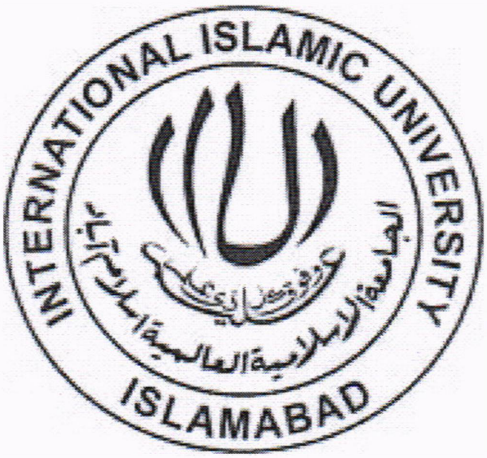
COMPUTATIONAL ANALYSIS OF HEAD AND NECK CANCER



Researcher:

Adnan Tahir

3-FBAS/MSBI/S11

Department of Bioinformatics and Biotechnology
Faculty of Basic and Applied Sciences
International Islamic University, Islamabad.
(2013)

Accession No. 10785

MS
571.978
ADC
Head-Cancer.
Neck-Cancer
Cancer Biology

DATA ENTERED
27/9/13

COMPUTATIONAL ANALYSIS OF HEAD AND NECK CANCER

Thesis submitted in the partial fulfillment of the
requirement for the degree of
MS in Bioinformatics

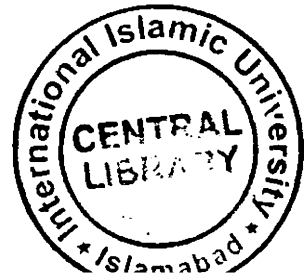
By:

Adnan Tahir

3-FBAS/MSBI/S11

Department of Bioinformatics and Biotechnology
Faculty of Basic and Applied Sciences
International Islamic University, Islamabad.

(2012)



COMPUTATIONAL ANALYSIS OF HEAD AND NECK CANCER

Researcher

Adnan Tahir
MS Bioinformatics
Reg. No. 3-FBAS/MSBI/S11

Supervisors

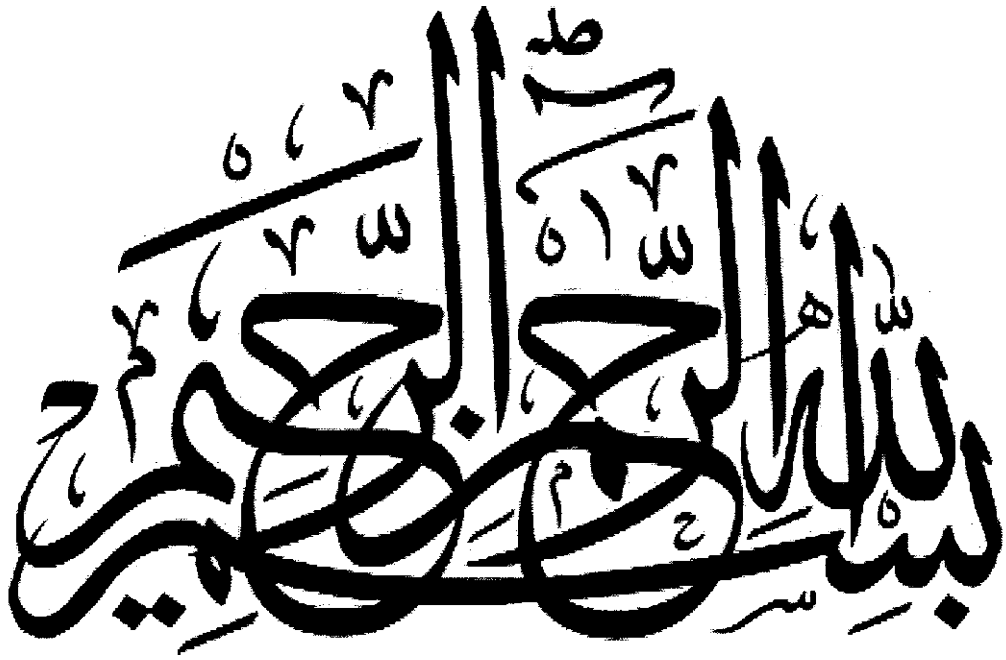
Dr. Jabar Zaman Khan Khattak
Assistant Professor
Department of Bioinformatics and Biotechnolo

Co-Supervisor

Dr. Asif Mir
Assistant Professor
Department of Bioinformatics and Biotechnolo

Department of Bioinformatics and Biotechnology
Faculty of Basic and Applied Sciences
International Islamic University, Islamabad.

(2012)



In the name of Allah Most Gracious and Most Beneficial

Department of Bioinformatics and Biotechnology

International Islamic University Islamabad

Dated: _____

FINAL APPROVAL

It is certificate that we have read the thesis submitted by **Mr. Adnan Tahir** and it is our judgment that this project is of sufficient standard to warrant its acceptance by the International Islamic University, Islamabad for the M.S Degree in Bioinformatics.

Chairman
Dr. Jabar Zaman Khan Khattak
Assistant Professor
Department of Bioinformatics and Biotechnology
International Islamic University Islamabad

Dean, FBAS
Dr. Muhammad Sher
International Islamic University Islamabad

19-2-13

Department of Bioinformatics and Biotechnology

International Islamic University Islamabad

Dated: 12-2-2013

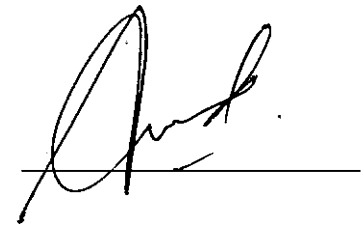
FINAL APPROVAL

It is certificate that we have read the thesis submitted by **Mr. Adnan Tahir** and it is our judgment that this project is of sufficient standard to warrant its acceptance by the International Islamic University, Islamabad for the M.S Degree in Bioinformatics.

COMMITTEE

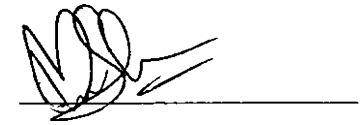
External Examiner

Dr. Mukhtar Ahmad
PMAS Arid Agriculture University, Rawalpindi

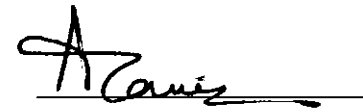


Internal Examiner

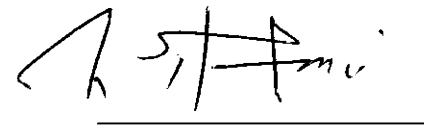
Dr. Imran Shabir
Department of Bioinformatics and Biotechnology
International Islamic University Islamabad



Supervisor
Dr. Jabar Zaman Khan Khattak
Assistant Professor
Department of Bioinformatics and Biotechnology
International Islamic University Islamabad



Co-Supervisor
Dr. Asif Mir
Assistant Professor
Department of Bioinformatics and Biotechnology
International Islamic University Islamabad



A thesis submitted to Department of Bioinformatics and Biotechnology,
International Islamic University, Islamabad as a partial
fulfillment of requirement for the award of the
Master in Sciences of Bioinformatics
(MSBI)

I would like to dedicate this thesis to my beloved father Tahir Mahmood, as a small symbol of my gratitude for their invaluable trust in me, without which this work would not be carried out.

DECLARATION

I hereby solemnly declare that the work ***"Computational Analysis of Head and Neck Cancer"*** present in the following thesis is my own effort, except where otherwise acknowledged and that the thesis is my own composition. No part of the thesis has been previously presented for any other degree.

