.0 server. Results are shown in Table 3.17

Table 3.17: cAMPdependent protein kinases phosphorylation sites in *LIPH*

Serial No.	Residue No.	Sequence	Score	Kinase
1	14	LLCLSRSDA	0.639261	PKA
2	16	CLSRSDAEE	0.769835	PKA
3	60	TINSSAFGN	0.549597	PKA
4	97	KGLLSVEDM	0.563966	PKA
5	196	RLDPSDAQF	0.801044	PKA

3.1.4.4 Casein kinase II phosphorylation site:

Casein kinase II phosphorylation sites in *LIPH* were predicted using kinasepho 2.0 server. Results are shown in Table 3.18

Table 3.18: Casein kinase II phosphorylation sites in *LIPH*

Serial No.	Residue No.	Sequence	Score	Kinase
1	16	CLSRSDAEE	0.794368	CK2
2	56	TCAQTINSS	0.622937	CK2
3	97	KGLLSVEDM	0.783815	CK2
4	124	ASSKTRKVA	0.533266	CK2
5	144	AEGASLDDI	0.881276	CK2
6	206	DVIHSDTDA	0.933153	CK2
7	258	LYLSSLRES	0.70737	CK2
8	272	YPCDSYQDY	0.649579	CK2
9	287	SCGTSQKES	0.687625	CK2
10	320	AFFDTAEES	0.614511	CK2
11	359	AGNTTESKI	0.710343	CK2

3.1.4.5 C-Mannosylation Site:

To predict C-mannosylation site NetCGlyc 1.0 Server was used but no C-mannosylation sites were predicted it means there is no C-Mannosylation site in *P2RY5* and *LIPH.*

3.1.4.6 Leucine Zipper pattern:

Leucine zipper is regarded as DNA binding protein and is characterized by a sporadic recurrence of leucine residues at every seventh position over a distance spanning eight helical turns. (Landschulz *et al.*, 1998). Leu–X_6–Leu–X_6–Leu–X_6–Leu is the repetition pattern of leucine zipper. No potential leucine zipper pattern was retrieved by using 2ZIP server in *P2RY5*.

3.1.4.6 .1 LEUCINE REPEATS which do NOT correspond to a Leucine Zipper:

```
1---------11--------21--------31--------41--------51--------
MLRFYLFISLLCLSRSDAEETCPSFTRLSFHSAVVGTGLNVRLMLYTRKNLTCAQTINSS

61--------71--------81--------91--------101-------111-------
AFGNLNVTKKTTFIVHGFRPTGSPPVWMDDLVKGLLSVEDMNVVVVDWNRGATTLIYTHA

121-------131-------141-------151-------161-------171-------
SSKTRKVAMVLKEFIDQMLAEGASLDDIYMIGVSLGAHISGFVGEMYDGWLGRITGLDPA

181-------191-------201-------211-------221-------231-------
GPLFNGKPHQDRLDPSDAQFVDVIHSDTDALGYKEPLGNIDFYPNGGLDQPGCPKTILGG

241-------251-------261-------271-------281-------291-------
FQYFKCDHQRSVYLYLSSLRESCTITAYPCDSYQDYRNGKCVSCGTSQKESCPLLGYYAD

301-------311-------321-------331-------341-------351-------
NWKDHLRGKDPPMTKAFFDTAEESPFCMYHYFVDIITWNKNVRRGDITIKLRDKAGNTTE

361-------371-------381-------391-------401-------411-------
SKINHEPTTFQKYHQVSLLARFNQDLDKVAAISLMFSTGSLIGPRYKLRILRMKLRSLAH
                                  L------L------L------L
421-------431-------441-------451
PERPQLCRYDLVLMENVETVFQPILCPELQL
```

3.2 Topology Prediction:

Topology can be thought of as a critical intermediate stage between amino acid sequence and 3D structure i.e. secondary structure (Heijne, 2006). It involves linking and arrangement of 2D structure elements into 3D space. Topology of structure does not change when structure of a protein is damaged. Topology of transmembrane protein is crucial to predict. It is more convenient and accurate than resolving 2D structure through X-ray Crrystallography because

membrane proteins are difficult to crystallize (Tusnaady and Simon, 1998). For toplology prediction HMMTop was used.

3.2.1 Topology of *P2RY5:*

Results of HMMTop for *P2RY5* are shown in table 3.19

Table 3.19: results of HMM top for *P2RY5*

N-Terminus	Transmembrane helices	Total entropy of the model	Entropy of the best path
OUT	7	17.0104	17.0124

3.2.1.1 Best Path:

```
seq MVSVNSSHCF YNDSFKYTLY GCMFSMVFVL GLISNCVAIY IFICVLKVRN  50
pred OOOOOOOooo oooooooooo ooHHHHHHHH HHHHHHHHHH HHHHHHiiii
seq ETTTYMINLA MSDLLFVFTL PFRIFYFTTR NWPFGDLLCK ISVMLFYTNM 100
pred iiHHHHHHHH HHHHHHHHHH HHoooooooo oooooooooo ooHHHHHHHH
seq YGSILFLTCI SVDRFLAIVY PFKSKTLRTK RNAKIVCTGV WLTVIGGSAP 150
pred HHHHHHHHHH HHiiiiii iiiiiiii iiiiHHHHHH HHHHHHHHHH
seq AVFVQSTHSQ GNNASEACFE NFPEATWKTY LSRIVIFIEI VGFFIPLILN 200
pred HHHHoooooo oooooooooo oooooooooo ooooHHHHHH HHHHHHHHHH
seq VTCSSMVLKT LTKPVTLSRS KINKTKVLKM IFVHLIIFCF CFVPYNINLI 250
pred HHHHHHHHii iiiiiiii iiiiiiii HHHHHHHHHH HHHHHHHHHH
seq LYSLVRTQTF VNCSVVAAVR TMYPITLCIA VSNCCFDPIV YYFTSDTIQN 300
pred HHHHoooooo oooooooooo ooooHHHHHH HHHHHHHHHH HHHHiiiii
seq SIKMKNWSVR RSDFRFSEVH GAENFIQHNL QTLKSKIFDN ESAA       344
pred iiiiiiiiI IIIIIIII IIIIIIII IIIIIIII IIII
```

3.2.1.2 Mutation Analysis:

Effect of mutations on Topology of *P2RY5* is shown in Table 3.20(a) to Table 3.20(b).

3.2.1.2 (a). p.D63V:

Table 3.20 (a): Topology of Mutated *P2RY5*

N-Terminus	Transmembrane helices	Total entropy of the model	Entropy of the best path
OUT	7	17.0096	17.0115

3.2.1.2 (a).1 The best path:

```
seq  MVSVNSSHCF YNDSFKYTLY GCMFSMVFVL GLISNCVAIY IFICVLKVRN    50
pred OOOOOOOooo oooooooooo ooHHHHHHHH HHHHHHHHHH HHHHHHiiii

seq  ETTTYMINLA MSVLLFVFTL PFRIFYFTTR NWPFGDLLCK ISVMLFYTNM   100
pred iiHHHHHHHH HHHHHHHHHH HHoooooooo oooooooooo ooHHHHHHHH

seq  YGSILFLTCI SVDRFLAIVY PFKSKTLRTK RNAKIVCTGV WLTVIGGSAP   150
pred HHHHHHHHHH HHiiiiiiii iiiiiiiiii iiiiHHHHHH HHHHHHHHHH

seq  AVFVQSTHSQ GNNASEACFE NFPEATWKTY LSRIVIFIEI VGFFIPLILN   200
pred HHHHoooooo oooooooooo oooooooooo ooooHHHHHH HHHHHHHHHH

seq  VTCSSMVLKT LTKPVTLSRS KINKTKVLKM IFVHLIIFCF CFVPYNINLI   250
pred HHHHHHHHii iiiiiiiiii iiiiiiiiii HHHHHHHHHH HHHHHHHHHH

seq  LYSLVRTQTF VNCSVVAAVR TMYPITLCIA VSNCCFDPIV YYFTSDTIQN   300
pred HHHHoooooo oooooooooo ooooHHHHHH HHHHHHHHHH HHHHiiiiii

seq  SIKMKNWSVR RSDFRFSEVH GAENFIQHNL QTLKSKIFDN ESAA         344
pred iiiiiiiiiI IIIIIIIIII IIIIIIIIII IIIIIIIIII IIII
```

3.2.1.2 (b). p.E189K:

Table 3.20 (b): Topology of Mutated *P2RY5*

N-Terminus	Transmembrane helices	Total entropy of the model	Entropy of the best path
OUT	7	17.0104	17.0123

3.2.1.2 (b).1 The best path:

```
 seq  MVSVNSSHCF YNDSFKYTLY GCMFSMVFVL GLISNCVAIY IFICVLKVRN    50
pred OOOOOOOooo oooooooooo ooHHHHHHHH HHHHHHHHHH HHHHHHiiii

seq  ETTTYMINLA MSDLLFVFTL PFRIFYFTTR NWPFGDLLCK ISVMLFYTNM   100
pred iiHHHHHHHH HHHHHHHHHH HHoooooooo oooooooooo ooHHHHHHHH

seq  YGSILFLTCI SVDRFLAIVY PFKSKTLRTK RNAKIVCTGV WLTVIGGSAP   150
pred HHHHHHHHHH HHiiiiiiii iiiiiiiiii iiiiHHHHHH HHHHHHHHHH

seq  AVFVQSTHSQ GNNASEACFE NFPEATWKTY LSRIVIFIKI VGFFIPLILN   200
pred HHHHoooooo oooooooooo oooooooooo ooooHHHHHH HHHHHHHHHH

seq  VTCSSMVLKT LTKPVTLSRS KINKTKVLKM IFVHLIIFCF CFVPYNINLI   250
pred HHHHHHHHii iiiiiiiiii iiiiiiiiii HHHHHHHHHH HHHHHHHHHH

seq  LYSLVRTQTF VNCSVVAAVR TMYPITLCIA VSNCCFDPIV YYFTSDTIQN   300
pred HHHHcooooo oooooooooo ooooHHHHHH HHHHHHHHHH HHHHiiiiii

seq  SIKMKNWSVR RSDFRFSEVH GAENFIQHNL QTLKSKIFDN ESAA  344
pred iiiiiiiiiI IIIIIIIIII IIIIIIIIII IIIIIIIIII IIII
```

3.2.1.2 (c). p.G146R:

Table 3.20 (c): Topology of Mutated *P2RY5*

N-Terminus	No of transmembrane helices	Total entropy of the model	Entropy of the best path
OUT	7	17.0113	17.0133

3.2.1.2 (c).1 The best path:

```
 seq  MVSVNSSHCF YNDSFKYTLY GCMFSMVFVL GLISNCVAIY IFICVLKVRN    50
pred OOOOOOOooo oooooooooo ooHHHHHHHH HHHHHHHHHH HHHHHHiiii

seq  ETTTYMINLA MSDLLFVFTL PFRIFYFTTR NWPFGDLLCK ISVMLFYTNM   100
pred iiHHHHHHHH HHHHHHHHHH HHoooooooo oooooooooo ooHHHHHHHH

seq  YGSILFLTCI SVDRFLAIVY PFKSKTLRTK RNAKIVCTGV WLTVIRGSAP   150
pred HHHHHHHHHH HHiiiiiiii iiiiiiiiii iiiiHHHHHH HHHHHHHHHH

seq  AVFVQSTHSQ GNNASEACFE NFPEATWKTY LSRIVIFIEI VGFFIPLILN   200
pred HHHHoooooo oooooooooo oooooooooo ooooHHHHHH HHHHHHHHHH

seq  VTCSSMVLKT LTKPVTLSRS KINKTKVLKM IFVHLIIFCF CFVPYNINLI   250
pred HHHHHHHHii iiiiiiiiii iiiiiiiiii HHHHHHHHHH HHHHHHHHHH

seq  LYSLVRTQTF VNCSVVAAVR TMYPITLCIA VSNCCFDPIV YYFTSDTIQN   300
pred HHHHoooooo oooooooooo ooooHHHHHH HHHHHHHHHH HHHHiiiiii

seq  SIKMKNWSVR RSDFRFSEVH GAENFIQHNL QTLKSKIFDN ESAA  344
pred iiiiiiiiiI IIIIIIIIII IIIIIIIIII IIIIIIIIII IIII
```

3.2.1.2 (d). p.N248Y:

Table 3.20 (d): Topology of Mutated *P2RY5*

N-Terminus	No of transmembrane helices	Total entropy of the model	Entropy of the best path
OUT	7	17.0102	17.0121

3.2.1.2 (d).1 The best path:

```
seq  MVSVNSSHCF YNDSFKYTLY GCMFSMVFVL GLISNCVAIY IFICVLKVRN    50
pred OOOOOOOooo oooooooooo ooHHHHHHHH HHHHHHHHHH HHHHHHiiii

seq  ETTTYMINLA MSDLLFVFTL PFRIFYFTTR NWPFGDLLCK ISVMLFYTNM   100
pred iiHHHHHHHH HHHHHHHHHH HHoooooooo oooooooooo ooHHHHHHHH

seq  YGSILFLTCI SVDRFLAIVY PFKSKTLRTK RNAKIVCTGV WLTVIGGSAP   150
pred HHHHHHHHHH HHiiiiiiii iiiiiiiiii iiiiHHHHHH HHHHHHHHHH

seq  AVFVQSTHSQ GNNASEACFE NFPEATWKTY LSRIVIFIEI VGFFIPLILN   200
pred HHHHoooooo oooooooooo oooooooooo ooooHHHHHH HHHHHHHHHH

seq  VTCSSMVLKT LTKPVTLSRS KINKTKVLKM IFVHLIIFCF CFVPYNIYLI   250
pred HHHHHHHHii iiiiiiiiii iiiiiiiiii HHHHHHHHHH HHHHHHHHHH

seq  LYSLVRTQTF VNCSVVAAVR TMYPITLCIA VSNCCFDPIV YYFTSDTIQN   300
pred HHHHoooooo oooooooooo ooooHHHHHH HHHHHHHHHH HHHHiiiiii

seq  SIKMKNWSVR RSDFRFSEVH GAENFIQHNL QTLKSKIFDN ESAA  344
pred iiiiiiiiiI IIIIIIIIII IIIIIIIIII IIIIIIIIII IIII
```

3.2.1.2 (e). p.L277P:

Table 3.20(e): Topology of Mutated *P2RY5*

N-Terminus	No of transmembrane helices	Total entropy of the model	Entropy of the best path
OUT	7	17.0111	17.0132

3.2.1.2 (e). 1 The best path:

```
 seq  MVSVNSSHCF YNDSFKYTLY GCMFSMVFVL GLISNCVAIY IFICVLKVRN    50
pred OOOOOOOooo oooooooooo ooHHHHHHHH HHHHHHHHHH HHHHHHiiii

seq  ETTTYMINLA MSDLLFVFTL PFRIFYFTTR NWPFGDLLCK ISVMLFYTNM   100
pred iiHHHHHHHH HHHHHHHHHH HHoooooooo oooooooooo ooHHHHHHHH

seq  YGSILFLTCI SVDRFLAIVY PFKSKTLRTK RNAKIVCTGV WLTVIGGSAP   150
pred HHHHHHHHHH HHiiiiiiii iiiiiiiiii iiiiHHHHHH HHHHHHHHHH

seq  AVFVQSTHSQ GNNASEACFE NFPEATWKTY LSRIVIFIEI VGFFIPLILN   200
pred HHHHoooooo oooooooooo oooooooooo ooooHHHHHH HHHHHHHHHH

seq  VTCSSMVLKT LTKPVTLSRS KINKTKVLKM IFVHLIIFCF CFVPYNINLI   250
pred HHHHHHHHii iiiiiiiiii iiiiiiiiii HHHHHHHHHH HHHHHHHHHH

seq  LYSLVRTQTF VNCSVVAAVR TMYPITPCIA VSNCCFDPIV YYFTSDTIQN   300
pred HHHHoooooo oooooooooo ooooHHHHHH HHHHHHHHHH HHHHiiiiii

seq  SIKMKNWSVR RSDFRFSEVH GAENFIQHNL QTLKSKIFDN ESAA  344
pred iiiiiiiiiI IIIIIIIIII IIIIIIIIII IIIIIIIIII IIII
```

3.2.2 *LIPH*:

Result of HMMTOP for *LIPH* is shown in Table 3.21.

Table 3.21: results of HMM top for *LIPH*

N-Terminus	No of transmembrane helices	Total entropy of the model	Entropy of the best path
OUT	7	17.0242	17.0244

3.2.2.1 Best Path:

```
 seq  MLRFYLFISL LCLSRSDAEE TCPSFTRLSF HSAVVGTGLN VRLMLYTRKN    50
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq  LTCAQTINSS AFGNLNVTKK TTFIVHGFRP TGSPPVWMDD LVKGLLSVED   100
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq  MNVVVVDWNR GATTLIYTHA SSKTRKVAMV LKEFIDQMLA EGASLDDIYM   150
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq  IGVSLGAHIS GFVGEMYDGW LGRITGLDPA GPLFNGKPHQ DRLDPSDAQF   200
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq  VDVIHSDTDA LGYKEPLGNI DFYPNGGLDQ PGCPKTILGG FQYFKCDHQR   250
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq  SVYLYLSSLR ESCTITAYPC DSYQDYRNGK CVSCGTSQKE SCPLLGYYAD   300
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq  NWKDHLRGKD PPMTKAFFDT AEESPFCMYH YFVDIITWNK NVRRGDITIK   350
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq  LRDKAGNTTE SKINHEPTTF QKYHQVSLLA RFNQDLDKVA AISLMFSTGS   400
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq  LIGPRYKLRI LRMKLRSLAH PERPQLCRYD LVLMENVETV FQPILCPELQ   450
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO
```

3.2.2.2 Mutation Analysis:

Effect of the two reported mutations of *LIPH* is shown in Table 3.22 (a) and 3.22 (b).