Dated: 12-02-2013

Adnan.

Adnan Tahir

3-FBAS/MSBI/S11

TABLE OF CONTENTS

Acknowledgements

All praises for **ALLAH ALMIGHTY**, the Most Merciful, the Most Gracious, Who helped and blessed me all the way along till the completion of my project. Without His grace and mercy, I am nothing. Countless salutations be upon the **HOLY PROPHET HAZRAT MUHAMMAD (peace be upon him)**, the city of knowledge who guided his UMMAH to seek knowledge from cradle to grave.

At this Juncture, I thank to my parents **Mr. & Mrs. Tahir Mahmood** whose selfless sacrificial life and their great efforts and unceasing prayers have enabled me to reach the present position in life. My deepest gratitude goes to my family for their unflagging love support throughout my life, this thesis is simply impossible without them. My thanks are tendered to my loving, caring, and sincere parents, sisters, cousins and family members specially **Usman Rana, Asad Rana** and **Muhammad Hassan** for their ever-lasting cooperation and prayers.

I am immensely pleased to place on record my profound gratitude and heartfelt thanks to the Chairman and my Supervisor, **Dr. Jabar Zaman Khan Khattak**, Department of Bioinformatics and Biotechnology (IIUI). I deem it as my privilege to work under his able guidance.

I am highly indebted to my Co-Supervisor **Dr. Asif Mir** (IIUI), who created idea and methodology for this research and carried out by Dr. Jabar Zaman Khan Khattak and Ms. Naureen Aslam Khattak (PMAS, Arid Agriculture University, Rawalpindi).

I would like to express my heartfelt gratitude and admiration to **Ms. Naureen Aslam Khattak** for her steady help, guidance, and concentration. She has been endowing me with her constructive observations at every stage of this research.

I am forever grateful to my friends **Mr. Sheikh Arslan Sehgal, Mr. Syed Babar Jamal, Mr. Sajjad Ahmad** and **Mr. Waqar Arshad** for their constant support and encouragement throughout my research work and made my two years in IIUI awesome. It has been a great privilege to spend two years in the Department of Bioinformatics and Biotechnology, and its members will always remain dear to me.

Finally, I thank all those who have helped me directly or indirectly in the successful completion of my thesis. Anyone missed in this acknowledgement are also thanked.

Thanks to All of You

Adnan Tahir

LIST OF ABBREVIATIONS

Å	Angstrom
BLAST	Basic Local Alignment Search Tool
CT	Computed Tomography
EGFR	Epidermal Growth Factor Receptor
FFT	Fast Fourier Transform
FGF3	Fibroblast Growth Factor 3
HGP	Human Genome Project
HNC	Head and Neck Cancer
HOPE	Have your Protein Explained
HPV	Human PapillomaVirus
ING1	Inhibitor of Growth 1
LOH	Loss of Heterozygosity
MD	Molecular Dynamics
MEGA	Molecular Evolutionary Genetics Analysis
MRI	Magnetic Resonance Image
NMR	Nuclear Magnetic Resonance
ORAOV1	Oral Cancer Overexpressed 1
OSCC	Oral Squamous Cell Carcinoma
ProSA	Protein Structure Analysis
PTEN	Phosphate and Tensin Homolog
SCCHN	Squamous Cell Carcinoma of the Head and Neck
STRING	Search Tool for the Retrieval of Interacting Genes
TNF	Tumour Necrosis Factor
TNFRSF10B	Tumour Necrosis Factor Receptor Superfamily, member 10B
UniProt	Universal Protein Resource
VMD	Visual Molecular Dynamics

LIST OF FIGURES

LIST OF TABLES

Project in Brief

Project Title

Computational Analysis of Head and Neck Cancer

Objectives

The main objectives of the project are to

- Identify and generate the 3D models of head and neck cancer nonstructural proteins through structural Bioinformatics.
- Analysis and interpretation of head and neck cancer candidate genes.
- Identify the evolutionary relationship on the basis of structural and nonstructural proteins of four Head and Neck Cancer genes by the application of a systematic Bioinformatics approach.
- Find protein-protein interactions of head and neck cancer candidate genes with appropriate interacting partners by Bioinformatics tools.
- Find inhibitor molecule for mutated protein and that might be used in drugs designing.
- Mutational analysis of the mutated candidate proteins leading to head and neck cancer.

Undertaken by

Mr. Adnan Tahir

Supervised by

Dr. Jabar Zaman Khan Khattak
Assistant Professor
Department of Bioinformatics and Biotechnology
International Islamic University, Islamabad.

Dr. Asif Mir
Assistant Professor
Department of Bioinformatics and Biotechnology
International Islamic University, Islamabad.

Operating System

Microsoft Windows 7

Technology Used

Bioinformatics Tools

MODELLER 9v10 and I-Tassar for 3D structure prediction
Rampage and ERRAT for Model Evaluation
BlAST, ChemDraw, HOPE server, STRING
MEGA 5 for Phylogenetic analysis
GRAMM-X and PatchDock for protein-protein docking
Autodock Vina for protein-ligand docking
PyMOL, VMD, Chimera, Ligplot for visualization

Start Date

March 12, 2012

Completion Date

November 12, 2012

Abstract

Head and neck cancer (HNC) is a complex and morbid disease that is characterized by biological, clinical and pathologic heterogeneity. Progression and evolution of HNC result from numerous alterations of molecular and cellular pathway in squamous epithelium. Genes directly involved in disease and its pathway were scrutinized based on a literature survey and *in-silico* analysis. Four candidate genes were shortlisted as most susceptible genes for disease etiology namely *TNFRSF10B*, *ING1*, *PTEN* and *ORAOV1*. MODELLER 9v10 was employed to predict 3D structures of *TNFRSF10B*, *ING1* and *PTEN* proteins. Threading based I-Tasser server was applied in structure prediction of *ORAOV1* protein. Mutational analysis of *PTEN* and *ING1* showed alterations that were located within the domain surface lead to the disturbance in domain and binding affinities with other molecules. Molecular Evolutionary Genetics Analysis (MEGA5) demonstrated that TNFRSF10A and ING5 genes were ancestor in *TNFRSF10B* and *ING1* neighbour joining tree respectively. Ciona intestinalis was an outgroup in *PTEN* and *ORAOV1* orthologous tree. Primates, mammals, rodents, birds and teleosts were their appropriate place and concealing with the specie tree. Human was closely related to chimpanzee and gorilla showing >70% sequence similarities even 100% similarities are also identified.

Protein-ligand docking revealed LYS148, THR153, and GLY141 residues of receptor protein showed hydrogen interactions with prostaglandin F2alpha ligand. Hydrophobic interactions were identified among ALA122, LEU137, VAL118 and PRO152 with ligand.

Protein-protein docking was employed for *TNFRSF10B*, *ING1* and *ORAOV1* proteins and functional partners of receptor proteins were used as a ligand proteins. *TNFRSF10B* receptor protein residues (ILE58, SER90 and ALA62) showed ionic interactions with *TNFSF10* ligand residues (ARG130, SER156 and ARG130). Ionic interactions were observed in the receptor protein *ING1* residues (CYS240, ASP167 and ASP167) with *TP53* residues (CYS231, ARG60 and GLU137) respectively. *ORAOV1-FGF3* Docked complex showed ionic interactions of GLU28, ARG24, TYR23, ILE53, ARG24 and ARG24 receptor residues with ARG-144, GLU-82, GLU-82, ARG-132, ARG-132 and THR-136 ligand residues respectively.

The current computational analysis will help to analyse and identify the complexity of HNC and hypothesized that docking studies of candidate genes will assist in drug development to cure the 6th most occurring cancer i.e. Head and Neck Cancer. There is an urgent need to develop specific databases and tools to identify candidate genes and their mechanisms in pathways for the reduction of complexity of the disease.

CHAPTER 01

INTRODUCTION

Sequencing of Human Genome Project (HGP) and other species produced a vast amount of biological data (Fuchs, 2002). Large number of proteins is identified based on sequencing data but addition to known sequence, predicted structure is still a mystery for many proteins and the process of identifying structural insights is favourably under consideration. Due to the increasing number of protein sequences there was need to develop an automated method for deriving the three dimensional (3D) structures of proteins from sequences (Chou, 2004). Therefore, it is extremely desired to develop a field that will allow access, interpret and analysis of the biological data by computational analysis. To overcome these hurdles, a new field emerged known as bioinformatics that provides *in-silico* solutions of biological problems. The major aims of bioinformatics are, to store and organize data for scientists to access, to develop tools for data analysis and then interpret the results.

Over the past few years, development of genomics and proteomics based applications have greatly contributed in boosting the development of bioinformatics techniques. A few decades before, this field expanded in different number of markets mainly in industrial and agricultural biotechnology and in pharmaceutical industry (Rashid, 2006).

Currently, bioinformatics is mainly focusing on newly identified gene functions, structure prediction by using protein sequences encoded by nucleotide sequences, finding regulatory elements, repetitive sequence analysis, inference of evolutionary distances for phylogenetic analysis, finding enzymes active sites, mutational analysis, primer designing, finding interactions of proteins leading to the protein-protein docking.

Head and Neck Cancer (HNC) cover a wide range of malignant tumours that appear in larynx, nasal cavity, pharynx, oral cavity and paranasal sinuses. Almost all of these tumours originate from epithelial cells and called Squamous Cell Carcinoma of the Head and Neck (SCCHN) (Argiris and Eng, 2003). It is the sixth most common cancer worldwide (Parkin *et al.*, 2002) but more frequent in South East Asian countries (Argiris *et al.*, 2004; Ries *et al.*, 2006). HNC is the second most frequent type prevalent cancer in Pakistan (Faheem, 2007), represents 40.1% incidence of all the cancers (Hanif *et al.*, 2009). Worldwide, HNC incidence variations is

highly attributed to demographic differences in the habits of chewing tobacco, exposure to smoking and drinking alcohol (Barnes *et al.,* 2005). This view has evolved with the advent of twenty first century because the identification of new subtypes of SCCHN that differs not only in etiology but also in clinical effects and pathogenesis (French *et al.,* 2004; Westra, 2009; Ang *et al.,* 2010).

There is always a lifetime risk of dying from second primary tumours, respiratory and cardiac diseases to the survivors of SCCHN (Argiris *et al.,* 2004). Every year, 3-5% second primary tumours arise that affects the aerodigestive tract.

About 3-5% second primary tumours develop and affect the aerodigestive tract in every year. There is no proved imaging and well known biomarker is present for surveillance of patient (Khuri *et al.,* 2006). Tobacco and alcohol are the main risk factors (Argiris and Eng, 2003), but Human Papillomavirus (*HPV*) can also cause specific subsets of SCCHN (D'Souza *et al.,* 2007). 75% of all the SCCHN are due to the consumption of alcohol and tobacco and combination of these greatly increasing the risk of SCCHN (Blot *et al.,* 1988; Vineis *et al.,* 2004). Different genetic polymorphisms are reported in enzymes involved in the metabolization of alcohol and tobacco greatly increases the risk of SCCHN (Sturgis and Wei, 2002; Hashibe *et al.,* 2006).

The familial inheritance and periodically development of SCCHN has been reported (Suarez *et al.,* 2006). Cancer susceptibility syndromes like Fanconi's anaemia, Li-Fraumeni syndrome, colorectal cancer and ataxia telangiectasia increase the SCCHN risk in individuals (Foulkes *et al.,* 1996).

HPV primarily types 16 and 18 are newly identified as risk factors of SCCHN. *HPV* type 16 has higher effects than type 18 (D'Souza *et al.,* 2007). The involvement of *HPV* in SCCHN is strongly associated with tonsil cancers, intermediary with oropharynx cancer and weakly with larynx and oral cavity cancer (Hobbs *et al.,* 2006). Immunohistochemistry and in-situ hybridization techniques are applied for the detection of Oncogenic *HPV* subtypes. *HPV* genomic DNA is present in about 25% of the SCCHN (Trizna and Schantz, 1992).

Early recognition of signs and symptoms are very important for SCCHN diagnosis. Population screening is the only evidence based method for oral cancer screening in high risk

regions (Sankaranarayanan *et al.,* 2005). For specialist assessment of biopsy in malignant tumours, referral criteria have been developed to confirm the diagnosis (Witcher *et al.,* 2007). Therapeutic decision depends upon the most important factor accurate staging.

Computed Tomography (CT) and MRI (Magnetic Resonance Image) for neck imaging are staging methods employed in radiological and surgeon assessment (Patel and Shah, 2005). Combined CT and PET scans provides more accurate results in the identification of malignant tumours in head and neck (Branstetter *et al.,* 2005). PET-CT scanning is also valuable for detection and assessment of response of recurrent or persistent SCCHN. Radiotherapy and surgery have been the main approaches for treatment of SCCHN (Argiris *et al.,* 2008).

1.1. Pathogenesis

Genetic events results in the inactivation of tumour suppressor genes and activation of proto oncogenes or both, leading to the development of SCCHN. Epigenetic and genetic variations in invasive and premalignant lesions could be identified by molecular techniques. Molecular techniques permit the delineation of a theoretical succession model for SCCHN carcinogenesis (Ha and Califano, 2006; Perez-Ordonez *et al.,* 2006). Inflammatory and stromal cells contribute in the carcinogenesis and the resistance of treatment. Immortalization and telomere maintenance is controlled by telomerase that has been found to be reactivated in premalignant and in 90% of SCCHN lesions (McCaul *et al.,* 2002).

In 70-80% of SCCHN, a very frequent genetic aberrations 9p21 loss is observed (Mao *et al.,* 1996). Promoter hypermethylation, point mutations and homozygous deletions inactivates the p16, and loss of 3p could be early events in carcinogenesis of SCCHN (Rocco and Sidransky, 2001; Perez-Ordonez *et al.,* 2006). In 50% of all the SCCHN cases, point mutations in TP53 and loss of heterozygosity of 17p are observed (Balz *et al.,* 2003). Disruptive TP53 mutations in tumour DNA are known to be associated with SCCHN surgical treatment having reduced chances of survival (Poeta *et al.,* 2007). Overexpression of cyclin D1 and amplification of 11q13 are also identified in SCCHN and could associate with the behaviour of more aggressive tumour (Capaccio *et al.,* 2000). EGFR (Epidermal Growth Factor Receptor) is the member of ErbB

growth factor receptor tyrosine kinase family which is central to the biology of SCCHN (Karamouzis *et al.*, 2007).

Malignant cells of HNC inhibit or evade the immune recognition defences and flee from destruction by immune systems (Kuss *et al.*, 2004). Many mechanisms of immune evasion are proposed including escaping from immune recognition system and elimination (Ferris *et al.*, 2005; Lopez-Albaitero *et al.*, 2006) by different tumour factors, immunosuppressive cells activity, activity of T-lymphocyte and cytokines mediating systemic and local effects (Argiris *et al.*, 2008).