3.2.2.2(a) W108R

Table 3.22 (a): Topology of Mutated *LPIH*

N-Terminus	No of transmembrane helices	Total entropy of the model	Entropy of the best path
OUT	0	17.0240	17.0241

3.2.2.2(a) .1 The best path:

```
 seq  MLRFYLFISL LCLSRSDAEE TCPSFTRLSF HSAVVGTGLN VRLMLYTRKN    50
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq  LTCAQTINSS AFGNLNVTKK TTFIVHGFRP TGSPPVWMDD LVKGLLSVED   100
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq  MNVVVVDRNR GATTLIYTHA SSKTRKVAMV LKEFIDQMLA EGASLDDIYM   150
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq  IGVSLGAHIS GFVGEMYDGW LGRITGLDPA GPLFNGKPHQ DRLDPSDAQF   200
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq  VDVIHSDTDA LGYKEPLGNI DFYPNGGLDQ PGCPKTILGG FQYFKCDHQR   250
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq  SVYLYLSSLR ESCTITAYPC DSYQDYRNGK CVSCGTSQKE SCPLLGYYAD   300
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq  NWKDHLRGKD PPMTKAFFDT AEESPFCMYH YFVDIITWNK NVRRGDITIK   350
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq  LRDKAGNTTE SKINHEPTTF QKYHQVSLLA RFNQDLDKVA AISLMFSTGS   400
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq  LIGPRYKLRI LRMKLRSLAH PERPQLCRYD LVLMENVETV FQPILCPELQ   450
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq  L  451
pred O
```

3.2.2.2(b) M1T:

Table 3.22 (b): Topology of Mutated *LIPH*

N-Terminus	No of transmembrane helices	Total entropy of the model	Entropy of the best path
OUT	0	17.0238	17.0240

3.2.2.2(b).1 The best path:

```
seq  TLRFYLFISL LCLSRSDAEE TCPSFTRLSF HSAVVGTGLN VRLMLYTRKN    50
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq  LTCAQTINSS AFGNLNVTKK TTFIVHGFRP TGSPPVWMDD LVKGLLSVED   100
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq  MNVVVVDWNR GATTLIYTHA SSKTRKVAMV LKEFIDQMLA EGASLDDIYM   150
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq  IGVSLGAHIS GFVGEMYDGW LGRITGLDPA GPLFNGKPHQ DRLDPSDAQF   200
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq  VDVIHSDTDA LGYKEPLGNI DFYPNGGLDQ PGCPKTILGG FQYFKCDHQR   250
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq  SVYLYLSSLR ESCTITAYPC DSYQDYRNGK CVSCGTSQKE SCPLLGYYAD   300
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq  NWKDHLRGKD PPMTKAFFDT AEESPFCMYH YFVDIITWNK NVRRGDITIK   350
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq  LRDKAGNTTE SKINHEPTTF QKYHQVSLLA RFNQDLDKVA AISLMFSTGS   400
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq  LIGPRYKLRI LRMKLRSLAH PERPQLCRYD LVLMENVETV FQPILCPELQ   450
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq  L  451
pred O
```

3.4 Secondary structure prediction:

Secondary structure is important to be predicted as it determines tertiary structure of a protein. HNN was used to predict secondary structure of *P2RY5* and *LIPH*. Reported mutations were inserted to analyze effect of mutation on secondary structure of protein.

3.4.1 Secondary structure of *P2RY5*:

Predicted secondary structure of *P2RY5* is shown in Table 3.23 and Figure 3.15(a).Graph in Figure 3.15(b) is representing helices, extended strands and coils. Secondary structure of P2Y5 is represented by alpha helices and extended strands. There is no other form of secondary structure present in *P2Y5*.

```
MVSVNSSHCFYNDSFKYTLYGCMFSMVFVLGLISNCVAIYIFICVLKVRNETTTYMINLAMSDLLFVFTL
ceeccccceccccchhhhhhhhhhhhhhhhhhhhhhhhehhhheeccccchhhhhhhhhhhhhhhhcc
PFRIFYFTTRNWPFGDLLCKISVMLFYTNMYGSILFLTCISVDRFLAIVYPFKSKTLRTKRNAKIVCTGV
ccceeeeeccccchhhhhhhhhhhhhhhcccceeeeeeechhhhheecccccccchccccceeeeeee
WLTVIGGSAPAVFVQSTHSQGNNASEACFENFPEATWKTYLSRIVIFIEIVGFFIPLILNVTCSSMVLKT
eeeeecccceeeeeccccccccchhhhhchhhhhhhhhhhhhhhhhhhhhhhhhhhcccccheehhh
LTKPVTLSRSKINKTKVLKMIFVHLIIFCFCFVPYNINLILYSLVRTQTFVNCSVVAAVRTMYPITLCIA
ccccecccccchhhhhhhhhhhheeeeeeccccchhhhhhhhhcccccchchhhhhhhhhcccceeeeee
VSNCCFDPIVYYFTSDTIQNSIKMKNWSVRRSDFRFSEVHGAENFIQHNLQTLKSKIFDNESAA
Eccccccceeeeeccchhchheeccccccccccccccccchhhhhhchhhhhhheeccccccc
```

Table 3.23: Results of HNNfor *P2RY5*

Structure	Occurrences	Percentage
Alpha helix (Hh)	151	43.90%
3_{10} helix (Gg)	0	0.0%
Pi helix (Ii)	0	0.0%
Beta bridge (Bb)	0	0.0%
Extended strand (Ee)	65	18.90%
Beta turn (Tt)	0	0.0%
Bend region (Ss)	0	0.0%
Random coil (Cc)	128	37.21%

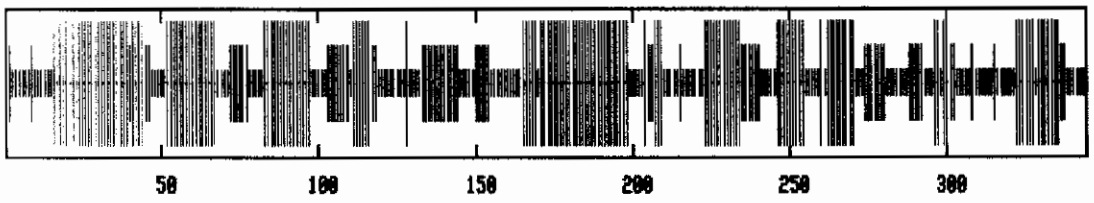


Figure 3.15 (a): Diagram indicating locations of helices, sheets and coils in *P2RY5* secondary structure

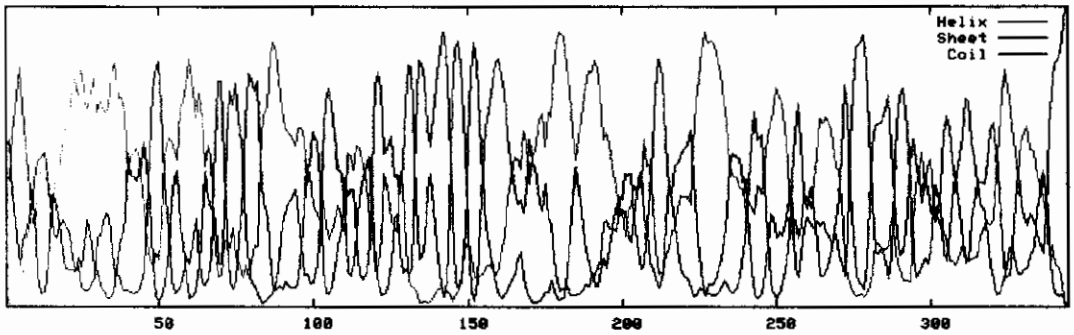


Figure 3.15(b): graph showing helices, sheets and coils in *P2RY5* secondary structure

3.4.1.1 Mutation Analysis:

Effect of 6 reported missense mutations in P2Y5 on its secondary structure was checked. Table 3.24(a) to Table 3.24(f). Figure 3.16 (a) (b) to Figure 3.20(a) (b).

3.4.1.1(a). Aspartic acid to valine (p.D63V):

```
MVSVMSSHCFYMDSFKYTLYGCMFSMVFVLGLISNCVAIYIFICVLKVRNETTTYMINLAMSVLLFVFTL
ceeccccccecccccchhhhhhhhhhhhhhhhhhhhhhhhhhhehhhheeccccchhhhhhhhhhhhhhhcc
PFRIFYFTTRNWPFGDLLCKISVMLFYTNMYGSILFLTCISVDRFLAIVYPFKSKTLRTKRNAKIVCTGV
cceeeeeeccccchhhhhhhhhhhhhhhhccccceeeeeeechhhhhheecccccccccchccccceeeeeee
WLTVIGGSAPAVFVQSTHSQGNNASEACFENFPEATWKTYLSRIVIFIEIVGFFIPLILNVTCSSMVLKT
eeeeecccccceeeeeccccccccchhhhhchhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccheehhh
LTKPVTLSRSKINKTKVLKMIFVHLIIFCFCFVPYNIMLILYSLVRTQTFVNCSVVAAVRTMYPITLCIA
ccccceccccccchhhhhhhhhhhhheeeeeeccccchhhhhhhhhhcccccchchhhhhhhhhhcccceeeeee
VSNCCFDPIVYYFTSDTIQNSIKMKNWSVRRSDFRFSEVHGAENFIQHNLQTLKSKIFDNESAA
Ecccccccceeeeecchhchheeccccceccccccecccccchhhhhhchhhhhhheeccccccc
```

Table 3.24 (a): HNN result for Mutated Secondary Structure

Structure	Occurrences	Percentage
Alpha helix (Hh)	151	43.90%
Extended strand (Ee)	65	18.90%
Random coil (Cc)	128	37.21%

No effect of this mutation on secondary structure of *P2RY5* is predicted by HNN.

3.4.1.1(b). Glutamate to lysine (p.E189K):

```
MVSVNSSHCFYNDSFKYTLYGCMFSMVFVLGLISNCVAIYIFICVLKVRNETTTYMINLAMSDLLFVFTL
ceeccccccecccccchhhhhhhhhhhhhhhhhhhhhhhhehhhheecccccchhhhhhhhhhhhhhhhcc
PFRIFYFTTRNWPFGDLLCKISVMLFYTNMYGSILFLTCISVDRFLAIVYPFKSKTLRTKRNAKIVCTGV
cceeeeeeccccchhhhhhhhhhhhhhhhcccccceeeeeeechhhhhheecccccccccchccccceeeeeee
WLTVIGGSAPAVFVQSTHSQGNNASEACFENFPEATWKTYLSRIVIFIKIVGFFIPLILNVTCSSMVLKT
eeeeeccccceeeeeccccccccccchhhhhcchhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccheehhh
LTKPVTLSRSKINKTKVLKMIFVHLIIFCFCFVPYNINLILYSLVRTQTFVNCSVVAAVRTMYPITLCIA
ccccceccccccchhhhhhhhhhhheeeeeeccccchhhhhhhhhcccccchchhhhhhhhhccceeeeeee
VSNCCFDPIVYYFTSDTIQNSIKMKNWSVRRSDFRFSEVHGAENFIQHNLQTLKSKIFDNESAA
Ecccccccceeeeeccchhchheeccccceccccccceccccchhhhhhchhhhhhheeccccccc
```

Table 3.24 (b): HNN result for Mutated Secondary Structure

Structure	Occurrences	Percentage
Alpha helix (Hh)	150	43.60%
Extended strand (Ee)	65	18.90%
Random coil (Cc)	129	37.50%

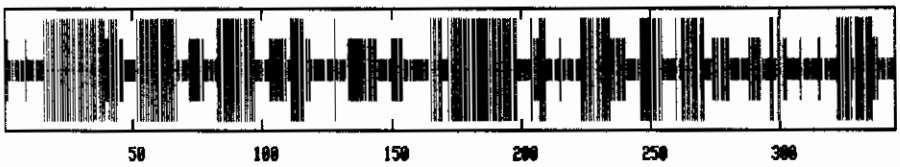


Figure 3.16 (a): Diagram indicating locations of helices, sheets and coils in mutated *P2RY5* secondary structure

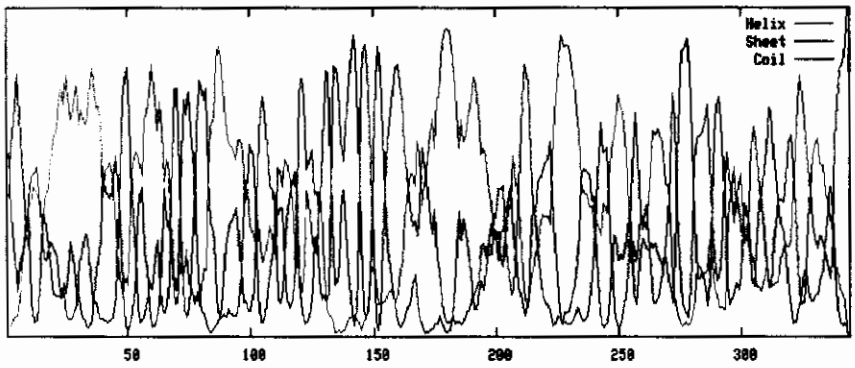


Figure 3.16(b): Graph showing helices, sheets and coils in mutated *P2RY5* secondary structure

This mutation results in the replacement of glutamic acid with lysine at position number 189. glutamic acid participates in formation of alpha helix in *P2RY5*, replaced amino acid is also making alpha helix but at position number 172 phenyl alanine that was making an alpha helix is now part of random coil. Random coil is not considered as true secondary structure so; phenyl alanine is no longer participating in secondary structure. P2Y5 is a member of GPCR. Through interaction with Guanine binding proteins, these receptors transduce extra cellular signals. *P2RY5* is a member of purine and pyrimidine nucleotides receptors family. Alpha helices are crucial for binding of particular protein with nucleotides. This mutated structure might have reduced binding with Guanine binding proteins which leads to reduction in trasnduced extra cellular signals.

3.4.1.1(c). Serine to therionine (p.S3T):

```
MVTVNSSHCFYNDSFKYTLYGCMFSMVFVLGLISNCVAIYIFICVLKVRNETTTYMINLAMSDLLFVFTL
ceeeccccceccccchhhhhhhhhhhhhhhhhhhhhhhhhehhhheeccccchhhhhhhhhhhhhhhcc
PFRIFYFTTRNWPFGDLLCKISVMLFYTNMYGSILFLTCISVDRFLAIVYPFKSKTLRTKRNAKIVCTGV
cceeeeeeccccchhhhhhhhhhhhhhhhcccccceeeeeeechhhhhheeccccccccchccccceeeeeee
WLTVIGGSAPAVFVQSTHSQGNNASEACFENFPEATWKTYLSRIVIFIEIVGFFIPLILNVTCSSMVLKT
eeeeeccccceeeeeccccccccchhhhhchhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccheehhh
LTKPVTLSRSKINKTKVLKMIFVHLIIFCFCFVPYNINLILYSLVRTQTFVNCSVVAAVRTMYPITLCIA
ccccceccccccchhhhhhhhhhhheeeeeeccccchhhhhhhhhcccccchchhhhhhhhhcccceeeeee
VSNCCFDPIVYYFTSDTIQNSIKMKNWSVRRSDFRFSEVHGAENFIQHNLQTLKSKIFDNESAA
Ecccccccceeeeeccchhchheeccccceccccccceccccchhhhhhchhhhhhheeccccc
```

Table 3.24 (c): HNN result for Mutated Secondary Structure

Structure	Occurrences	Percentage
Alpha helix (Hh)	151	43.90%
Extended strand (Ee)	66	19.19%
Random coil (Cc)	127	36.92%

This mutation results in the replacement of serine with therionine at position number 3. Serine participates in formation of extended strand in *P2RY5*, replaced amino acid is also making extended strand but at position number 4 valine, which was making random coil is now making an extended strand. Random coil is not considered as true secondary structure so; valine is now participating in secondary structure. Extended strands can be associated with beta sheets or they can be isolated. Polar groups of polypeptide backbone often get satisfied by achieving alpha helices and beta sheet conformation. Isolated strands show similar characteristics shared with loops and beta sheet. (Narayanan *et al.*, 2003) valine that is now a part of extended strand might be forming hydrogen bond with some distant residue. This additional residue in extended strand might leads to changed tertiary structure.

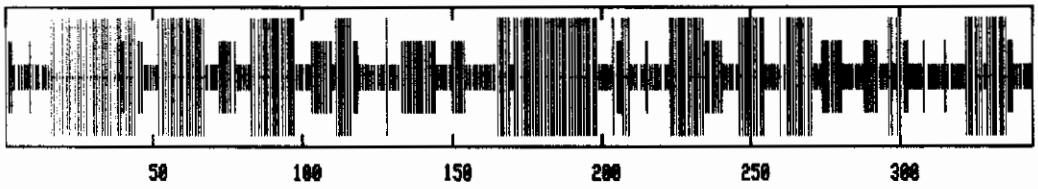


Figure 3.17 (a): Diagram indicating locations of helices, sheets and coils in mutated *P2RY5* secondary structure

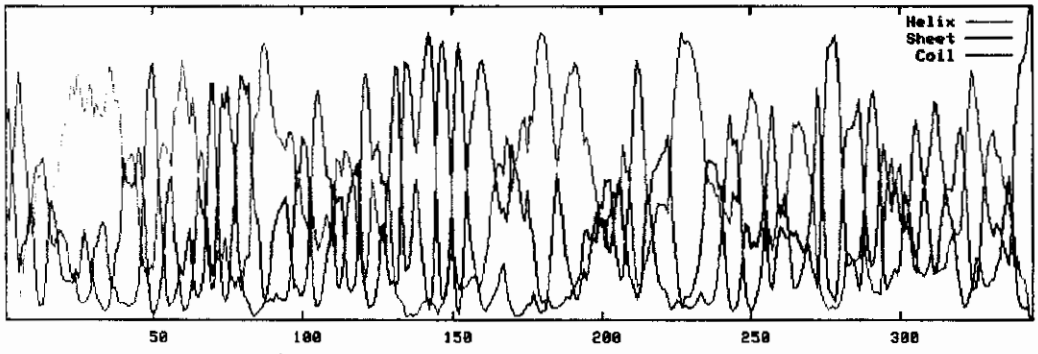


Figure 3.17(b): graph showing helices, sheets and coils in mutated *P2RY5* secondary structure

3.4.1.1(d). Glycine to arginine (p.G146R):

MVSVNSSHCFYNDSFKYTLYGCMFSMVFVLGLISNCVAIYIFICVLKVRNETTTYMINLAMSDLLFVFTL
ceeccccccecccccchhhhhhhhhhhhhhhhhhhhhhhhhehhhheecccccchhhhhhhhhhhhhhhhcc
PFRIFYFTTRNWPFGDLLCKISVMLFYTNMYGSILFLTCISVDRFLAIVYPFKSKTLRTKRNAKIVCTGV
cceeeeeeccccchhhhhhhhhhhhhhhcccccceeeeeeechhhhhheecccccccccchccccceeeeeee
WLTVIRGSAPAVFVQSTHSQGNNASEACFENFPEATWKTYLSRIVIFIEIVGFFIPLILNVTCSSMVLKT
eeeeeecccceeeeeeccccccccchhhhhchhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccheehhh
LTKPVTLSRSKINKTKVLKMIFVHLIIFCFCFVPYNINLILYSLVRTQTFVNCSVVAAVRTMYPITLCIA
ccccceccccccchhhhhhhhhhhheeeeeeccccchhhhhhhhhcccccchchhhhhhhhhcccceeeeee
VSNCCFDPIVYYFTSDTIQNSIKMKNWSVRRSDFRFSEVHGAENFIQHNLQTLKSKIFDNESAA
ecccccccceeeeeccchhchheecccccecccccceccccchhhhhhchhhhhhheeccccccc

Table 3.24 (d): HNN result for Mutated Secondary Structure

Structure	Occurrences	Percentage
Alpha helix (Hh)	151	43.90%
Extended strand (Ee)	67	19.48%
Random coil (Cc)	126	36.63%

This mutation results in conversion of glycine at position number 146 to arginine. Glycine at position number 146 in *P2RY5* is a part random coil but due to mutation, replaced amino acid is now participating in extended strand formation. Arginine that is now a part of extended strand might be forming hydrogen bond with some distant residue. This additional residue in extended strand might leads to changed tertiary structure.