1.2. Candidates Genes

Generally, candidate genes have known biological functions and regulate the biological processes of traits directly or indirectly. In association analysis, these genes could be confirmed in the effects of causative variants of genes. Currently, candidate gene approach has been applied in different research areas to understand the relationship of genes with diseases, to identify genetic association and in the selection of drug target and biomarker in various organisms from animals to humans (Tabor *et al.*, 2002). This approach is highly subjective in the selection of candidates from different potential possibilities. Many web-based resources and databases allow users to mine existing data for new targets of candidate genes (Zhu and Zhao, 2007).

1.2.1. Tumour Necrosis Factor Receptor Superfamily, member 10B (*TNFRSF10B*)

Tumour Necrosis Factor (TNF) is a mediator pro-inflammatory cytokine that is involved in the progression and development of cancer. The family of TNF inhibits tumour formation through apoptosis, but TNFs deregulation encourages metastasis, migration and invasion of tumour cells tissues (Matthew *et al.*, 2012).

TNFRSF10B gene contains 10 coding exons and the allelic loss of 8p was observed in 20 primary HNCs during sequence analysis of all these coding exons. In 1998, 2-bp insertion in this gene at residue 1065 was found that introduces a premature stop codon lead to truncated protein

in SCCHN. Sequence comparison between patient and normal tissues confirmed that germline contains truncating mutation in the absence of p53 mutation (Pai *et al.*, 1998).

1.2.2. Inhibitor of Growth 1 (*ING1*)

ING1 is a member of newly identified tumour suppressor family involved in cellular senescence, oncogenesis, apoptosis and control of DNA damage repair (Garkavtsev *et al.*, 1996; Kuzmichev *et al.*, 2002). Garkavtsevc and Riabowol, (1997) mapped the *ING1* gene at 13q33-q34 location by fluorescence in situ hybridization. Zeremski *et al.* (1997) localized the *ING1* gene to 13q34 by using radiation hybrid mapping technique. Gunduz *et al.* (2000) found the loss of heterozygosity in 68% of tumours from 32 informative SSCH cases at 13q33-q34 location of *ING1* gene. Three silent mutations and three missense mutations are identified in an *ING1* gene in 6 of 23 tumours with the allelic loss at 13q33-q34 region. These missense mutations were recognized within PHD finger domain and nuclear localization motif in protein of *ING1* which possibly abrogate its normal function.

P53-dependent signalling pathways are associated with different functions of *ING1* gene. P53 genes including proapoptotic factor Bax, and cell cycle progression p21WAF1 inhibitor were activated by over expression of *ING1* (Garkavtsevc and Riabowol, 1997; He *et al.*, 2005). Latest studies showed that *ING1* is involved in the control of cellular responses to genotoxic stresses and to DNA damage (Cheung *et al.*, 2001; Scott *et al.*, 2001). In case of DNA damage *ING1* translocates to the nucleolus, starts the arrest of cell cycle and then repair to DNA damage and it induces apoptosis. According to Kuzmichev *et al.* (2002), Skowyra *et al.* (2001) and Doyon *et al.* (2004) this gene is also very important in chromatin remodeling.

Mutations in *ING1* gene results in the progression of rapid cancer, instability of genome and less survival rate for the patients of cancer. Silent and missense mutations are found in 13% SCCHN, 12.9% esophageal Squamous Cell Carcinoma (SCC) and 19.6 malignant melanoma (Campos *et al.*, 2004).

1.2.3. Phosphate and Tensin Homolog (*PTEN*)

In 1997, *PTEN* was identified as a candidate tumour suppressor gene on chromosome 10 by mapping of homozygous deletion on 10q23 (Li *et al.*, 1997; Steck *et al.*, 1997; Li and Sun, 1997). It is frequently mutated in many human cancers (Bonneau and Longy, 2000). It inhibits the progression of cell cycle by activating proapoptotic molecules and down-regulating cyclin D1 so that playing an important role in apoptosis and cell survival (Tsugawa *et al.*, 2002; Weng *et al.*, 2001)

PTEN is a good phosphoinositide D3-phosphatase and a poor protein phosphatase (Maehama and Dixon, 1998). *PTEN* antagonizes phosphoinositide 3-kinase (PI3K) which in turn activates the antiapoptotic and proloferative signalling pathways (Stambolic *et al.*, 1998; Wang *et al.*, 2000). It is regulated through phosphorylation of a cluster of threonine and serine residues in its C terminus during post translational modifications (Vazquez *et al.*, 2001; Birle *et al.*, 2002; Miller *et al.*, 2002). One alteration known as loss of heterozygosity (LOH) mostly occurs in various human tumours at 10q23 (Li *et al.*, 1997). An ala121-to-gly mutation was found in 1 laryngeal carcinoma and 1 oropharyngeal by using 52 SCCHN tumour samples (Poetsch *et al.*, 2002).

1.2.4. Oral Cancer Overexpressed 1 (*ORAOV1*)

ORAOV1 gene is a candidate proto-oncogene in different SCCs. First, Huang and his colleagues mapped the *ORAOV1* at chromosomal band 11q13. It was considered the candidate oncogene for the progression and development of different human SCCs because it was supposed as a main reason in 11q13 gene amplification (Huang *et al.*, 2002).

Gene amplification often increases the gene expression level and used as a common finding for a broad range of tumour types (Schwab, 1999). In SCCHN, high levels of 11q13 DNA amplifications are verified by different molecular studies (Garnis *et al.*, 2004; Jarvinen *et al.*, 2008; Martin *et al.*, 2008). High level of expression in cell lines of OSCC (Oral Squamous Cell Carcinoma) carrying the 11q13 amplification has been discovered by northern blot analysis.

Expression level was low in those cells that are not carrying the amplification and in human normal oral keratinocytes (Huang *et al.*, 2002).

Many functional studies explained the role of *ORAOV1* in the OSCC tumorigenesis by taking part in the tumour angiogenesis and regulation of cell growth (Jiang *et al.*, 2008). Worldwide, SCCHN deaths are resulted by one of the most common malignancies oral squamous cell carcinoma (Neville and Day, 2002). *ORAOV1* may be a very important biological biomarker in SCCs (Huang *et al.*, 2002). Rearrangements in 11q13 regions are considered the independent prognostic factors and most common tumour related chromosome region in many SCCs (Choi *et al.*, 2007; Baras *et al.*, 2009).

Highly characterized candidate genes of head and neck cancer are selected. Different bioinformatics applications will employ to understand the insights of selected candidate genes. Structural bioinformatics will assist in the structural analysis of candidate proteins. 3D structures of selected proteins were not already reported in biological databases including the Protein Data Bank (PDB). Threading and comparative modelling will perform for structure prediction that will lead to the better understanding of proteins function.

Phylogenetic analysis will perform by using orthologs and paralogs of selected genes that presents evolutionary relationship between genes. Phylogenetic tree will infer ancestral gene specie, outgroup and divergent gene species.

Interactions of candidate proteins will determine with other ligands, chemical compounds, enzymes and proteins. Interacted ligands and proteins will be further analysed and will be utilized as a ligand in molecular docking. Protein-ligand and protein-protein docked complexes will be visualized, interpreted and evaluated by bioinformatics techniques. Virtual screening leads to the drug designing to stop and cure disease. Overview of current work is shown in figure 1.1.