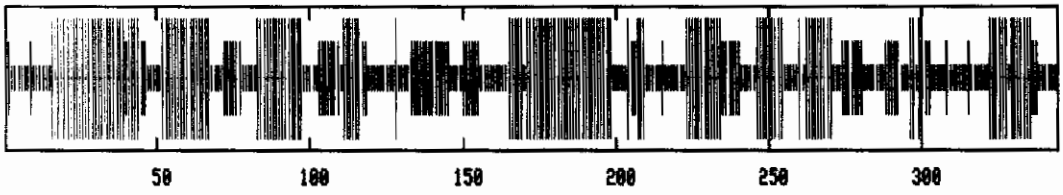


Figure 3.18 (a): Diagram indicating locations of helices, sheets and coils in mutated *P2RY5* secondary structure

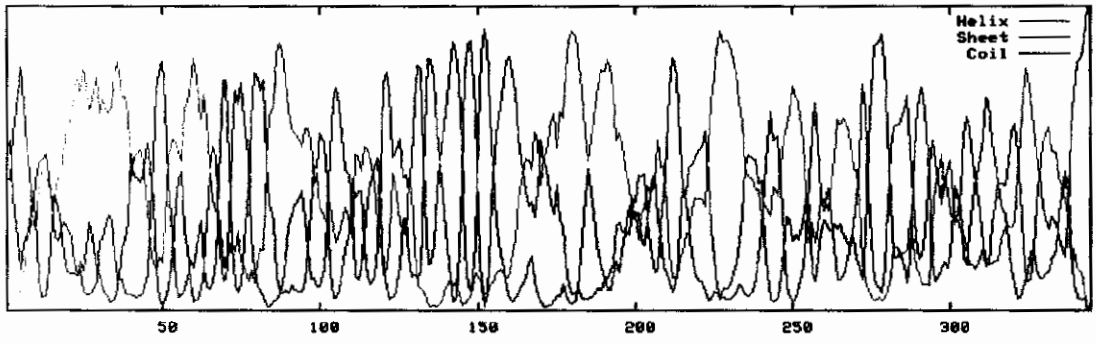


Figure 3.18(b): graph showing helices, sheets and coils in mutated *P2RY5* secondary structure

3.4.1.1(e). Arginine to Tyrosine (p.N248Y):

```
MVSVNSSHCFYNDSFKYTLYGCMFSMVFVLGLISNCVAIYIFICVLKVRNETTTYMINLAMSDLLFVFTL
ceeccccceccccchhhhhhhhhhhhhhhhhhhhhhhehhhheeccccchhhhhhhhhhhhhhhhcc
PFRIFYFTTRNWPFGDLLCKISVMLFYTNMYGSILFLTCISVDRFLAIVYPFKSKTLRTKRNAKIVCTGV
cceeeeeeccccchhhhhhhhhhhhhhhhcccceeeeeeechhhhhheecccccccchccccceeeeeee
WLTVIGGSAPAVFVQSTHSQGNNASEACFENFPEATWKTYLSRIVIFIEIVGFFIPLILNVTCSSMVLKT
eeeeeccccceeeeeccccccccchhhhhchhhhhhhhhhhhhhhhhhhhhhhhhhhcccccheehhh
LTKPVTLSRSKINKTKVLKMIFVHLIIFCFCFVPYNIYLILYSLVRTQTFVNCSVVAAVRTMYPITLCIA
ccccceccccccchhhhhhhhhhhhheeeheccchchhhhhhhhhhcccccchchhhhhhhhhcccceeeeee
VSNCCFDPIVYYFTSDTIQNSIKMKNWSVRRSDFRFSEVHGAENFIQHNLQTLKSKIFDNESAA
Ecccccccceeeeccchhchheeccccceccccceccccchhhhhhchhhhhhheeccccccc
```

Table 3.24 (e): HNN result for Mutated Secondary Structure

Structure	Occurrences	Percentage
Alpha helix (Hh)	154	44.77%
Extended strand (Ee)	63	18.31%
Random coil (Cc)	127	36.92%

This mutation results in conversion of aspargine at position number 248 to tyrosine, aspargine at position number 248 in *P2RY5* forms alpha helix, replaced amino acid is also forms alpha helix but due to this mutation Isoleucine and phenyl alanine at position number 236 and 238 respectively, which make extended strand in normal structure are now participating in alpha helix formation. Tyrosine at position number 245 that is a part of random coil is now making alpha helix. Due to this single mutation there is reduction in extended strands and increase in number of alpha helices that leads to altered tertiary structure.

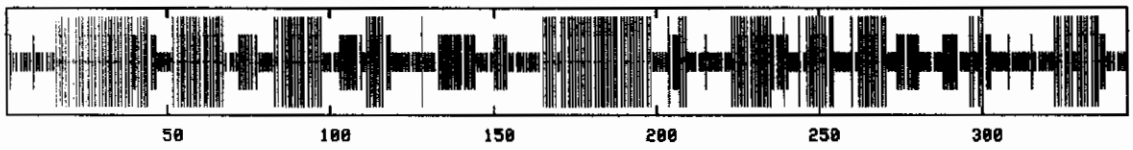


Figure 3.19 (a): Diagram indicating locations of helices, sheets and coils in mutated *P2RY5* secondary structure

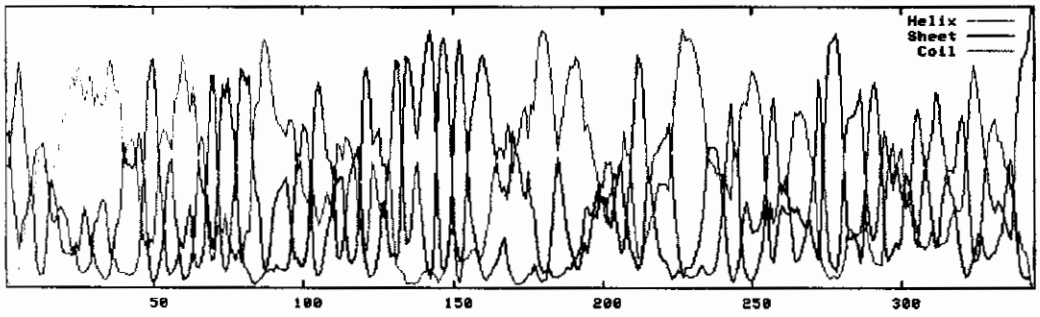


Figure 3.19(b): Graph showing helices, sheets and coils in mutated *P2RY5* secondary structure

3.4.1.1(f). Leucine to Proline (p.L277P):

MVSVNSSHCFYNDSFKYTLYGCMFSMVFVLGLISNCVAIYIFICVLKVRNETTTYMINLAMSDLLFVFTL
ceeccccccecccccchhhhhhhhhhhhhhhhhhhhhhhhhehhhheeccccchhhhhhhhhhhhhhhhcc
PFRIFYFTTRNWPFGDLLCKISVMLFYTNMYGSILFLTCISVDRFLAIVYPFKSKTLRTKRNAKIVCTGV
cceeeeeeccccchhhhhhhhhhhhhhhcccccceeeeeeechhhhheecccccccccchccccceeeeeee
WLTVIGGSAPAVFVQSTHSQGNNASEACFENFPEATWKTYLSRIVIFIEIVGFFIPLILNVTCSSMVLKT
eeeeeccccceeeeeccccccccccchhhhhchhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccheehhh
LTKPVTLSRSKINKTKVLKMIFVHLIIFCFCFVPYNINLILYSLVRTQTFVNCSVVAAVRTMYPITPCIA
ccccceccccccchhhhhhhhhhhheeeeeeccccchhhhhhhhhcccchhchhhhhhhhhhcccccceee
VSNCCFDPIVYYFTSDTIQNSIKMKNWSVRRSDFRFSEVHGAENFIQHNLQTLKSKIFDNESAA
Ecccccccceeeeeccchhchheeccccceccccccecccccchhhhhhchhhhhhheeccccc

Table 3.24 (f): HNN result for Mutated Secondary Structure

Structure	Occurrences	Percentage
Alpha helix (Hh)	153	44.48%
Extended strand (Ee)	62	18.02%
Random coil (Cc)	129	37.50%

This mutation results in conversion of leucine at position number 277 to proline, leucine at position number 277 in *P2RY5* forms extended strand, mutated amino acid is no longer participating in formation of extended but it forms random coil. This mutation also results in random coil at position 276 and 275 where therionine and isoleucine makes extended strands in normal structures. As a result of this mutation therionine and isoleucine no longer remain part of secondary structure conformation. So there is reduction in extended strands that result in altered tertiary structure.

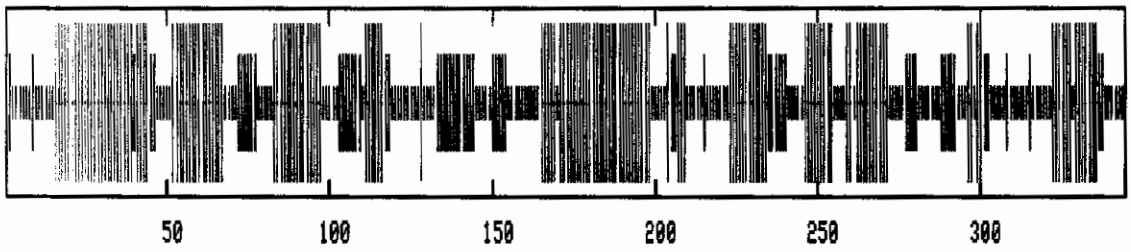


Figure 3.20 (a): Diagram indicating locations of helices, sheets and coils in mutated *P2RY5* secondary structure

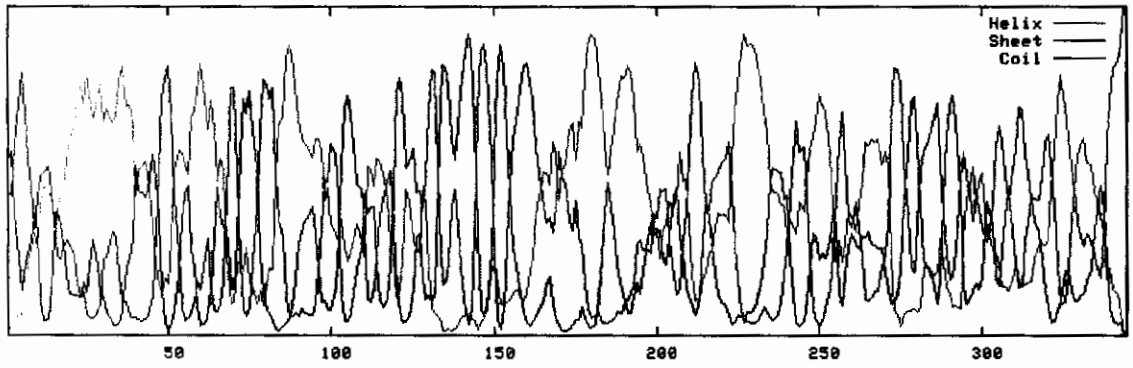


Figure 3.20(b): Graph showing helices, sheets and coils in mutated *P2RY5* secondary structure

3.4.2 LIPH:

Predicted secondary structure for *LIPH* is as follows:

```
MLRFYLFISLLCLSRSDAEETCPSFTRLSFHSAVVGTGLNVRLMLYTRKNLTCAQTINSSAFGNLNVTKK
cceeeeeeeeehhccccccccccchheeeccceeeccchheehhhhccccceeccccccccccccece
TTFIVHGFRPTGSPPVWMDDLVKGLLSVEDMNVVVVDWNRGATTLIYTHASSKTRKVAMVLKEFIDQMLA
eeeeeeccccccccccchhhhhhhhheeeccceeeeeeccccccceeeeeccccchhhhhhhhhhhhhhhhh
EGASLDDIYMIGVSLGAHISGFVGEMYDGWLGRITGLDPAGPLFNGKPHQDRLDPSDAQFVDVIHSDTDA
hcccchhhhehhhcccchhhhhhhhhhhhcceeeeccccccccccccccccccccccccceeeeeeccccc
LGYKEPLGNIDFYPNGGLDQPGCPKTILGGFQYFKCDHQRSVYLYLSSLRESCTITAYPCDSYQDYRNGK
ccccccccccccccccccccccccccccchhhheccccheeeehhccccceeeeeccccccccccccc
CVSCGTSQKESCPLLGYYADNWKDHLRGKDPPMTKAFFDTAEESPFCMYHYFVDIITWNKNVRRGDITIK
eeeccccccccccccccchhhhhhhccccccccccceccccccceeeeeeehhheeccccccccceeee
LRDKAGNTTESKINHEPTTFQKYHQVSLLARFNQDLDKVAAISLMFSTGSLIGPRYKLRILRMKLRSLAH
eeccccccccccccccccchhhhhhhhhhhhhchhhhhhheeeeehccccccchhhhhhhhhhhhhhcc
PERPQLCRYDLVLMENVETVFQPILCPELQL
Cccchceeeeeehhchhhhhcccccccccc
```

Composition on predicted structure is shown in Table 3.25 Graph representing alpha helices, beta sheets and coils, is shown in Figure 3.21(a) And predicted secondary structure is depicted in Figure 3.21(b)

Table 3.25: HNN result for *LIPH*

Secondary structure element	Total occurrences	Percentage
Alpha helix (Hh)	117	25.94%
3_{10} helix (Gg)	0	0.0%
Pi helix (Ii)	0	0.0%
Beta bridge (Bb)	0	0.0%
Extended strand (Ee)	92	20.40%
Beta turn (Tt)	0	0.0%
Bend region (Ss)	0	0.0%
Random coil (Cc)	242	53.66%
Ambigous states (?)	0	0.0
Other states	0	0.0

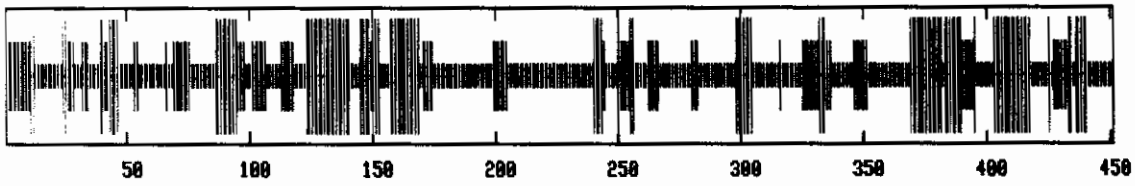


Figure 3. 21(a): Graph representing position of helices, extended strands and coils

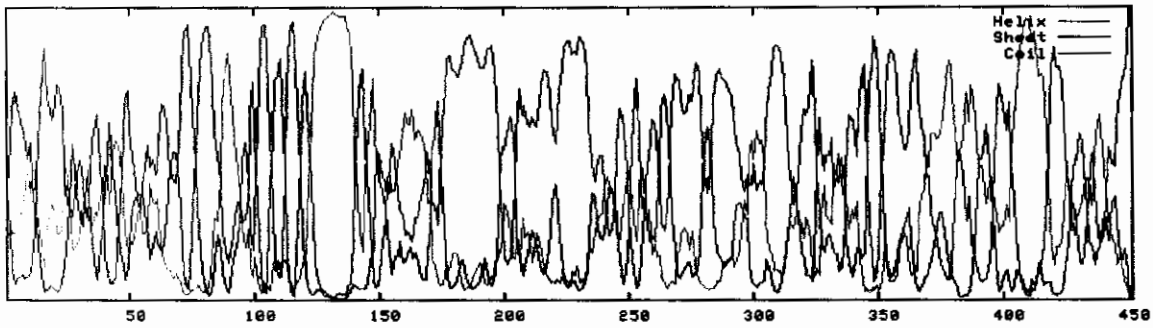


Figure 3. 21(b) : Predicted secondary structure of LIPH

3.4.2.1 Mutation analysis:

After insertion of two reported mutations results that were retrieved are shown in

Table 3.26 (a) for mutation p. W108R and Table 3.26(b) for mutation p. M1T Respectively.

(Figure 3.22(a)(b) and Figure 3.23(a)(b))

3.4.2.1 (a) p. W108R

```
MLRFYLFISLLCLSRSDAEETCPSFTRLSFHSAVVGTGLNVRLMLYTRKNLTCAQTINSSAFGNLNVTKK
ccceeeeeeeehhccccccccccchheeeccceeeccccchheehhhhccccceeccccccccccccecce
TTFIVHGFRPTGSPPVWMDDLVKGLLSVEDMNVVVVDRNRGATTLIYTHASSKTRKVAMVLKEFIDQMLA
eeeeeeccccccccccchhhhhhhhheeeccceeeeeeccccceeeeeeccccchhhhhhhhhhhhhhhhh
EGASLDDIYMIGVSLGAHISGFVGEMYDGWLGRITGLDPAGPLFNGKPHQDRLDPSDAQFVDVIHSDTDA
hcccchhhhehhhccccchhhhhhhhhhhhccceeeeccccccccccccccccccccccccceeeeeeccccc
LGYKEPLGNIDFYPNGGLDQPGCPKTILGGFQYFKCDHQRSVYLYLSSLRESCTITAYPCDSYQDYRNGK
ccccccccccccccccccccccccccccccchhhheccccccheeeehhccccceeeeeccccccccccccc
CVSCGTSQKESCPLLGYYADNWKDHLRGKDPPMTKAFFDTAEESPFCMYHYFVDIITWNKNVRRGDITIK
eeeccccccccccccccchhhhhhhccccccccccccecccccccceeeeeeehhheeccccccccceeee
LRDKAGNTTESKINHEPTTFQKYHQVSLLARFNQDLDKVAAISLMFSTGSLIGPRYKLRILRMKLRSLAH
eeccccccccccccccccchhhhhhhhhhhhhchhhhhhheeeeehccccccchhhhhhhhhhhhhhcc
PERPQLCRYDLVLMENVETVFQPILCPELQL
ccccchceeeeeehhchhhhhccccccccc
```

Table 3.26 (a): Secondary structure composition of mutated structure

Structure	Occurrences	Percentage
Alpha helix (Hh)	117	25.94%
Extended strand (Ee)	93	18.90%
Random coil (Cc)	241	53.44%

This mutation results in conversion of tryptophan at position number 108 to arginine. Due to this mutation therionine at position number 113 that is not a part of secondary structure in normal LIPH is now participating in formation of extended strand. So extended strand is now starting at 108 locations instead of 109 and it is 6 residues long instead of five. This extended strand of 6 will affect tertiary structure.

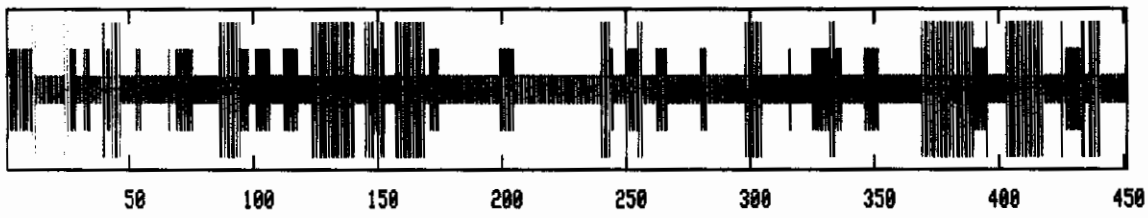


Figure 3.22 (a): Graph representing position of helices, extended strands and coils of mutated LIPH

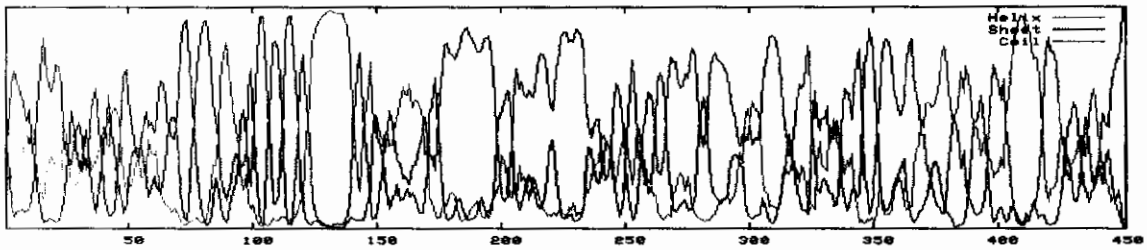


Figure 3.22 (b): Predicted secondary structure of mutated LIPH

3.4.2.1 (b) p. M1T:

```
TLRFYLFISLLCLSRSDAEETCPSFTRLSFHSAVVGTGLNVRLMLYTRKNLTCAQTINSSAFGNLNVTKK
cceeeeeeeeeehccccccccccchheeeccceeeccchheehhhhccccceeccccccccccecce
TTFIVHGFRPTGSPPVWMDDLVKGLLSVEDMNVVVVDWNRGATTLIYTHASSKTRKVAMVLKEFIDQMLA
eeeeeeccccccccccchhhhhhhhheeeccceeeeeeccccccceeeeeccccchhhhhhhhhhhhhhhhh
EGASLDDIYMIGVSLGAHISGFVGEMYDGWLGRITGLDPAGPLFNGKPHQDRLDPSDAQFVDVIHSDTDA
hcccchhhhehhhccccchhhhhhhhhhhhcceeeeccccccccccccccccccccccccceeeeeeccccc
LGYKEPLGNIDFYPNGGLDQPGCPKTILGGFQYFKCDHQRSVYLYLSSLRESCTITAYPCDSYQDYRNGK
ccccccccccccccccccccccccccccchhhheccccheeeehhccccceeeeeccccccccccccc
CVSCGTSQKESCPLLGYYADNWKDHLRGKDPPMTKAFFDTAEESPFCMYHYFVDIITWNKNVRRGDITIK
eeeccccccccccccccchhhhhhhccccccccccceccccccceeeeeeehhheeccccccccceeee
LRDKAGNTTESKINHEPTTFQKYHQVSLLARFNQDLDKVAAISLMFSTGSLIGPRYKLRILRMKLRSLAH
eeccccccccccccccccchhhhhhhhhhhhhchhhhhhheeeeehccccccchhhhhhhhhhhhhhhcc
PERPQLCRYDLVLMENVETVFQPILCPELQL
Ccccchceeeeeehhchhhhhccccccccc
```

Table 3.26 (b): Secondary structure composition of mutated structure

Structure	Occurrences	Percentage
Alpha helix (Hh)	116	25.72%
Extended strand (Ee)	93	20.62%
Random coil (Cc)	242	53.66%

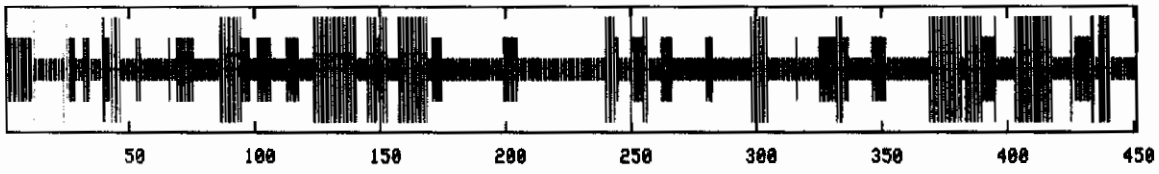


Figure 3.23 (a): Graph representing position of helices, extended strands and coils of mutated LIPH

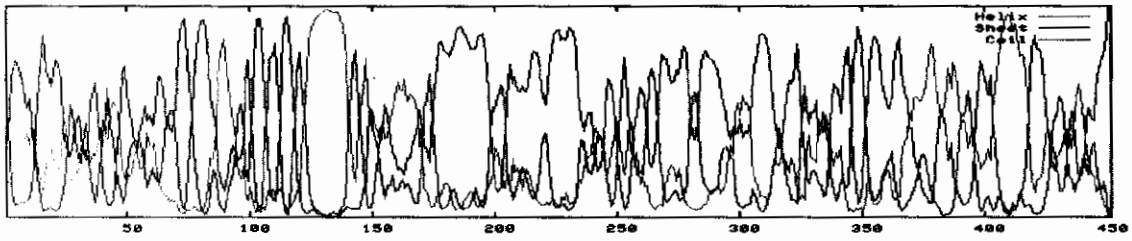


Figure 3.23 (b): Predicted secondary structure of mutated *LIPH*

This mutation results in conversion of methionine at position number 1 to therionine. This mutation results in decrease in alpha helices within protein structure. Cystien at position number 12 makes alpha helix in normal LIPH but due to this mutation it now making random coil.

3.5 Physiochemical Properties:

3.5.1 *P2RY5*:

Mutations also have effects on physio-chemical properties of proteins. Physio-chemical properties of normal *P2RY5* are shown in table 3.27 while table 3.28(a) to Table 3.28(f) shows properties of mutated Protein.

Table 3.27: Physio-chemical parameters of *P2RY5*

No. of amino acids	Molecular weight	Theoretical pI	Formula	Total no. of atoms
344	39391.5	9.22	$C_{1821}H_{2811}N_{443}O_{478}S_{26}$	5579

3.5.1.1 Mutation Analysis:

3.5.1.1(a)Aspartic acid to valine (p.D63V):

Table 3.28 (a): Physio-chemical parameters of mutated *P2RY5*

No. of amino acids	Molecular weight	Theoretical pI	Formula	Total no. of atoms
344	39375.6	9.27	$C_{1822}H_{2815}N_{443}O_{476}S_{26}$	5582

3.5.1.1(b)Glutamate to lysine (p.E189K):

Table 3.28(b): Physio-chemical parameters of mutated *P2RY5*

No. of amino acids	Molecular weight	Theoretical pI	Formula	Total no. of atoms
344	39390.6	9.32	$C_{1822}H_{2816}N_{444}O_{476}S_{26}$	5584

3.5.1.1(c) Serine to therionine (p.S3T):

Table 3.28 (c): Physio-chemical parameters of mutated *P2RY5*

No. of amino acids	Molecular weight	Theoretical pI	Formula	Total no. of atoms
344	39405.6	9.22	$C_{1822}H_{2813}N_{443}O_{478}S_{26}$	5582

3.5.1.1(d)Glycine to arginine (p.G146R):

Table 3.28 (d): Physio-chemical parameters of mutated *P2RY5*

No. of amino acids	Molecular weight	Theoretical pI	Formula	Total no. of atoms
344	39490.7	9.27	$C_{1825}H_{2820}N_{446}O_{478}S_{26}$	5595

3.5.1.1(e)Arginine to Tyrosine (p.N248Y):

Table 3.28 (e): Physio-chemical parameters of mutated *P2RY5*

No. of amino acids	Molecular weight	Theoretical pI	Formula	Total no. of atoms
344	39440.6	9.21	$C_{1826}H_{2814}N_{442}O_{478}S_{26}$	5586

3.5.1.1(f)Leucine to Proline (p.L277P):

Table 3.28 (f): Physio-chemical parameters of mutated *P2RY5*

No. of amino acids	Molecular weight	Theoretical pI	Formula	Total no. of atoms
344	39375.5	9.22	$C_{1820}H_{2807}N_{443}O_{478}S_{26}$	5574

From the above results, it is clear that like other properties of proteins, physio-chemical properties of proteins are also get affected by mutations.

3.5.2 *LIPH*:

Table 3.29 expresses physiochemical properties of LIPH

Table 3.29: Physio-chemical parameters of *LIPH*

Number of amino acids	Molecular weight	Theoretical pI	Formula	Total number of atoms
451	50859.3	7.15	$C_{2275}H_{3535}N_{609}O_{663}S_{26}$	7108

3.5.2.1 Mutation Analysis:

Mutations effect on physiochemical properties of *LIPH* are revealed in Table 3.30(a) and Table 3.30(b).

3.5.2.1 (a)Tryptophan to arginine (p.W108R):

Table 3.30(a): Physio-chemical parameters of mutated *LIPH*

Number of amino acids	Molecular weight	Theoretical pI	Formula	Total number of atoms
451	50829.3	7.53	$C_{2270}H_{3537}N_{611}O_{663}S_{26}$	7107

3.5.2.1 (b)Methionine to Leucine (p. M1T):

Table 3.30(b): Physio-chemical parameters of mutated *LIPH*

Number of amino acids	Molecular weight	Theoretical pI	Formula	Total number of atoms
451	50829.2	7.11	$C_{2274}H_{3533}N_{609}O_{664}S_{25}$	7105

3.5 Tertiary Structure:

3.5.1 *P2RY5*:

3D structure of *P2RY5* was intended to be predicted by comparative modeling to accomplish this similarity searching was conducted.

3.5.1.1. Similarity Searching:

Similarity search was conducted to investigate whether target sequence has similarity to other sequences or not. Sequence information can be utilized to find their relatives in order to have clues for structure and function of that particular protein. Similarity search manifests extent of similarity between input sequence and sequences that are stored in databases. Following tools for were used for alignment.

3.5.1.1. 1. BLAST

3.5.1.1. 2 FASTA3

3.5.1.1. 1. Basic Local Alignment Search Tool (BLAST):

BLAST results for *P2Y5* were obtained by selecting PDB as a database to be searched. Blast retrieved those PDB files that show similarity with *P2Y5* (Table 3.31)

Table 3.31: results of BLAST for *P2RY5*

Accession No.	Protein name	Origin of protein	E-value	Identity %
2R4SA	Chain A, Crystal Structure Of The Human Beta2 Adrenoceptor	Homo sapiens	5e-11	22
2R4RA	Chain A, Crystal Structure Of The Human Beta2 Adrenoceptor	Homo sapiens	5e-11	22
3D4SA	Chain A, Cholesterol Bound Form Of Human Beta2 Adrenergic Receptor	Unknown	4e-10	24

2RH1A	Chain A, High Resolution Crystal Structure Of Human B2-Adrenergic G Protein-Coupled Receptor.	Unknown	6e-10	24
2VT4A	Chain A, Turkey Beta1 Adrenergic Receptor With Stabilising Mutations And Bound Cyanopindolol.	Meleagris gallopavo (turkey)	4e-09	22
1JFP	Chain A, Structure Of Bovine Rhodopsin	Bos taurus	2e-07	20
3C9M|A	Chain A, Structure Of A Mutant Bovine Rhodopsin In Hexagonal Crystal Form.	Bos taurus	2e-07	20
2J4Y|A	Chain A, Crystal Structure Of A Rhodopsin Stabilizing Mutant Expressed In Mammalian Cells.	Bos taurus	2e-07	20
2Z73|A	Chain A, Crystal Structure Of Squid Rhodopsin	Todarodes pacificus	4e-04	29
2ZIY|A	Chain A, Crystal Structure Of Squid Rhodopsin.	Todarodes pacificus	4e-04	29

3.5.1.1. 2 FASTA3:

Result of FASTA for *P2RY5 is* shown in Table 3.32.

Table 3.32: results of FASTA3 for *P2RY5*

Accession No.	Protein name	Origin of protein	Sequence similarity (%)	Sequence identity (%)	e-Value
P43657, A4FTW9, B3KVF2, O15133, Q3KPF5, Q53FA0,	Oleoyl-L-alpha-lysophosphatidic acid receptor	Human	100	100	1e-157
B3KVQ5	cDNA FLJ41064 fis	Human	100.0	99.7	2.7e-157
Q8BMC0, A6H6N5, Q3TEJ6	Oleoyl-L-alpha-lysophosphatidic acid receptor	Mus musculus (Mouse)	98.3	93.3	9.8e-148
Q4G072	Oleoyl-L-alpha-lysophosphatidic acid receptor;	Rattus norvegicus (Rat)	98.3	93.3	9.8e-148
A6QL49	*P2RY5* protein	Bos taurus (Bovin)	95.3	88.9	9.1e-139

FASTA retrieved sequences stored in databases that have similarity with P2Y5. From BLAST result it is clear that no Pdb file in Protein database exists that have more than 30% identity with *P2Y5*. No template was found to predict 3D structure of *P2Y5* through comparative modeling therefore threading approach was used to predict *P2Y5* 3D structure. SAM T02was used to predict tertiary structure of protein. Predicted structure is shown in figure 3. 24

3.5.1.2. Tertiary Structure:

Tertiary structure of P2RY5 is shown in Figure 3.24

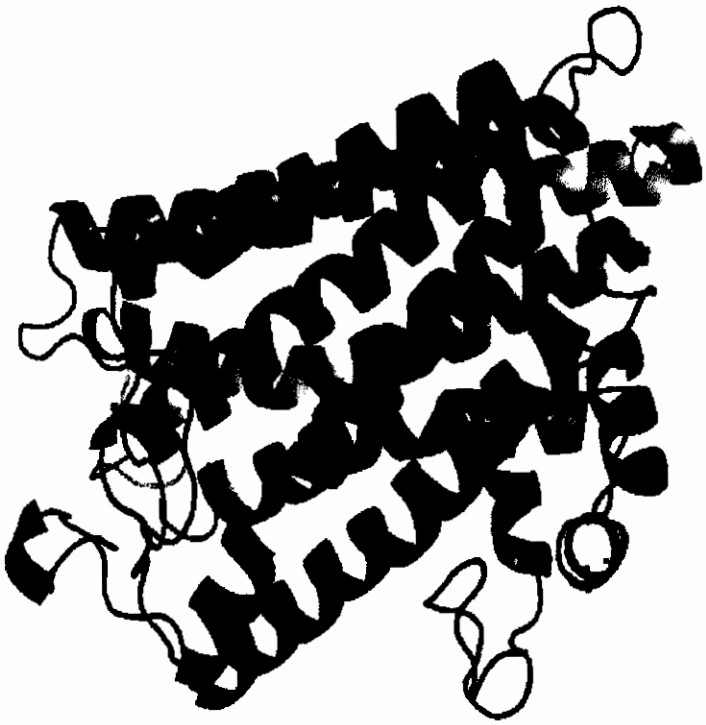


Figure 3.24: 3D Structure of normal P2Y5

3.5.1.3 Mutation Analysis:

Effect of mutations on 3D structure of *P2RY5* is shown in Figure 3.25 (a) (b) (c) to figure 3.30 (a) (b) (c). Mode of mutated amino acid have been changed to make effect of mutation more prominent.