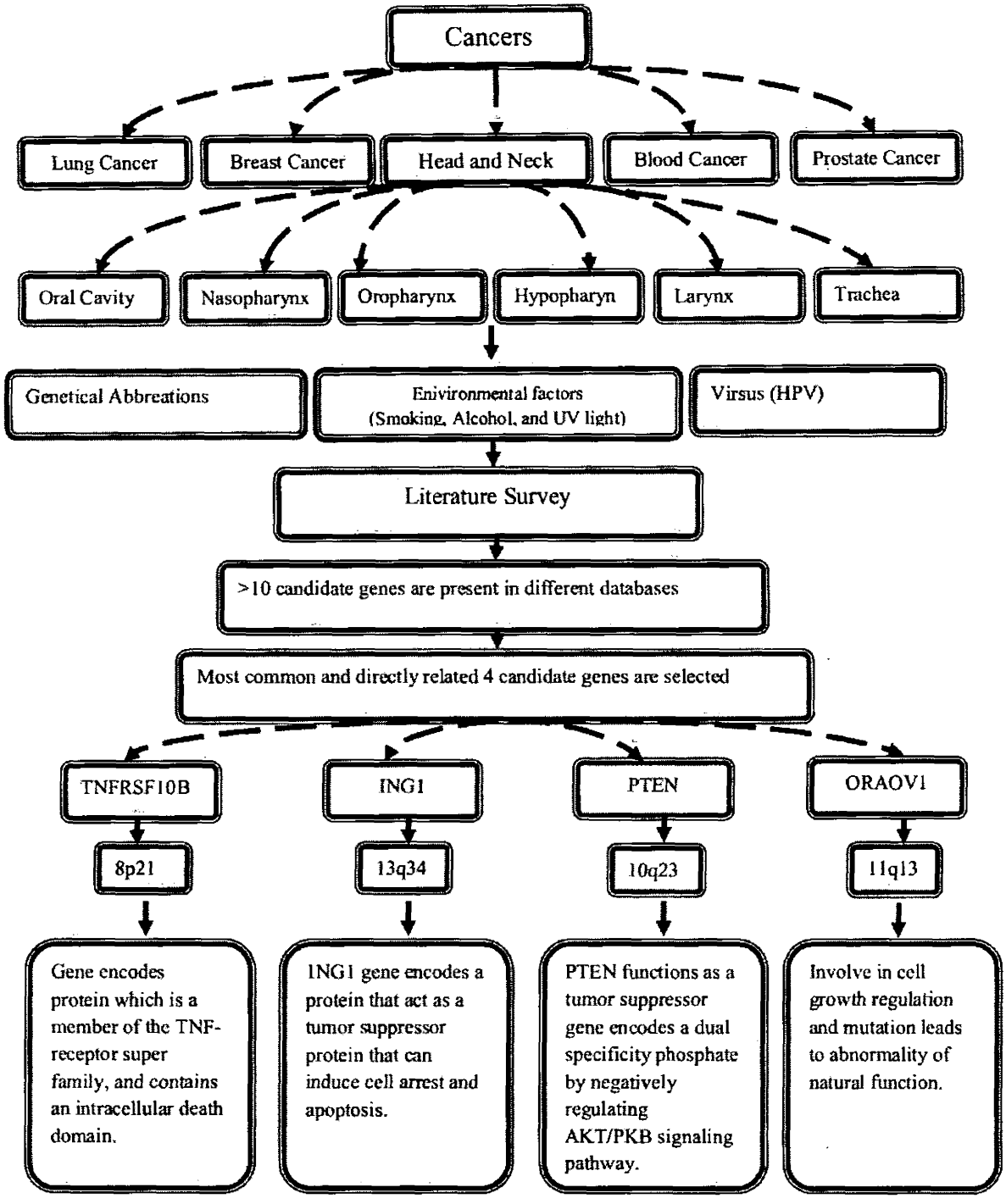


Figure 1.1: Schematic representation showing flow of data on Head and Neck Cancer

CHAPTER 02

MATERIALS & METHODS

Current research methodology illustrated computational analysis to proposed candidate genes for head and neck cancer. Four most susceptible genes (*TNFRSF10B*, *ING1*, *PTEN*, and *ORAOV1*) were recruited from mutational databases and cross checked through literature. Candidate genes were further analysed for their 3D structures and interactions. Protein sequences of candidate genes are retrieved from the Uniprot Knowledgebase. 3D structures of selected candidate proteins were not reported in the Protein Data Bank (PDB). Basic Local Alignment Search Tool (BLAST) technique was applied to find out the best templates of targeted protein for comparative modelling. Best templates were selected on the basis of query coverage, E-value and identity. 3D structures of templates were retrieved from the PDB as input of homology modelling. The comparative modelling approach was used to predict the structures of targeted proteins. MODELLER was applied for homology modelling. Threading approach was also used for structure prediction of *ORAOV1* protein which has identity less than 30% with the template. Selected 3D models of target proteins are further evaluated by evaluation tools like Rampage, ERRAT and ProSA.

To determine the evolutionary relationships of genes, Ensembl BLAST was used to find the paralogs and orthologs of genes in different species. Paralogs were selected on the basis of alignment and then orthologs sequences of selected paralogs were retrieved. Alignment of these paralogs and orthologs were carried out by the ClustalW method in Molecular Evolutionary Genetics Analysis (MEGA 5). Complete deletion option, p-distance value and bootstrap was used in tree construction by neighbour-joining method.

Docking studies were carried out for the understanding of ligand-receptor interactions and their binding. Ligand of target proteins was searched from ZINC [ucsf], PubChem, Chemspider, KEGG, and PDB databases. Chemical structure was drawn in ChemDraw and saved into .PDB format as input of AutoDock Vina. AutoDock tool was applied to find the interactions and bonding of receptor and ligand. A docked complex of receptor and ligand was visualized and determined the hydrophobic, ionic and hydrogen interactions by Visual molecular dynamics (VMD) tool. String server was used to determine the interactions of target proteins with other proteins. Protein-protein docking was carried out by GRAMM-X, HEX and PatchDock server. PYMOL tool was used for visualization and evaluation of docked complex of receptor and ligand protein.

2.1. Pubmed

Pubmed was widely used to retrieve information about head and neck cancer. Candidate genes of SCCHN are selected based on literature review retrieved from Pubmed.

2.2. UniProtKB

UniProtKB (Universal Protein Resource Knowledgebase) contains consistent, accurate and rich annotated proteins functional information. UniProtKB was used to retrieve the protein sequences, entry names, its origins, annotations, and functional information.

2.3. Basic Local Alignment Search Tool (BLAST)

BLAST is a very powerful tool used in the comparison of nucleotide and protein sequences in same and different species. Sequence similarities may be used to find the evolutionary relationship, predict protein structures, functions and their families. The BLAST is widely used tool to find the sequence similarities by searching many sequence databases. The BLAST tool searches similar regions in query sequence against different databases on the basis of local alignment (Altschul *et al.*, 1990). BLAST parameters allow user for deletions and insertions during alignments. Different types of BLAST are present including protein BLAST, nucleotide BLAST etc.

2.4. Protein Data Bank (PDB)

PDB is freely accessible database for scientists for retrieving 3D structures and structural information of proteins as an input for docking studies. PDB archive is the collection of the 3D structures of proteins, nucleic acids and complex assemblies (Raghava *et al.*, 2011). It also contains the ligands of target proteins determined by X-Ray crystallography and NMR techniques. PDB achieve contain 84508 3D structures of macromolecular till 11 September, 2012.

2.5. Structure Prediction

Threading and comparative modeling approach was employed to predict 3D structures of selected proteins through I-Tasser and MODELLER respectively.

2.5.1. Iterative-Tasser (I-Tasser)

I-Tasser is an online server for the structure and function of proteins by threading method. I-Tasser is known as no 1 server for structure prediction seen in CASP7, CASP8 and CASP9 experiments. It built the 3D structures by using LOMETS for multiple threading alignments, Iterative-Tasser for assembly simulations and then BioLiP protein functional database for matching predicted models to determine the functional insights.