3.5.1.2. 1(a). p.D63V:

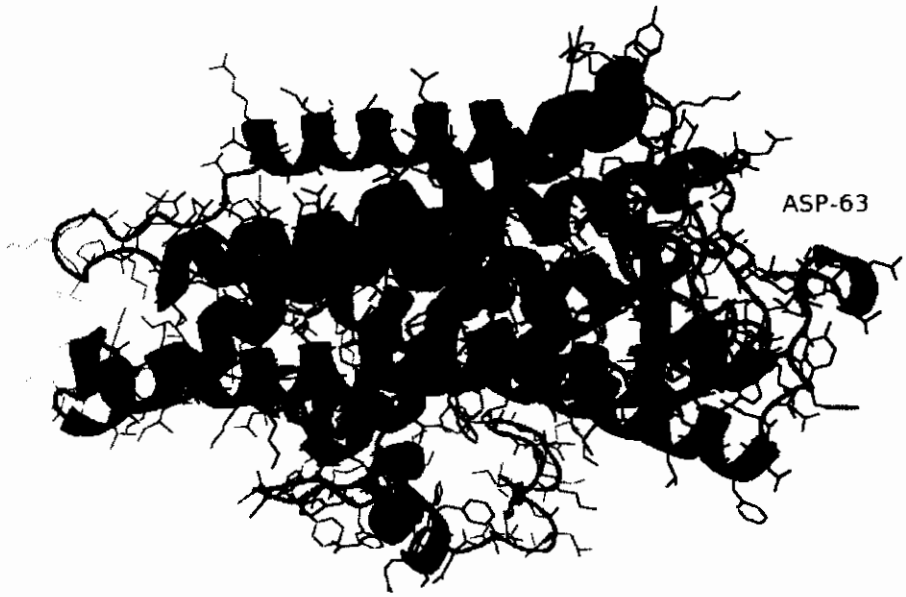


Figure 3.25 (a): Normal Structure with Aspartic acid at position 63

Figure 3.25 (b): Mutated Structure with Valine at position No. 63

3.5.1.2. 1(b). p.E189K

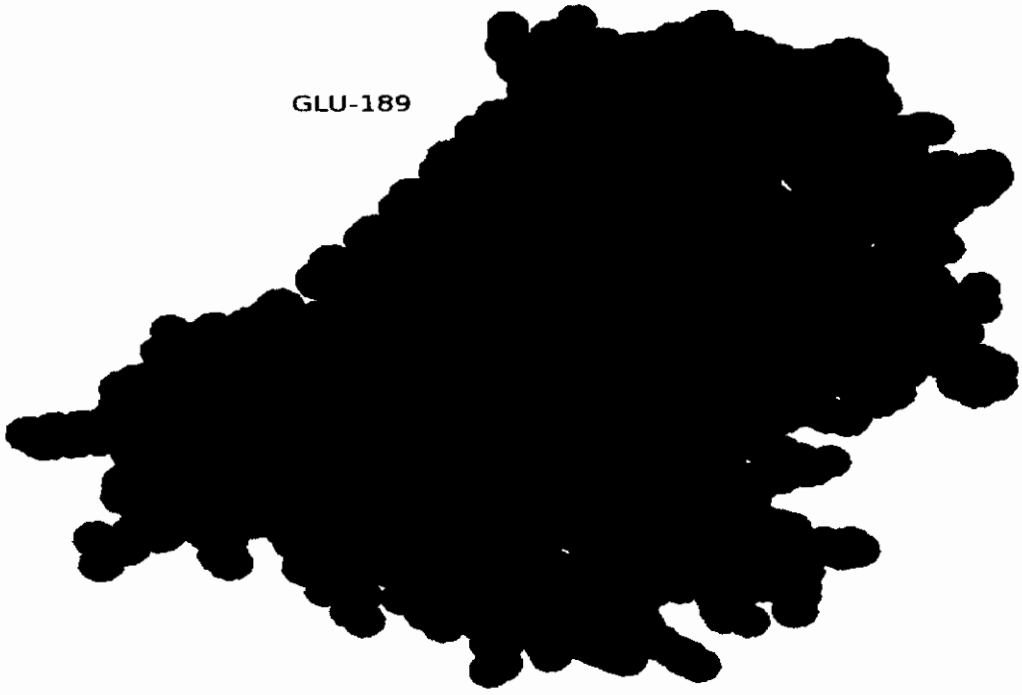


Figure 3.26 (a): Normal Structure with Glutamic acid at position 189

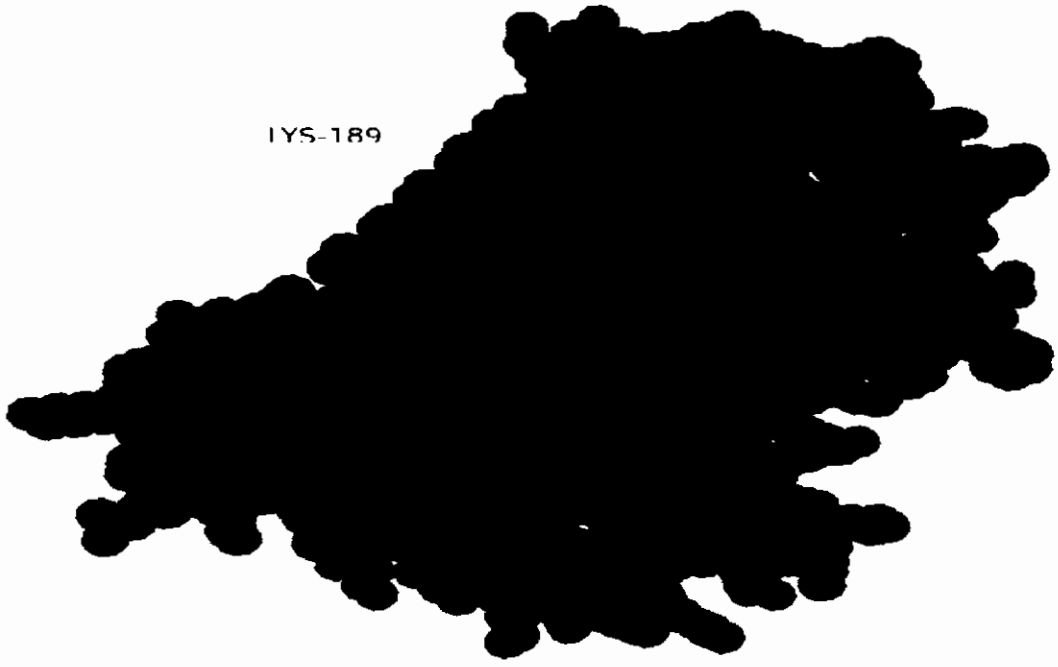


Figure 3.26 (b): Mutated Structure with lysine at position 189

3.5.1.2. 1(c). p.S3T

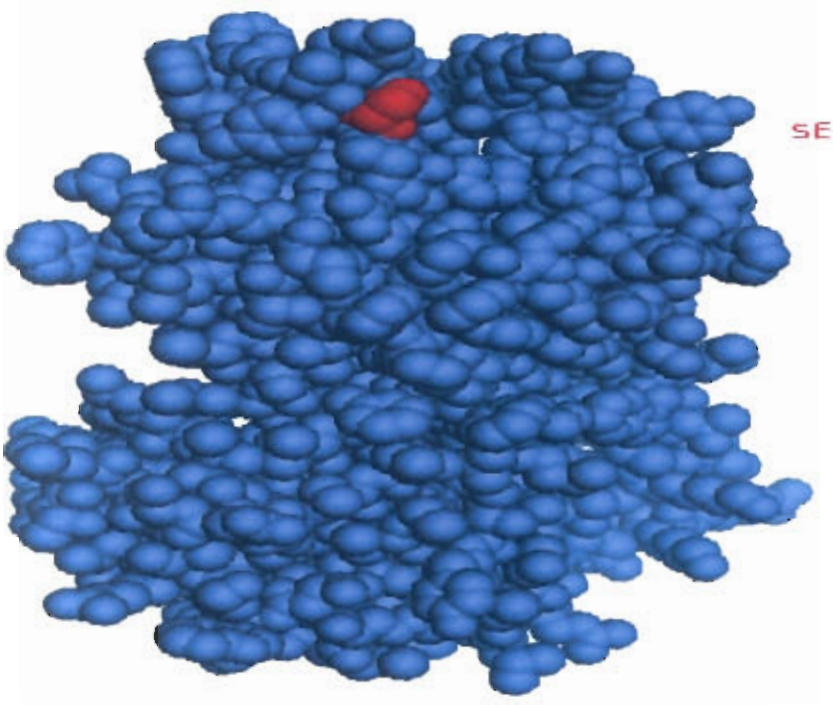


Figure 3.27 (a): Normal Structure with Serine at position 3

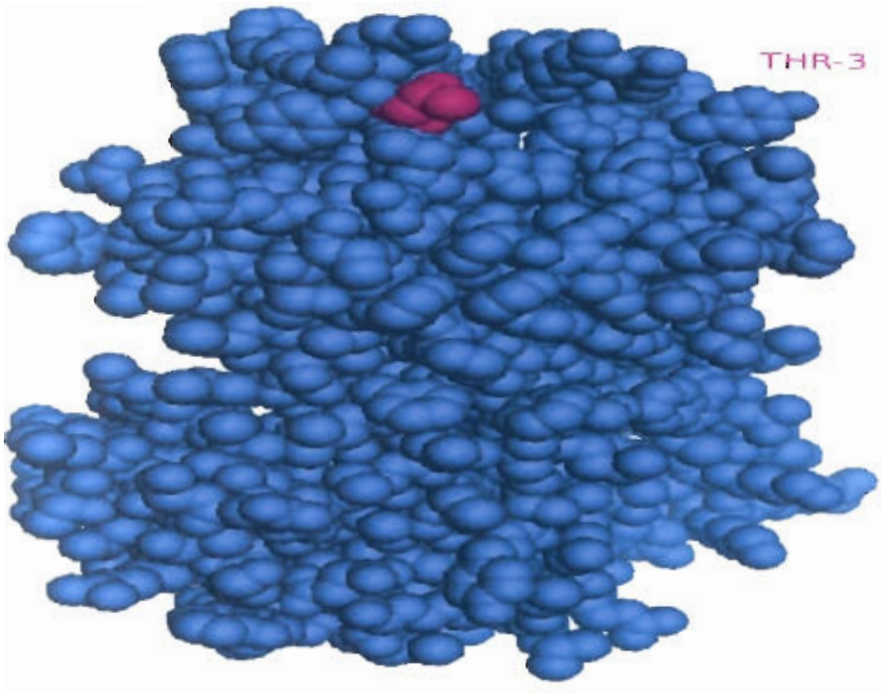


Figure 3.27 (b): Mutated Structure with Therionine at position 3

3.5.1.2. 1(d). p.G146R:

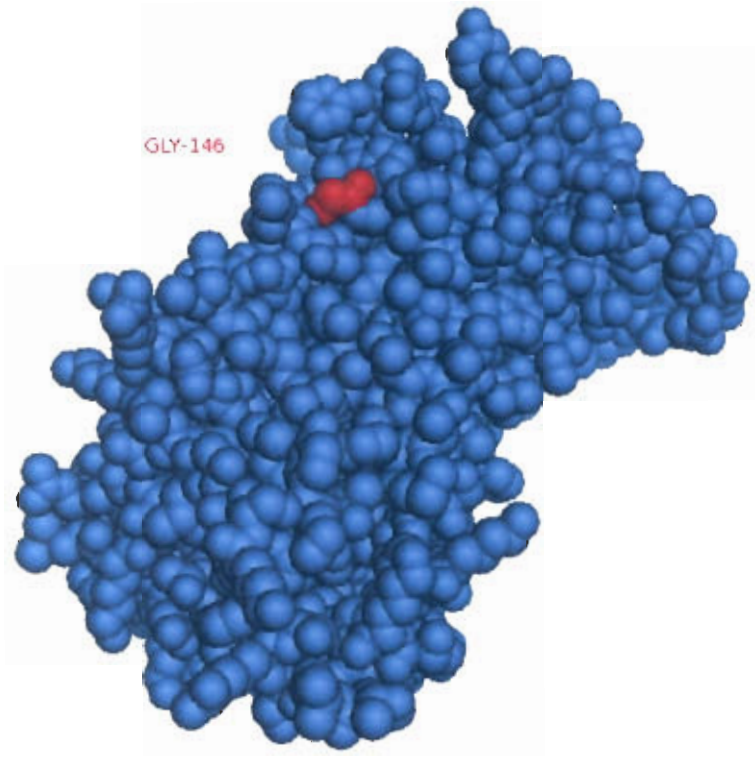


Figure 3.28 (a): Normal Structure with glycine at position No.146

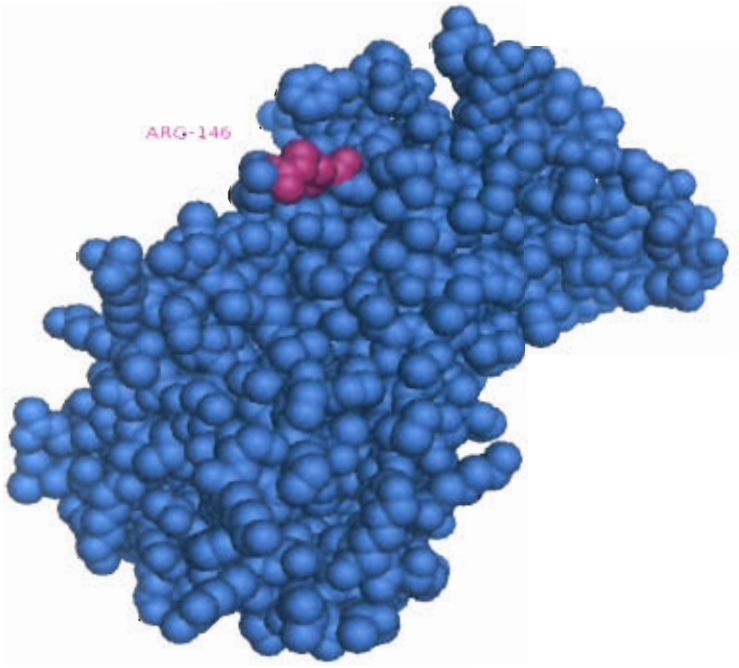


Figure 3.28 (b): Mutated Structure with Aspargine at position No.146

3.5.1.2. 1(e). p.N248Y:

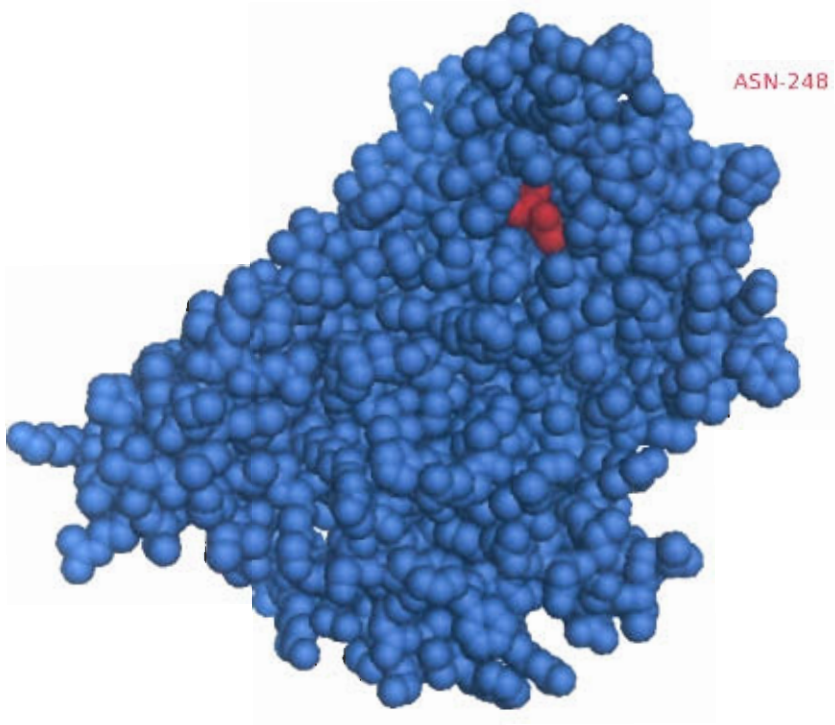


Figure 3.29 (a): Normal Structure with Aspargine at position No.248

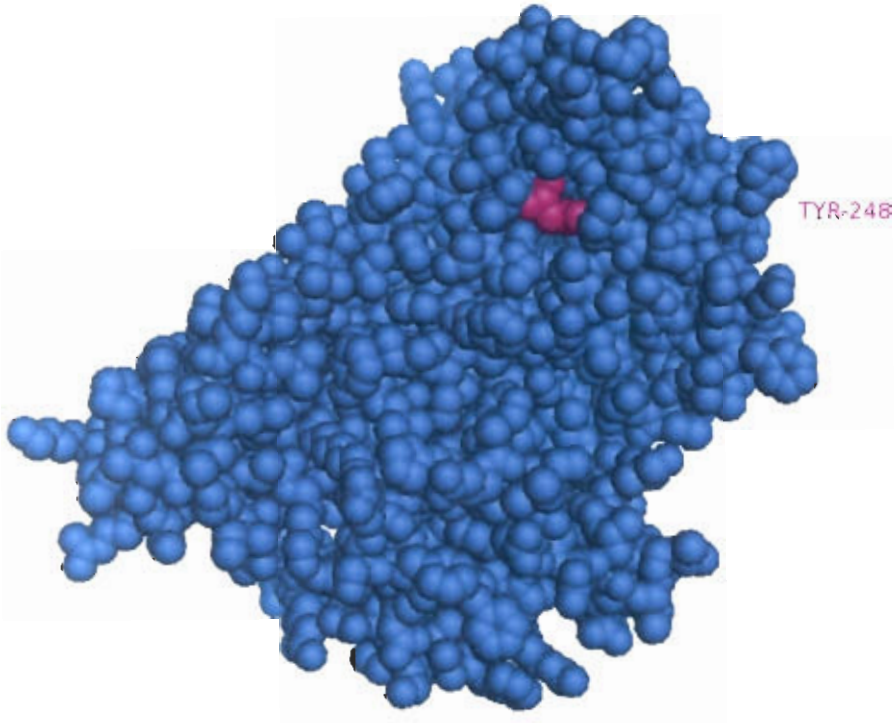


Figure 3.29 (b): Mutated structure with Tyrosine at position No.248

3.5.1.2. 1(f). p.L277P:

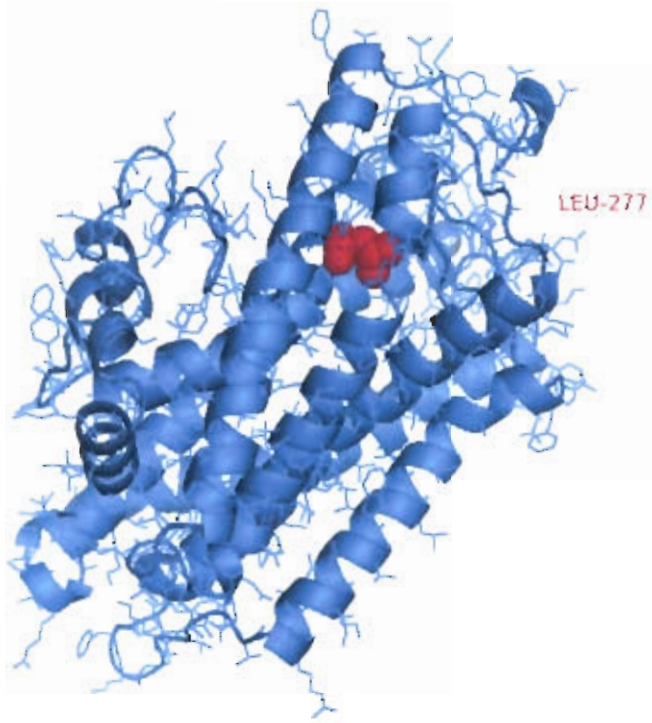


Figure 3.30 (a): Normal Structure with Leucine at position No.277

Figure 3.30 (b): Mutated structure with Proline at position No. 277

3.5.2. 3D structure Prediction of LIPH:

Homology modeling was used to predict 3D structure of *LIPH*. To search templates for 3D structure similarity searching was conducted with the help of BLAST. BLAST results are show in Table 3.33

3.5.2.1 BLAST results for *LIPH:*

Table 3.33: Blast results for *LIPH*

Accession No.	Name of prtein	Protein Origin	E-value
1W52X	Chain X, Crystal Structure Of A Proteolyzed Form Of Pancreatic Lipase Related Protein 2 From Horse	Equus caballus (horse)	6e-56
1BU8A	Chain A, Rat Pancreatic Lipase Related Protein 2	Rattus norvegicus (Norway rat)	3e-55
1GPLA	Chain A, Rp2 Lipase	Cavia porcellus (domestic guinea pig)	3e-53
1LPAB	Chain B, Interfacial Activation Of The Lipase-Procolipase Coupled Receptor.	Homo sapiens (human)	3e-52
2PVSA	Chain A, Structure Of Human Pancreatic Lipase Related.	Homo sapiens (human)	2e-51

3.5.2.2 FSATA results For *LIPH*:

Result of FASTA fro LIPH is shown in table 3.34.

Table 3.34: results of FASTA3 for *LIPH*

Accession No.	Name of prtein	Protein Origin	E-value
1W52X	Chain X, Crystal Structure Of A Proteolyzed Form Of Pancreatic Lipase Related Protein 2 From Horse	Equus caballus (horse)	6e-56
1BU8A	Chain A, Rat Pancreatic Lipase Related Protein 2	Rattus norvegicus (Norway rat)	3e-55
1GPLA	Chain A, Rp2 Lipase	Cavia porcellus (domestic guinea pig)	3e-53
1LPAB	Chain B, Interfacial Activation Of The Lipase-Procolipase Coupled Receptor.	Homo sapiens (human)	3e-52
2PVSA	Chain A, Structure Of Human Pancreatic Lipase Related.	Homo sapiens (human)	2e-51

LIPH has significant similarity with 1BU8 and 1W52, so these two proteins were selected as templates. Through modeler tertiary structure of *LIPH* was predicted. 10 models were generated using 1BU8 as a template and the model with smallest value of objective function

was selected. 20 models were generated using 1W52 as a template. Z-score of prediction using 1Bu8 and 1W52 are shown in table 3.35

Table 3.35: Z-scores of Templates for LIPH

Template	Z-score
1BU8	-1.156
1W52	-2.473

As 1W52 has less Z-score so, 3D structure that was predicted using 1W52 as a template was selected. Tertiary structure of LIPH in a ribbon view is shown in figure 3.31

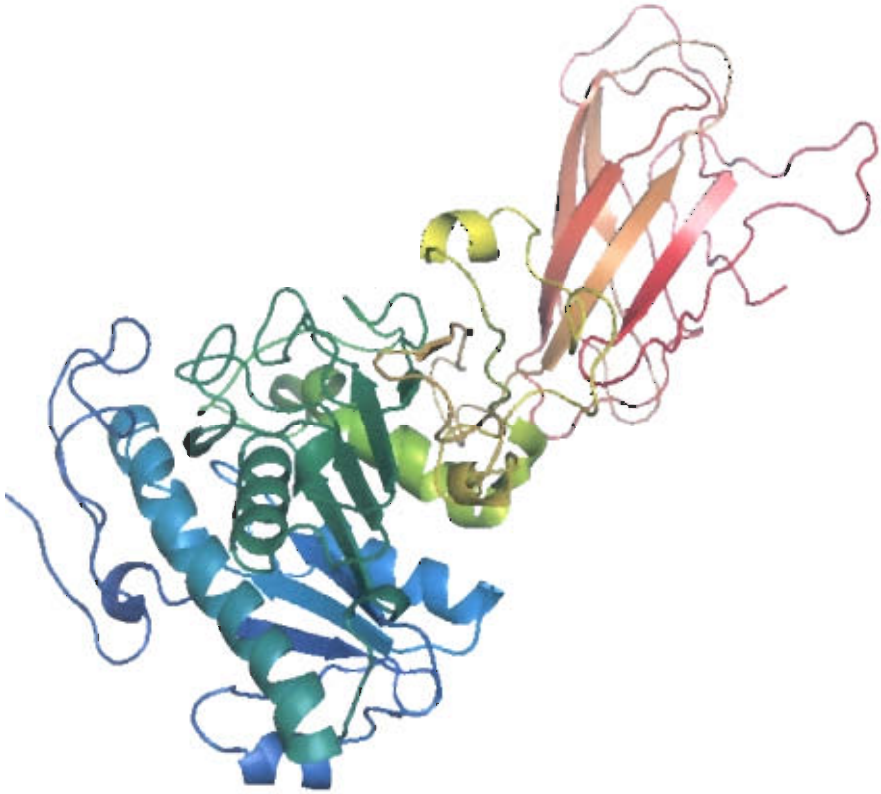


Figure 3. 31 Tertiary Structure of Normal LIPH

3.5.2.2.3 Effect of mutations:

Effects of two reported mutations are shown in Figure 3. 32 (a), (b), (c) for mutation p.W108R and Figure 3. 32 (a), (b), (c) for mutation p.M1T. Mode of mutated residues are displayed in different mode.

3.5.2.2.3 (a) p.W108R:

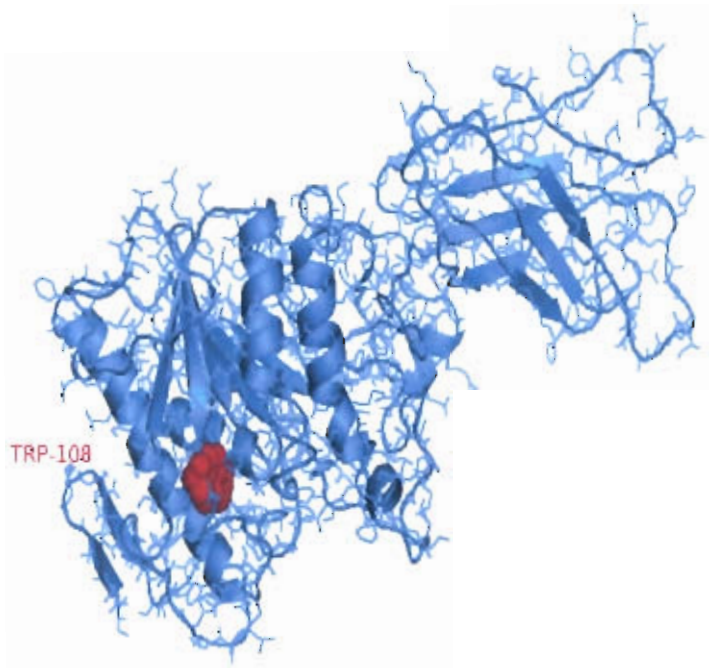


Figure 3.32 (a) Normal Structure with Tryptophan at position No.108

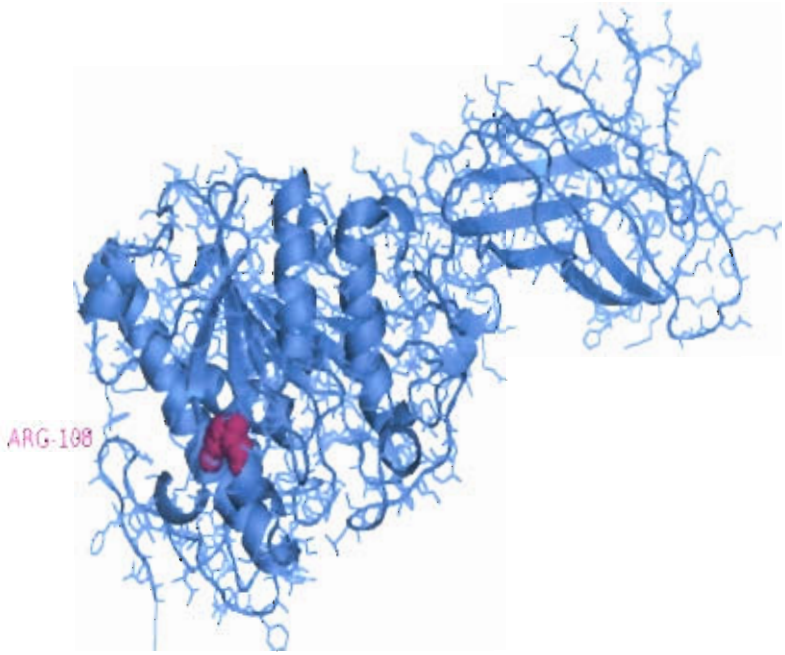


Figure 3.32 (b) Mutated Structure with Arginine at position No. 108

3.5.2.2.3 (b) p.M1T:

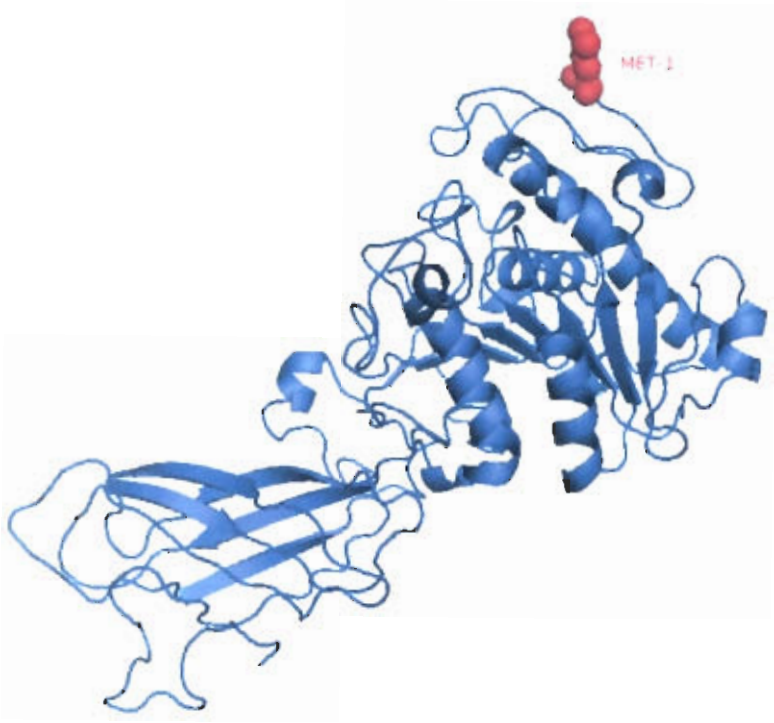


Figure 3.33(a) Normal Structure with methionine at position No. 1

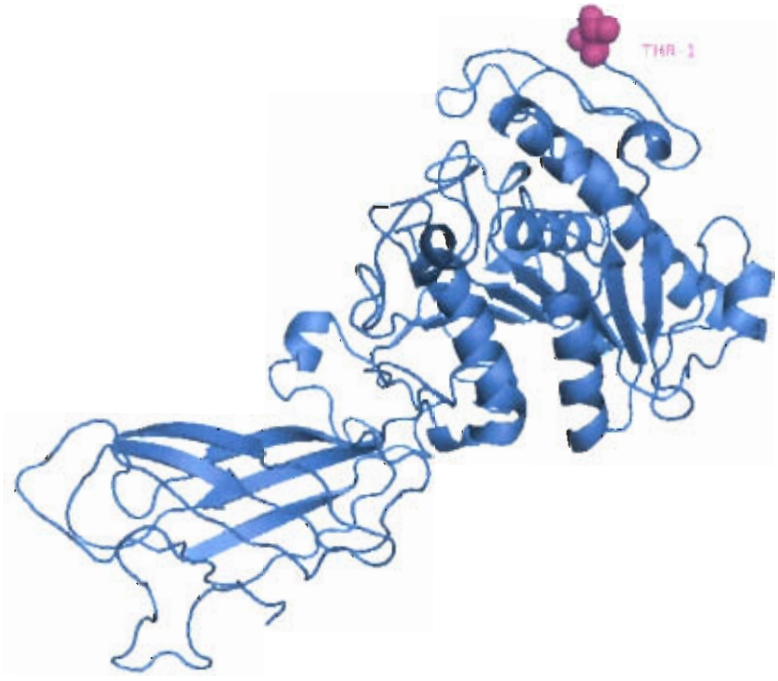


Figure 3.33(b) Mutated structure with therionine at position No. 1

3.6 Structure Evaluation:

To evaluate predicted structure following tools were used:

3.6.1 Procheck

3.6.2 Whatif

3.6.3 PROSA

Results of Procheck for *LIPH* and *P2RY5* are shown in Table 3.36 and 3.37 respectively.

Table 3.36: Results of Procheck for *LIPH*

Core	Allowed	Disallowed	General
80.0%	14.7%	1.5%	2.8%

Table 3.37: Results of Procheck for *P2RY5*

Core	Allowed	Disallowed	General
90.5%	8.5%	.6%	.3%

3.6.3 PROSA

3.6.3 .1 *P2RY5:*

Result of PROSA for *P2RY5* is shown in Figure 3.34. Z-score for *P2RY5* 3D model is -1.17

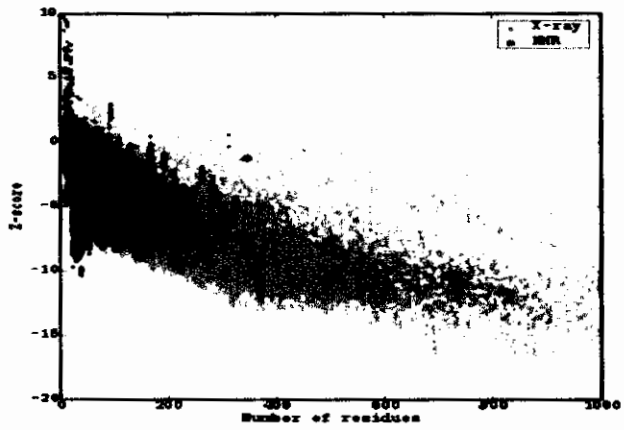


Figure 3.34 PROSA result for *P2RY5*

3.6.3 .2 *LIPH*:

Result of PROSA for *LPIH* is shown in Figure 3.35. Z-score for *P2RY5* 3D model is -1.17 .Z-score for *LIPH* 3D model is -5.38

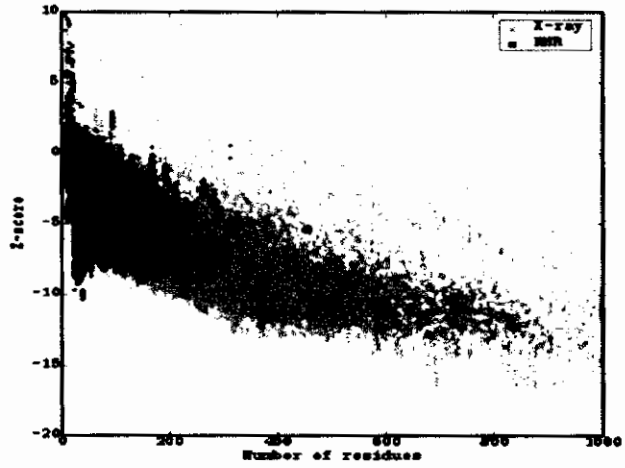


Figure 3.35 PROSA result for *LIPH*

3.7 Pattrens in 3D structure:

Patterns that were predicted by prosite are visualized in 3D structures through CN3D. Patterns are high lightened in yellow color.

3.7.1 *P2RY5*:

Patterns of *P2RY5* in 3D are depicted from Figure 3.36- Figure 3.39

1. **[RK]-x(2,3)-[DE]-x(2,3)-Y**

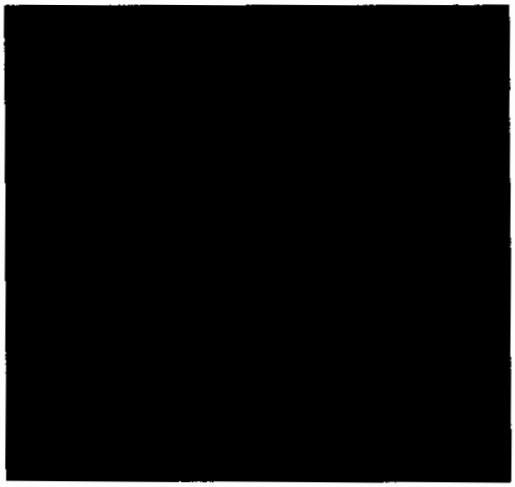


Figure 3.36 *P2RY5* Pattern

2. **[ST]-x-[RK].**

Figure 3.37 *P2RY5* Pattern

3. **G-{EDRKHPFYW}-x(2)-[STAGCN]-{P}**

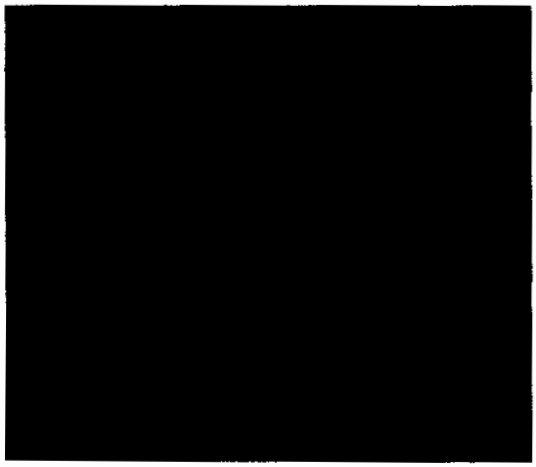


Figure 3.38 *P2RY5* Pattern

4. **N-{P}-[ST]-{P}.**

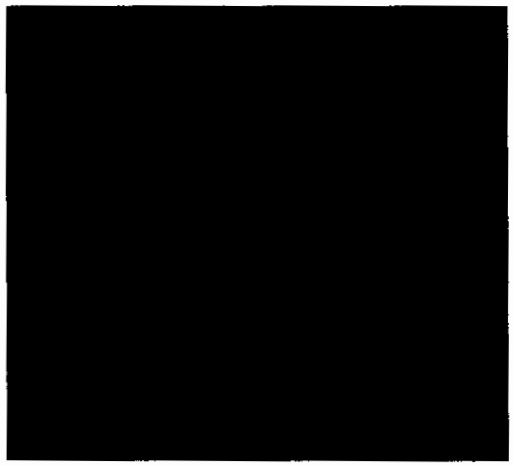


Figure 3.39 *P2RY5* Pattern

3.7.2 Patterns of LIPH in 3D structure:

CN3D was used to visualize patterns retrieved by Prosite in 3D structure. (Figure 3.40-Figure 3.43).