2.5.2. MODELLER 9v10

MODELLER is a homology or comparative modelling program for 3D structure predictions of proteins. MODELLER automatically calculates the model having non hydrogen atoms by taking the alignment of the target sequence with known structure and a script file. MODELLER predicts the structure by the satisfaction of spatial restraints. In addition to the structure prediction, MODELLER also performs sequence comparisons, alignments, clustering, structure evaluation, and predicts the de novo modelling of loops (Eswar *et al.,* 2008; Chen *et al.,* 2011).

2.6. Structure Assessment

The assessment of predicted structures was carried out by ERRAT, Rampage and ProSA tools to check the reliability of 3D structures.

2.6.1. ERRAT

The ERRAT structure evaluation algorithm is specially developed for the assessment of crystallographic model and its refinement. Its mechanism involves the examination of the statistics of non-bonded interactions lies in different atoms. ERRAT generates an output plot that

presents the quality of a model and shows sliding window v/s error function. An error function is extremely useful in decision making regarding the reliability of the model.

2.6.2. Rampage

Ramachandran plot of structure predicted protein is visualized and assessed by Rampage program (Sian *et al.*, 2011). The plot presents the distribution of residues in favoured, allowed and outlier regions. For pro, gly, pre-pro and other general residues, psi/phi plots were also derived.

2.6.3. Protein Structure Analysis (ProSA)

ProSA is a well known tool for validation and refinement of protein structures. It determines the potential errors in 3D protein structures. It can recognize the errors in theoretical and experimental determined structures (Wiederstein and Sippl, 2005).

2.7. Phylogenetic Analysis

To generate phylogenetic trees, protein sequences of orthologs and paralogs of selected proteins were retrieved from the Ensembl genome browser. MEGA tool is used for constructing and inferring phylogenetic trees.

2.7.1. Ensembl Genome Browser

Wellcome Trust Sanger Institute EMBL-EBI developed Ensembl to maintain and produce automatic annotations of eukaryotes. Ensembl is a free online program that provides the information of vertebrates and other eukaryotic species.

2.7.2. Molecular Evolutionary Genetics Analysis (MEGA)

MEGA is an integrated tool for sequence alignments, calculate rates of molecular evolution, Phylogenetic trees inferring, sequences inferring, database mining and to test evolutionary hypotheses (Tamura *et al.*, 2011).

2.8. Molecular Docking

Molecular docking studies were performed on candidate proteins to determine their interactions and binding affinities with their respective ligands and proteins.

2.8.1. Ligand Retrieval

Appropriate and suitable ligand of target protein selection is a crucial step in protein-ligand docking. Several online ligand databases are available for ligand retrieving.

2.8.1.1. ZINC

ZINC database contains about 21 million ligands for target receptor proteins used in virtual screening. User searches ligands of target proteins in this database by name of the protein, its structure or catalogue number and then retrieves appropriate ligand (Irwin and Shoichet, 2005).

2.8.1.2. KEGG

The KEGG database consists of enzymes, chemical compounds, reactions etc. and their information. User can easily search, select and retrieve appropriate ligands, enzymes, chemicals etc. in this database. Ligand of target protein retrieved from KEGG LIGAND database.

2.8.2. ChemDraw

ChemDraw is a desktop application used for drawing chemical structures based on drag and drop method. ChemDraw suit is employed for drawing structure and saved into .pdb format as an input of docking (Mendelsohn *et al.*, 2004).

2.8.3. AutoDock Vina

AutoDock Vina is a free and easy program used for virtual screening, molecular docking, and drug discovery with a high performance and accuracy. Autodock Vina is applied for docking of predicted proteins with suitable ligands to find out the interaction between them.

2.8.4. Visual Molecular Dynamics (VMD)

VMD tool is designed to visualize and analyse the lipids, nucleic acids proteins etc. it presents a molecule in different colours and rendering methods including lines, bonds, ribbons, spheres, cylinders, backbone etc. VMD can be employed in analysis and animation of the Molecular Dynamics (MD) simulation trajectory. Here VMD used for the analysis of protein-ligand interactions.

2.9. Protein Interactions

Protein interactions are the major process in life at the molecular level. Normally, every protein interacts with any other proteins (Tovchigrechko *et al.,* 2002). So that protein interactions are crucial in understanding of the protein's nature and their functionality. Here the interactions of selected proteins were determined to exploit the functions of proteins with respect to head and neck cancer.

2.9.1. STRING

STRING (Search Tool for the Retrieval of Interacting Genes) is an online server to determine the interactions of proteins between same and different species. It provides a network of interacted proteins with query sequence and gives a highest score of highest matched protein. It also provides the most interacted protein sequence and its function (Szklarczy *et al.*, 2011).

2.9.2. GRAMM-X

GRAMM-X is a web base software used for protein-protein docking and generates complexes. It develops original GRAMM-X methodology for Fast Fourier Transformation (FFT) to utilize the smoothened potentials, knowledge-based scoring and refinement of stages. This server examined by many weeks of benchmarking and publicly uses (Tovchigrechko and Vakser, 2006). Target and ligand proteins submitted to the server for protein-protein docking.

2.9.3. HEX Server

Hex server is publicly available for protein-protein docking based on FFT. The server generates 1000 docking predictions of protein-protein docked complexes but individuals can access and download only first 20 predictions. Hex server is used in protein-protein docking.

2.9.4. PyMOL

PyMOL is a desktop program for visualizing the molecular system. This software is used for visualization and analysis of protein-protein docked complexes.

Overview of the methodology applied in the current work is presented in figure 2.1. Different bioinformatics tools, databases and online servers applied in current work are shown in table 2.1.

Table 2.1: Tools involved in the present work with their functions are listed in the table.

No.	Tool name	Function
1	UniProt	Protein Data Retrieval
2	BioGPS	Expression profiling
3	BLASTp	Template selection
4	MODELLER	Structure prediction
5	Chimera	Structure visualization
6	Rampage, ERRAT	Model Evaluation
7	AutoDock	Docking Analysis
8	HOPE server	Mutational analysis
9	STRING	Protein-Protein Interactions
10	KEGG	Ligand Retrieval
11	ChemDraw	Ligand Drawing
12	GRAMM-X	Protein-Protein Docking
13	VMD	Binding interactions of docked protein-ligand complexes
14	LigPlot	For Protein-Ligand Interaction
15	PyMOL	Binding interactions of docked Protein-protein complexes
16	ENSEMBL	Phylogenetic Sequences Retrieving
17	MEGA	Phylogenetic Analysis

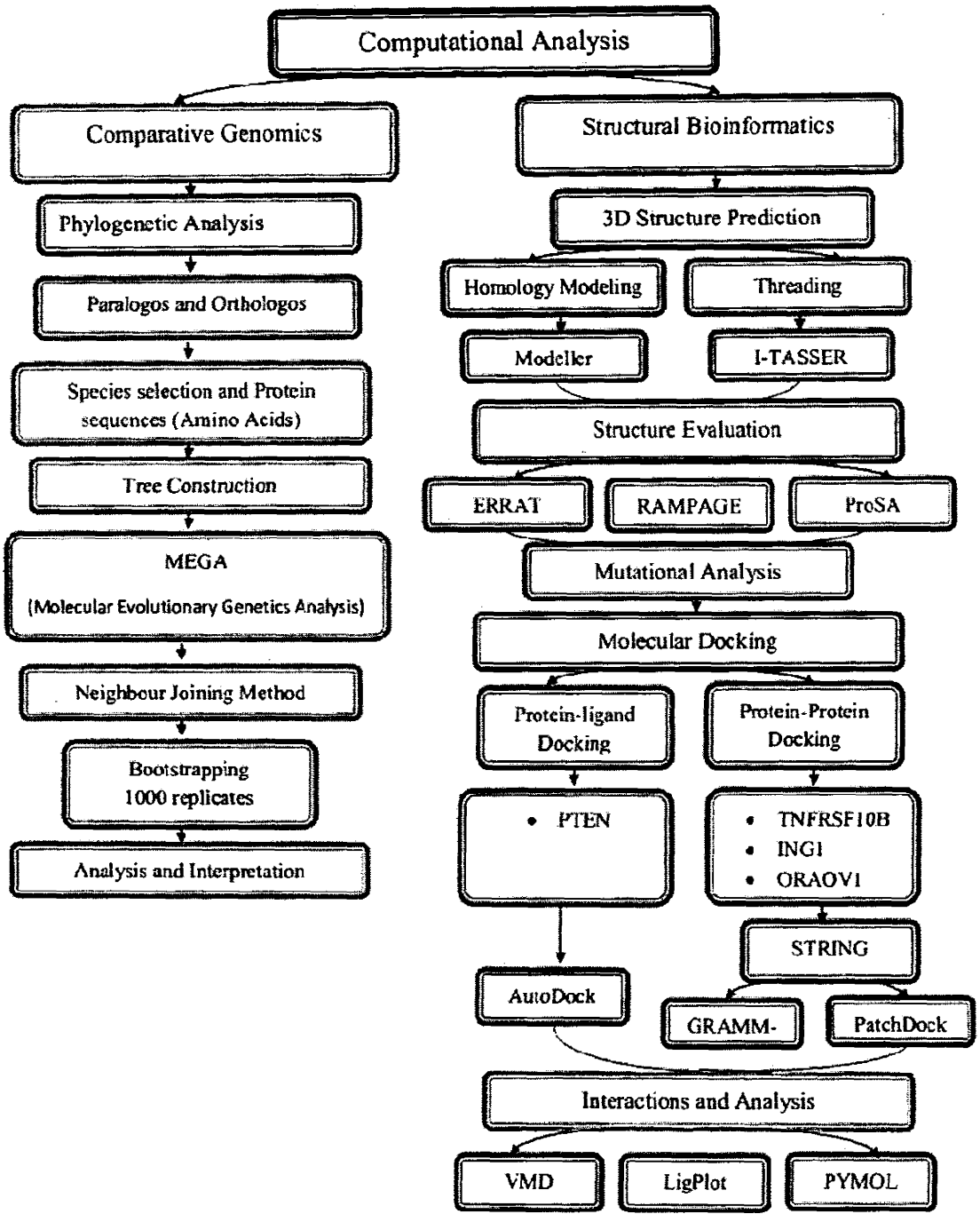


Figure 2.1: Schematic representation of the methodology applied in the present work

CHAPTER 03

RESULTS

Candidate gene approach was employed to identify the most plausible gene for disease pathogenecity. *In-silico* analysis were carried out on top four candidate genes (*TNFRSF10B*, *PTEN*, *ING1* and *ORAOV1*). Chromosomal locations, total base pairs, molecular functions, biological processes and cellular locations of prioritized genes are shown in table 3.1 and 3.2.

MODELLER was used to predict the 3D structures of *TNFRSF10B*, *ING1* and *PTEN* protein. I-Tasser based on the threading approach was applied to predict the 3D structure of *ORAOV1* protein. Evaluation tools Rampage, ERRAT and ProSA are showing the reliability of the models.

Evolutionary history and ancestral relationship of genes were determined by Molecular Evolutionary Genetic Analysis (MEGA 5). Bootstrap values representing the reliability of topologies of trees. All the species are in their appropriate place according to the specie tree.

Molecular docking permits to analyse and interpret the interactions of receptor proteins with ligands through protein-ligand and protein-protein docking methods. Protein-ligand docking was performed for *PTEN* and its ligand was retrieved from KEGG Ligand database. AutoDock Vina was used to generate docked complex of *PTEN* protein with appropriate ligand. STRING and STITCH 3 online databases were utilized to find the interactor proteins of target predicted proteins. Protein-protein docking of *TNFRSF10B*, *ING1* and *ORAOV1* were performed by docking servers GRAMM-X and HEX. Docked complexes of protein-protein and protein-ligand were visualized and analysed by PyMOL and VMD visualizing tools respectively.

Table 3.1: Prioritized set of genes with chromosomal loci, total base pairs, starting and ending base pairs.

Gene name	Chromosome #	Start Base Pair	Ending Base Pair	Total Base Pairs
TNFRSF10B	8	22877646	22926692	49,047
ING1	13	111365083	111373421	8,339
PTEN	10	89622870	89731687	108,818
ORAOV1	11	69467844	69490184	22,341

Table 3.2: Molecular functions, biological processes and cellular locations of candidate genes are mentioned in this table.

Gene	Molecular Function	Biological Process	Cellular Location
TNFRSF10B	Receptor activity Protein binding TRAIL binding	Apoptotic process Extrinsic apoptotic signalling pathway via death domain receptors. Regulation of apoptotic process. Activation of cysteine-type endopeptidase activity involved in the apoptotic process.	Plasma membrane Integral to membrane
ING1	Zinc ion binding Methylated histone residue binding.	Negative regulation of cell proliferation. Regulation of cell death. Negative regulation of cell growth.	Nucleus
PTEN	Protein serine/threonine phosphatase activity. Phosphatidylinositol-3-phosphatase activity Protein binding	Regulation of cyclin-dependent protein kinase activity. Angiogenesis Regulation of B cell apoptotic process Negative regulation of protein phosphorylation.	Nucleus Cytoplasm Cytosol Plasma membrane

3.1. Structure Prediction

Threading approach by using I-Tasser online software was applied to predict the 3D structure of *ORAOV1* protein having >30 template identity. Comparative modeling program MODELLER 9v10 was employed to build 3D structures of target proteins. Psi-BLAST was performed to find out possible templates of query sequence against PDB and templates with optimal alignment, high query coverage and least E-value was selected for homology modelling shown in table 3.3.

MODELLER 9v10 executed alignment and structure predictions of protein by using target protein sequence, template structure and script file in python as input. Predicted structure having lowest objective function was selected for further assessment by ERRAT, Rampage and ProSA evaluation tools.

3.1.1. *TNFRSF10B*

Protein sequence of *TNFRSF10B* in FASTA format was retrieved from UniprotKB with accession number E9PBT3. Table 3.4 represents the templates of *TNFRSF10B* protein sorted by their overall quality, total score and query coverage. The model was built by MODELLER by using 2ZB9 template with optimal alignment. Predicted model was visualized by Chimera 1.6 shown in figure 3.1. Assessment of predicted structures was performed by Rampage, ERRAT and ProSA shown in figure 3.2 to 3.4 respectively.

Table 3.3: Suitable templates of candidate proteins were selected based on E-value and query coverage by using BLASTp. *ORAOV1* template showed only 28% identity with query sequence so that threading approach was applied to predict structure of *ORAOV1* protein.

Protein Name	Cytogenetic location	Accession Number	Amino Acid length	Template	Query Coverage	Identity	E-Value
TNFRSF 10B	8p21	E9PBT3	90aa	2ZB9	84%	30%	5.1
ING1	13q34	Q5T9H0	279aa	4AFL	34%	36%	4e-15
PTEN	10q23	F2YHV0	175aa	1DFR	81%	89%	1e-98
ORAOV1	11q13	Q8WV07	137aa	3DOC	50%	28%	0.30

Table 3.4: Three templates of *TNFRSF10B* protein were found. 2ZB9 template having high score and query coverage was selected as a template in modelling.