1. **[RK](2)-x-[ST].**

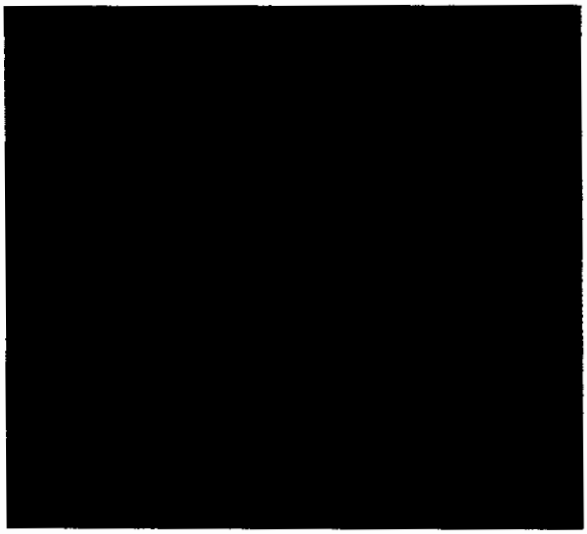


Figure 3.40 *LIPH* Pattern

2. **[ST]-x-[RK].**

Figure 3.41 *LIPH* Pattern

3. **G-{EDRKHPFYW}-x(2)-[STAGCN]-{P}.**

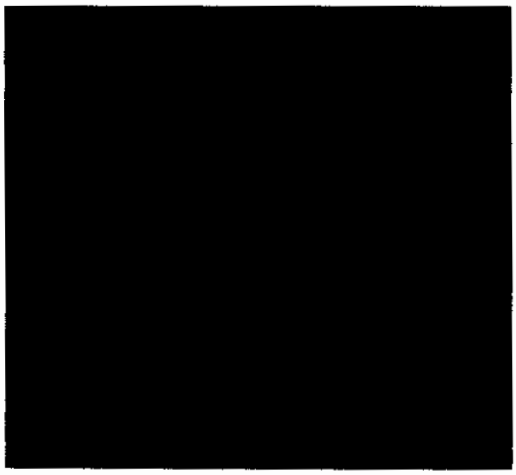


Figure 3.42 *LIPH* Pattern

4. **{P}-[ST]-{P}.**

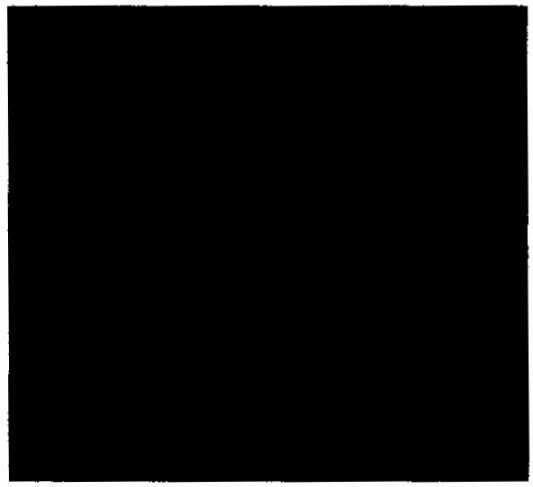


Figure 3.43 *LIPH* Pattern

3.8 Interaction Pathways:

3.8 Interaction Pathways:

3.8.1 *P2RY5*:

SMART was used to predict interaction pathway of *P2RY5*.

(Figure 3.44 and Table 3.38).

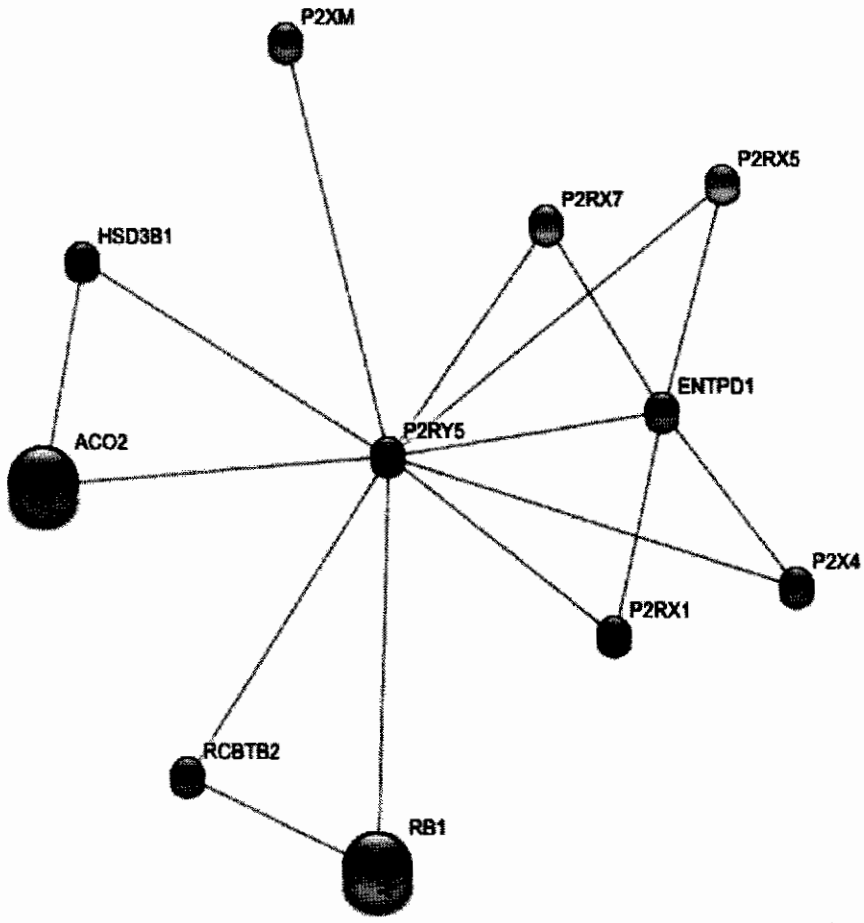


Figure 3.44 Interaction pathway of *P2RY5*

Table 3.38: *P2RY5* interacting proteins

Protein	Name
RB1	Retinoblastoma-associated protein
P2X4	P2X purinoceptor 4 (ATP receptor)
P2RX5	Tax1-binding protein 3 (Tax interaction protein 1)
P2RX1	P2X purinoceptor 1 (ATP receptor)
P2RX7	P2X purinoceptor 7 (ATP receptor)
RCBTB2	RCC1 and BTB domain-containing protein 2
P2XM	P2X purinoceptor 6 (ATP receptor)
HSD3B1	3 beta-hydroxysteroid dehydrogenase/Delta 5-->4-isomerase type I
ACO2	Aconitate hydratase, mitochondrial precursor
ENTPD1	Ectonucleoside triphosphate diphosphohydrolase 1

3.8.2 Interaction Pathway of *LIPH*:

Interaction pathway of *LIPH* was predicted by SMART and is shown in table 3. 37 and Figure 3.45

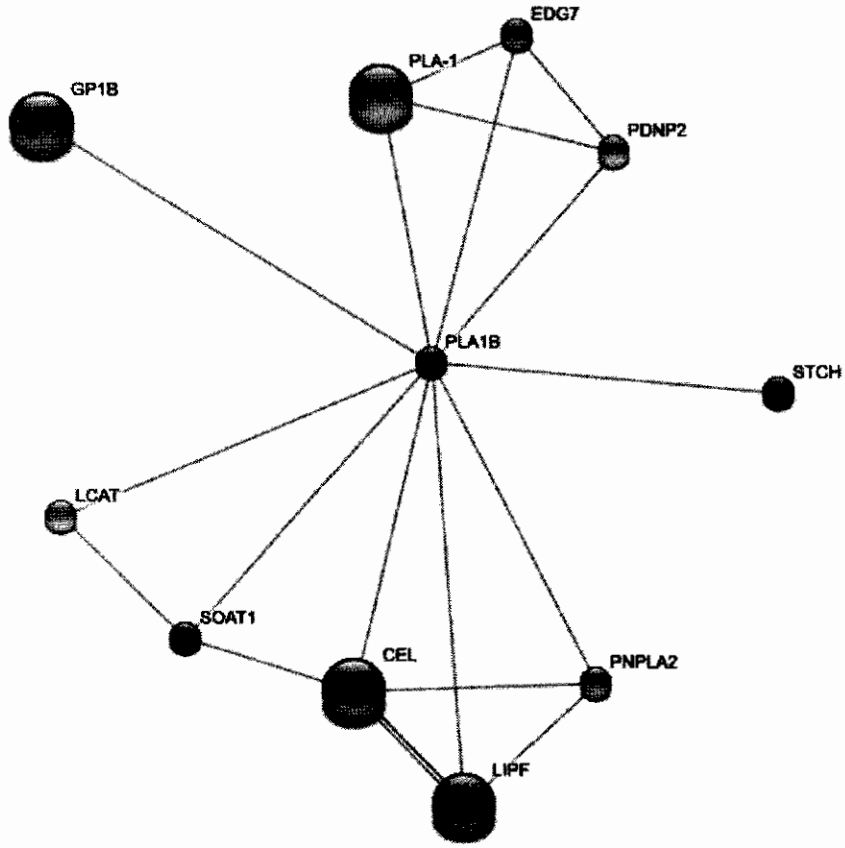


Figure 3.45 : Interaction pathway of LIPH

Table 3.39: Interacting proteins of LIPH

Protein	Full name
PLA-1	POU domain, class 2, transcription factor 3
EDG7	Lysophosphatidic acid receptor Edg-7
PDNP2	Ectonucleotide pyrophosphatase/phosphodiesterase family member 2 precursor
STCH	Stress 70 protein chaperone microsome-associated 60 kDa protein precursor
LCAT	Solute carrier family 12 member 4
GP1B	Platelet glycoprotein Ib alpha chain precursor
PNPLA2	patatin-like phospholipase domain containing 2
SOAT1	Sterol O-acyltransferase 1
LIPF	Gastric triacylglycerol lipase precursor
CEL	Bile salt-activated lipase precursor

3.9 Active Sites Prediction:

To predict active sites of *LIPH* PAR3D was used .*LIPH* is predicted to contain three residue metal-binding site.

Table 3. 38: Active sites in 3D structure of *LIPH*

Residue name	Residue number
ASP	178
ASP	207
HIS	248

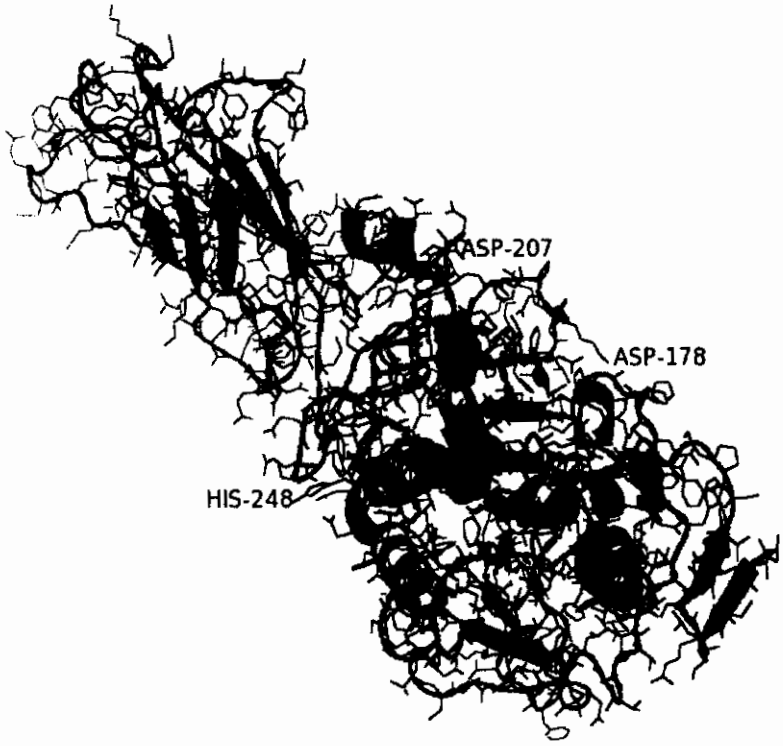


Figure 3.46: Active Site in LIPH

3.10 Ligands for LIPH

Ligands for *LIPH* were retrieved by screening KEGG Ligand database. 2D and 3D structures of ligands are shown in figure 3.49(a) to 3.52.

Aliphatic amide:

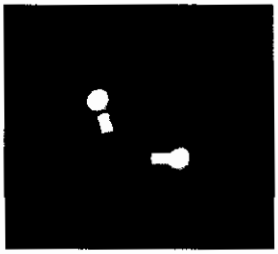
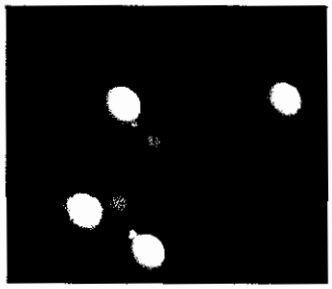


Figure 3.47 (a): Aliphatic amide (3D view)

Figure 3.47(b): Aliphatic amide (2D view)

Aliphatic aldoxime:



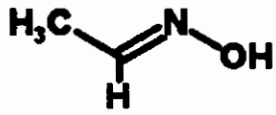


Figure 3.48(a): Aliphatic aldoxime (3D view)

Figure 3.48(b): Aliphatic aldoxime (2D view)

Aliphatic hydroxy acid

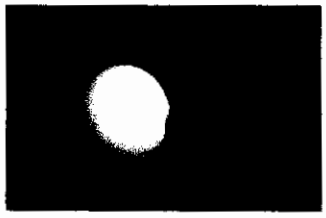


Figure 3.49(a): Aliphatic hydroxy acid (3D view)

Figure 3.49(b): Aliphatic hydroxy acid (2D view)

3.10.4 Aliphatic alcohol:

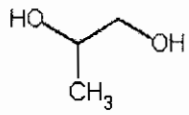


Figure 3.50 Aliphatic Alcohol (2D view)

4. DISCUSSION

Autosomal recessive hypotrichosis is a genetic hair disorder, known to be caused by *P2RY5*, *LIPH* and *DSG4*. Functional annotation of a protein is one of the most eminent problems in Molecular biology. In current study changings brought about by missense mutations in these proteins at molecular level were investigated to explore how changings at molecular level results in change of overall function. All of the work was done through bioinformatics tools. Effects of mutations were analyzed at domains, motifs, post translational modifications, physiochemical properties, secondary structure and Tertiary structure level.

Result of domain prediction shows that *P2RY5* belongs to GPCR rhodopsin like family because it shares 7 transmembrane helices with this family. GPCR is the largest known protein family and comprised of 3% of total human genes. GPCR family is further divided into 6 families. Family A encompasses rhodopsins, dopaminergic and adrenergic receptors and receptors for small organic ligands. Receptors for peptide hormones are included in Family B. Extra cellular calcium sensor and metabotropic glutamate receptors are incorporated in Family C. *P2RY5* share signatures with GPCR family A. It is the largest sub family of GPCR and is also known as rhodopsin family. Two features that are common in all of the receptors of family A are numerous highly conserved amino acids and a disulphide bridge that links the first and second extracellular loops (ECLs). Most of the receptors of this family also possess a PALMITOYLATED cysteine present in the carboxy-terminal tail.

Domain analysis showed that *P2RY5* is a membrane protein as it contains transmembrane helices and has a role in transduction of signals. Muations added 3 more domains into it, that divulged that 3 more folds are added to *P2RY5* that perform variable functions and results in divergence from normal function. *LIPH* share domains with lipase family. Members of this family are involved in lipid and energy metabolism. Lipase domain interacts with PLAT (Polycysterin 1, Lipoxygenase, alpha toxin) and COLIPASE domains. COLIPASE has enzyme regulator activity. PLAT domain is found in various lipid associated proteins. It mediates membrane attachment with the help of other protein

binding partners. So it can be said that COLIPASE regulates activity of LIPASE. LPA that is synthesized from *LIPH* will binds to *P2RY5* and PLAT provides assistance in this process. There is deletion of a motif in *LIPH* due to mutation also. Mutation also brought out changes in pattern of some post-translation modifications like phosphorylation.

Protein kinase C (PKC) is a family of kinases that phophorylates serine and therionine. It is involved in receptor desensitization, transcription regulation, in modulation of membrane structure events, mediation of immune responses, cell growth regulation and memory depending upon type of cell Serine at position number 159, 308, 312 and 317 are phophorylated. Mutations result in phophorylation of serine at position 148 also that will elicit a function change. Therionine at position number 129, 176, 259 and 332 are predicted to be phophorylated in *P2RY5* but due to mutations therionine at position number 276 also become phosphorylated that cause alteration in phsphorylation pattern that can be one of the causes of functional anomaly. Attachment of a myristoyl anchor to an N-terminal glycine by a myristoyltransferase leads to N-myristoylation that serves to modulate modified protein interaction with intracellular membranes or with other proteins. According to PRSOSITE this motif pattern exists in *P2RY5* but when NMT server was used to predict N-myristoylation sites in *P2RY5,* no result was retrieved.

PROSITE retrieved seven patterns for *LIPH.* cAMP and cGMP dependent protein kinases are activated by binding of cyclic AMP and cyclic GMP and then it phophorylate (addition of phosphate) specific residues in particular proteins to bring a functional change. Protein kinase C is conscientious for the phosphorylation of serine and therionine residues that are in closer proximity of basic C-terminal residues. Casein kinase II is a kinase independent of cyclic nucleotides and calcium that phosphorylates serine and/or therionine. According to result of 2ZIP there is no leucine zipper pattern in *LIPH.* Cell attachment sequence i.e. Arginine-Glycine-Aspartic acid is important in cell adhesion. Change in physiochemical properties of *LIPH* and *P2RY5* was also observed. Result of HMMTOP predicts that *P2RY5* has N-terminus at outside and has 7 transmembrane helices these 7 transmembrane helices are also predicted by interproscan and according to work of Attwood TK, Findlay JB (1994) *P2RY5* is a member of GPCR family that is

characterized by 7 transmembrane helices. N terminus of *LIPH* is also at outer side. *LIPH* does not contain any transnmembrane helices.

P2RY5 interacts with proteins that are shown in Table 3.47. RB1 is a tumor suppressor gene that controls the hair cells growth. It is responsible for transcription repression of E2F1 target genes. E2F1 interacts with underphosphorylated active form of RB1 and as a consequence repression of its transcription activity takes place that leads to cell cycle arrest. It is implicated in heterochromatin. It pedals histone "H4 Lys-20 trimethylation". It is also involved in inhibition of the inherent kinase activity of TAF1. P2X4, P2XM and P2RX5 are ATP receptors. Another receptor for ATP is P2RX1 with relatively high calcium permeability and it intercedes synaptic transmission between neurons and from neurons to smooth muscle and is linked to apoptosis it does so by increasing the intracellular concentration of calcium in the presence of ATP. P2RX7 is an ATP receptor that operates as a ligand gated ion channel and plays part in ATP-dependent lysis of macrophages in the course of the formation of membrane pores that are permeable to large molecules. RCBTB2 shows Ran guanyl-nucleotide exchange factor activity and is involved in protein binding. HSD3B1 catalyzes the oxidative conversion of ketosteroids. The 3-beta-HSD enzymatic system plays a crucial role in the biosynthesis of all classes of hormonal steroids. ACO2 is responsible for 4 iron, 4 sulfur cluster binding, aconitate hydratase activity and iron ion binding. ENTPD1 regulates purinergic neurotransmission within nervous system by hydrolyzing ATP and other nucleotides.

LIPH interacts with proteins that are shown in table 3.39. PLA-1 shows sequence-specific DNA binding and transcription factor activity. EDG7 acts as G-protein coupled Receptor and as a Transducer. PDNP2 is involved in hydrolysis of lysophospholipids to form lysophosphatidic acid (LPA) in extracellular fluids. lysophosphatidylcholine is its key substrate. STCH has peptide-independent ATPase activity. LCAT Mediates electroneutral potassium-chloride cotransport when activated by cell swelling. GP1B Catalyzes the initial step in triglyceride hydrolysis in adipocyte and non-adipocyte lipid droplets. It also has acylglycerol transacylase activity, may act coordinately with LIPE/HLS within the lipolytic cascade, regulates adiposome size and involved in the adiposomes degradation, have role in energy homeostasis, have a role in the response of

an organism against starvation and have role in enhancing hydrolysis of triglycerides and providing free fatty acids to other tissues to be oxidized in situations of energy depletion. SOAT1 is an enzyme that catalyzes that reaction for the formation of fatty acid-cholesterol esters and is involved in assembly of lipoprotein and dietary cholesterol absorption. It also exhibits acyltransferase activity and act as a ligase. LIPF has lipid binding and triacylglycerol lipase activity. CEL is an enzyme that is responsible for catalyzing fat and vitamin assimilation and acts in combination with colipase and pancreatic lipase for the absolute digestion of dietary triglycerides.

Secondary structure that was predicted deciphered number of alpha helices and beta sheets present in *LIPH* and *P2RY5*. As a result of mutations number of residues that form particular secondary structure, changes which result in structural change that leads to functional change as alpha helices play role in DNA binding and due to mutation they get reduced in *P2RY5*. Reduction in number of helices means reduction in sites for DNA binding that will definitely be derogated as it happens in the case of *P2RY5*.

Knowledge of tertiary structure is monumental for predicting molecular basis of function of the protein. Experimental methods for determination of tertiary structures that are XCR and NMR are labor demanding, costly, time intense and there are some sample constraints also. Due to these reasons predictions method have gained a lot of vogue and become a central point in molecular biology research today.

Tertiary structure of *P2RY5, LIPH* was also predicted in the current study. Reported mutations in these genes were inserted into the best selected models. Mutated models were predicted and compared with the normal models. *LIPH* has 32 % identity with *1W52* and 36% identity with *1BU8*. After using these proteins as a template tertiary structure of protein was obtained and evaluated using what if, procheck and prosa. Model predicted by using *1W52* as a template had a good Z-score and was selected as a final model. Mutations were inserted in a predicted model by what-if server. *P2RY5* has no significant identity with stored PDB structures so comparative modeling could not be used in this case so tertiary structure of *P2RY5* was predicted through threading approach. Study of changings that mutation brought out is important to understand to make a therapeutic agent for the proper cure of disease. There was a lot of work done on

mutation identification in genes involved in autosomal recessive hypotrichosis in different populations but the affects that mutation brought in a protein at molecular level were not studied. Due to mutation analysis that is done in current study, insight into alterations brought about by mutations, is increased and it is now uncomplicated to identify target to inhibit function of abnormal protein. *LIPH* has sites for metal binding. Although some ligands have been screened but further ligands can be designed that can best inhibit function of mutated *LIPH*. Once function of *LIPH* is inhibited LPA can be given as a supplement that can then bind with *P2RY5* and hair growth can therefore be initiated in this way LAH2 can be treated.

Conclusion:

Change in a single amino acid drastically effects normal functioning of proteins. Missense mutations in *P2RY5* and *LIPH* have effect on their structures that leads to distorted function. Mutations have following affect on *LIPH* and *P2RY5:*

1. Isoelectric point of protein changes which signifies that there must be a change in interaction pattern of a proteins. Lack of proper interaction between Proteins will lead to imperfect function.
2. Domains are added that perform independent function and interrupt normal functioning of protein.
3. There is alteration in phosphorylation pattern of proteins. Alteration in post-translational modifications leads to variable functionality.
4. There is also a change in secondary structure.
5. Tertiary structure that is directly related to functionality of protein also changes.

LIPH can be targeted to surmount LAH2. So it can be concluded that changings at sequence level are propagated to tertiary level and as a consequence other properties of proteins also get altered and bioinformatics is an efficient, time effective and cost effective way of modeling a disease. It can serve as an efficient, inexpensive surrogate of expensive experimental methods. In a developing country like Pakistan Bioinformatics can have a greater impact on accelerating biological research, where resources are limited

5. REFERENCES

Al Aboud K., Al Hawsawi K., Al Aboud D., Al Githami A. (2002) Hereditary hypotrichosis simplex: report of a family. *Clin. Exp. Derm.* **27**: 654-656.

Ali G., Chishti M., Raza S., John P., Ahmad W.(2007) A mutation in the lipase H (*LIPH*) gene underlie autosomal recessive hypotrichosis. *Hum Genet*, **121**:319–325.

Altschul F., Madden L., Schaffer A., Zhang J., Zhang Z., Miller W., Lipman J.(1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.***25**: 3389-3402.

Aslam M., Chahrour M., Razzaq A., Haque S., Yan K., Leal S., Ahmad W. (2004) A novel locus for autosomal recessive form of hypotrichosis maps to chromosome 3q26.33–q27.3. *J Med Genet.* **41**: 849–852.

Attwood T, Findlay J (1994) Fingerprinting G-protein-coupled receptors *Protein Eng.* **7**:195-203.

Azeem, Z., Jelani M., Naz G., Tariq M., Wasif N., Naqvi S., Ayub M., Yasinzai M., Amin-ud-din M., Wali A., Ali G., Chishti M., Ahmad W. (2008) Novel mutations in G protein-coupled receptor gene (P2RY5) in families with autosomal recessive hypotrichosis (LAH3). *Hum. Genet.***123**:515-519.

Bauer E., Rivals E., Vingron M. (1998) Computational Approaches to Identify Leucine Zippers. *Nucleic Acids Res.* **26**:2740-2746.

Blom N., Gammeltoft S., Brunak S. (1999) Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. *Journal of Molecular Biology.* **294**:1351-1362.

Bork P. (1991) Shuffled domains in extracellular proteins. *FEBS Lett.* **286**:47-54.

Castro E., Sigrist J., Gattiker A., Bulliard V., Langendijk-Genevaux P., Gasteiger E., Bairoch A., Hulo N. (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins.*Nucleic Acids Res.* **34**: 362–365.

Castori M., Covaciu C., Rinaldi R., Grammatico P., Paradisi M.(2008) rare cause of syndromic hypotrichosis: Nicolaides-Baraitser syndrome, *J Am Acad Dermatol.* **59**:92-98.

Chothia C. (1992) Proteins. One thousand families for the molecular biologist. *Nature* **357**:543-544.

Duan Y., Kollman P. (2001) Computational protein folding: From lattice to all atoms. *IBM Systems Journal.***40:** 297 – 309.

Eisenhaber F., Eisenhaber B., Kubina W., Stroh S., Neuberger G., Schneider G., Wildpaner M. (2003) Prediction of lipid posttranslational modifications and localization signals from protein sequences: big-Π, NMT and PTS1. *Nucleic Acids Res.* **31**:3631–3634.

Eswar N., Ramakrishnan C., Srinivasan N. (2003) Stranded in isolation: structural role of isolated extended strands in proteins, *Protein Engineering.***16**: 331-339.

Eswar N., Webb B., Marti M., Madhusudhan M., Eramian D., Shen M., Pieper U., Sali A. (2006) Comparative protein structure modeling using MODELLER. *Curr Protoc Bioinformatics.* Chapter 5:Unit 5.6.

Grossmann A., Kolibaba K., Willis S., Corbin A., Langdon W., Deininger M., Druker B. (2004) Catalytic domains of tyrosine kinases determine the phosphorylation sites within c-Cbl. *FEBS Lett.* **577**: 555-562.

Guex N., Peitsch M. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis.* **15**: 2714-2723.

Gupta R., Jung E., Brunak S. (2004) Prediction of N-glycosylation sites in human proteins.

Heijne G. (2006) Membrane-protein topology *Nature Reviews Molecular Cell Biology* **7**:909-918.

Herzog H., Darby K., Hort Y., Shine J. (1996) Intron 17 of the Human Retinoblastoma Susceptibility Gene Encodes an Actively Transcribed G Protein-coupled Receptor Gene. *Genome Res.* **6**:858-861.

Hunter S., Apweiler R., Attwood T., Bairoch A., Bateman A., Binns D., Bork P., Das U., Daugherty L., Duquenne L., Finn R., Gough J., Haft D., Hulo N., Kahn D.,

Kelly E., Laugraud A., Letunic I., Lonsdale D., Lopez R., Madera M., Maslen J., McAnulla C., McDowall J., Mistry J., Mitchell A., Mulder N., Natale D., Orengo C., Quinn A., Selengut J., Sigrist C., Thimma M., Thomas P., Valentin F., Wilson D., Wu C., Yeats C. (2009) InterPro: the integrative protein signature database. *Nucleic Acids Res.* **37:**211-215.

Jelani M., Wasif, N., Ali G., Chishti M., Ahmad W. (2008) A novel deletion mutation in LIPH gene causes autosomal recessive hypotrichosis (LAH2). *Clin. Genet.* **74:**184-188.

Jin W., Broedl U., Monajemi H., Glick J., Rader D. (2002) Lipase H, a New Member of the Triglyceride Lipase Family Synthesized by the Intestine. *Genomics.* **80**:268-273.

John C., Obenauer, Lewis C. Cantley and Michael B. Yaffe. (2003) Scansite 2.0: proteome-wide prediction of cell signaling interactions using short sequence motifs.*Nucleic Acids Res.* **31**:3635–3641.

Julenius K. (2007) NetCGlyc 1.0: Prediction of mammalian C-mannosylation sites. *Glycobiology.* **17**:868-876

Kahraman A., Morris R., Laskowski R., Thornton J. (2007) Variation of geometrical and physicochemical properties in protein binding pockets and their ligands, *BMC Bioinformatics*, **8:**S8-S1.

Kanehisa M., Goto S., Kawashima S., Nakaya A. (2002) The KEGG databases at GenomeNet. *Nucleic Acids Research.* **30**:42-46

Kano K. , Arima N. , Ohgami M. , Aoki J. (2008) LPA and its analogs-attractive tools for elucidation of LPA biology and drug development, *Curr Med Chem.* **15**: 2122-2131.

Kazantseva A., Goltsov A., Zinchenko R., Grigorenko A., Abrukova P., Moliaka A., Kirillov Y., Guo A., Lyle Z., Ginter S., Rogaev E., Evgeny I. (2006) Human hair growth deficiency is linked to a genetic defect in the phospholipase gene LIPH. *Science* **314**:982-985.

Kljuic A. , Bazzi H. , Sundberg J. , Martinez A. , Shaughnessy R. , Mahoney M. , Levy M. , Montagutelli X. , Ahmad W. , Aita V., Gordon D. , Uitto J. , Whiting D. , Ott J. , Fischer S. , Gilliam T., Jahoda C., Morris R., Panteleyev A., Nguyen V., Christiano A. (2003) Desmoglein 4 in Hair Follicle Differentiation and Epidermal Adhesion: Evidence from Inherited Hypotrichosis and Acquired Pemphigus Vulgaris, *Cell* **113**:249–260.

Goyal K., Mohanty D., Mande S. (2007) PAR-3D: a server to predict protein active site residues. *Nucleic Acids Research.* **35**:503-505.

Landschulz W., Johnson P., McKnight S. (1988)The leucine zipper: a hypothetical structure common to a new class of DNA binding *Science*. **240**: 1759-1764.

Letunic I., Doerks T., Bork P. (2009) SMART 6: recent updates and new developments. *Nucleic Acids Res.* **3**:229-232.

Liang H., Hsu J. (2005) Recent developments in structural proteomics for protein structure determination. *Proteomics*, **5**:2056–2068.

McMillan D., Kathleen M. Wandover K., Richardson J., White P. (2002) Very Large G Protein-coupled Receptor-1, the Largest Known Cell Surface Protein, Is Highly Expressed in the Developing Central Nervous System. *Biol. Chem,* **277**: 785-792.

Morris L, MacArthur W, Hutchinson G, Thornton M (1992) Stereo chemical quality of protein structure coordinates. *Proteins*, **12**:345-364.

Naz G., Khan B., Ali G., Azeem Z., Wali A., Ansar M., Ahmad W. (2009) Novel missense mutations in lipase H (LIPH) gene causing autosomal recessive hypotrichosis (LAH2). *J Dermatol Sci.* **54**:12-16.

Pasternack S., Von Kugelgen I., Al Aboud K., Lee Y., Ruschendorf F., Voss K., Hillmer A., Molderings J., Franz T., Ramirez A., Nurnberg P., Nothen M., Betz R. (2008) G protein-coupled receptor P2Y5 and its ligand LPA are involved in maintenance of human hair growth. *Nature Genet.* **40:**329-334.

Pearson R., Lipman J. (1988) Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A.* **85**: 2444-2448.

Puntervoll P., Linding R., Gemund C., Davidson S, Mattingsdal M., Cameron S., Martin D., Ausiello G., Brannetti B., Costantini A, Ferre F., Maselli V., Via A., Cesareni G., Diella F., Furga G., Wyrwicz L., Ramu C., McGuigan C., Gudavalli R., Letunic I., Bork P., Rychlewski L., Kuster B., Citterich M., William N. Hunter, Aasland R., Toby J. (2003) ELM server: a new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Res.* **31**:3625–3630.

Rost B., Schneider R., Sander C. (1997) Protein fold recognition by prediction-based threading *JMB*. **270**:1-10.

Shimomura Y., Wajid M., Ishii Y., Shapiro L., Petukhova L., Gordon D., Christiano M. (2008) Disruption of P2RY5, an orphan G protein-coupled receptor, underlies autosomal recessive woolly hair. *Nature Genet.* **40**: 335-339.

Stein A., Panjkovich A., Aloy P. (2009) 3did Update: domain–domain and peptide-mediated interactions of known 3D structure, *Nucleic Acids Research*, **37**:300-304.

Song H., Ramus S., Shadforth D., Quaye L., Kjaer S., Dicioccio R., Dunning A., Hogdall E., Hogdall C., Whittemore A., McGuire V., Lesueur F., Easton D., Jacobs I., Ponder B., Gayther S., Pharoah P. (2006) Common Variants in RB1 Gene and Risk of Invasive Ovarian Cancer. *Cancer Res.***66**:10220-10226.

Tariq M., Ayub M., Jelani M., Basit S., Naz G., Wasif N., Raza S., Naveed A., Khan S., Azeem Z., Yasinzai M., Wali A., Ali G., Chishti M., Ahmad W.(2009) Mutations in the P2RY5 gene underlie autosomal recessive hypotrichosis in 13 Pakistani families. *Br J Dermatol.***106**: 1006-1010.

Tusnady G., Simon I. (2001) The HMMTOP transmembrane topology prediction server. *Bioinformatics.* **17**:849-850.

Wali A., Chishti S., Ayub M., Yasinzai M., Kafaitullah, Ali G., John P., Ahmad W.(2007)Localization of a novel autosomal recessive hypotrichosis locus (LAH3) to chromosome 13q14.11–q21.32. *Clin Genet.* **72:**23–29.

Vriend G. (1990) WHAT IF: a molecular modeling and drug design program. J *Mol Graph.* **8**: 52-56.

Wiederstein M., Sippl M. (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **35**:407-410.

Wong Y., Lee T., Liang H., Huang C., Wang T., Yang, Huei Chu C., Huang H., Tat Ko M., Hwang J. (2007) KinasePhos 2.0: a web server for identifying protein kinase-specific phosphorylation sites based on sequences and coupling patterns. *Nucleic Acids Res.* **35**:588–594.

Yamamoto A., Noda I., Nagasawa M.(1970) Comparison of Various Methods of Determining Molecular Weight Distribution, *Polymer Journal*, **1**:304-311.