Accession ID	Total score	Query coverage	E-value	Identity
2ZB9	25.8	84%	5.4	30%
3NKE	25.8	46%	4.1	32%
3NKD	25.0	46%	8.1	32%

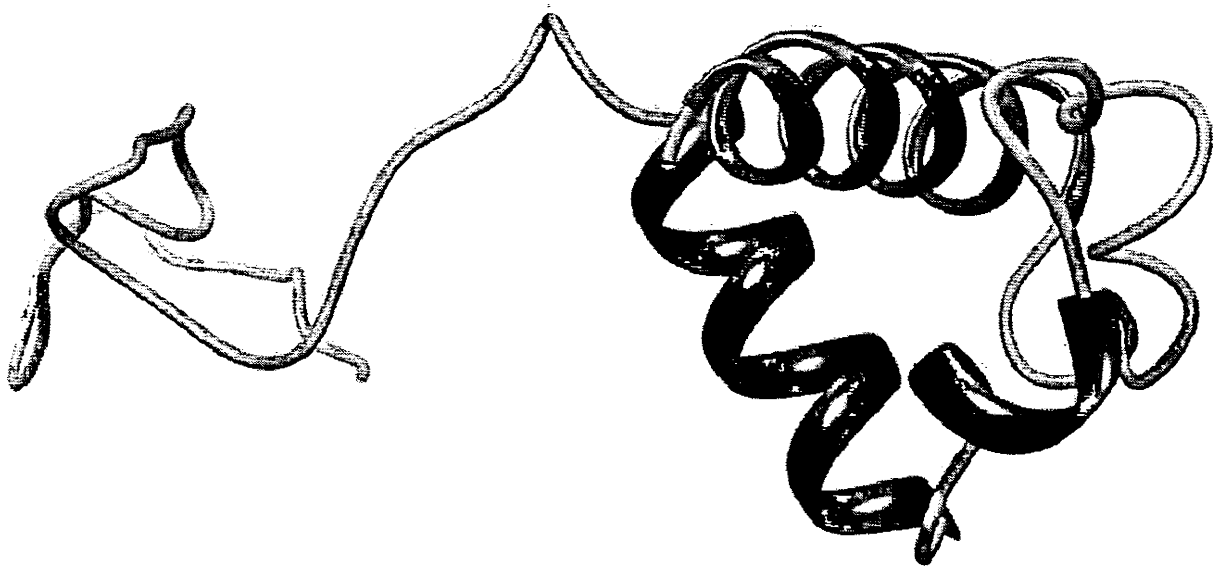


Figure 3.1: 3D structure of *TNFRSF10B* protein visualized by the UCSF CHIMERA visualizing tool. A model is presented in smoothing ribbons and sticks. Helixes are represented by blue colour and coils by green colour.

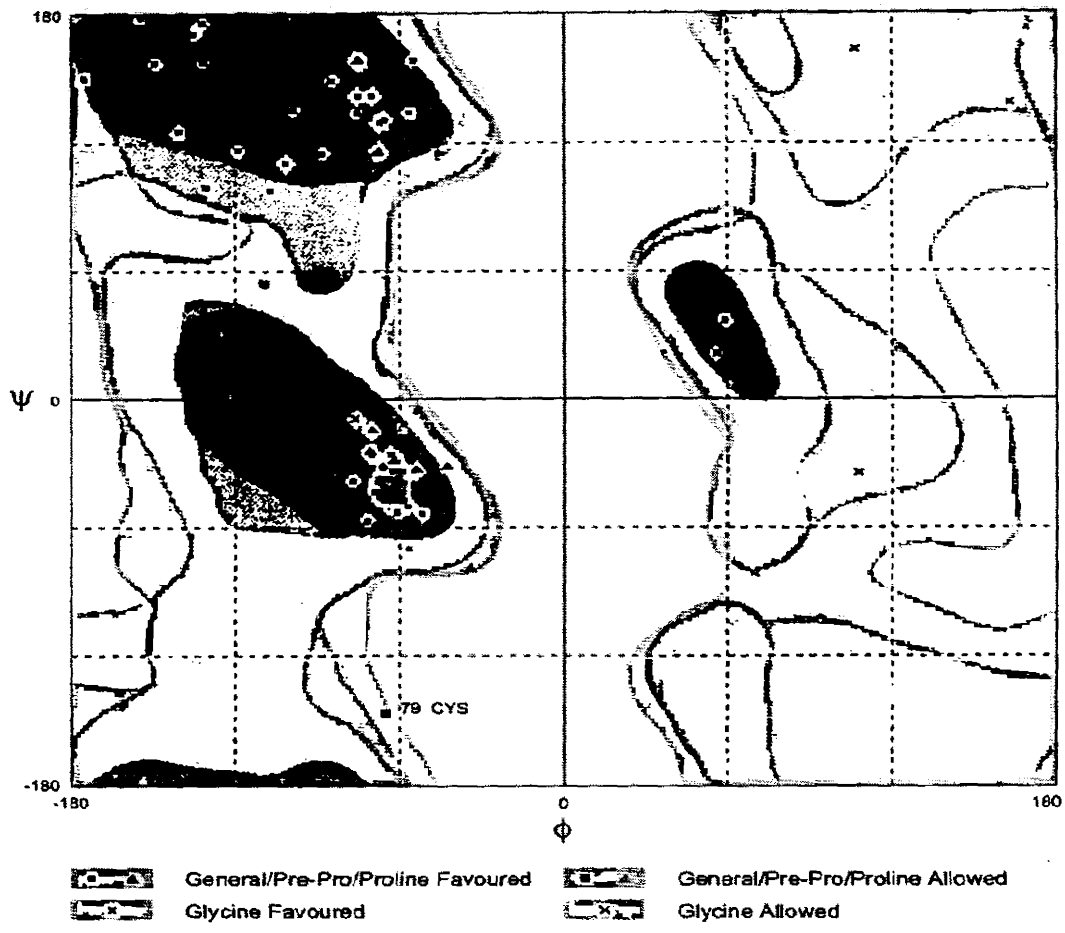


Figure 3.2: RAMPAGE result of *TNFRSF10B* protein showing the distribution of residues in favoured, allowed and outlier regions. Only 1 residue of model in outlier region representing the reliability of predicted structure.

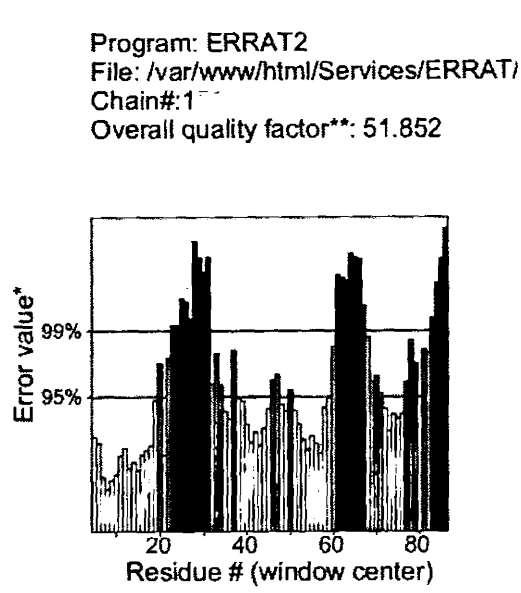
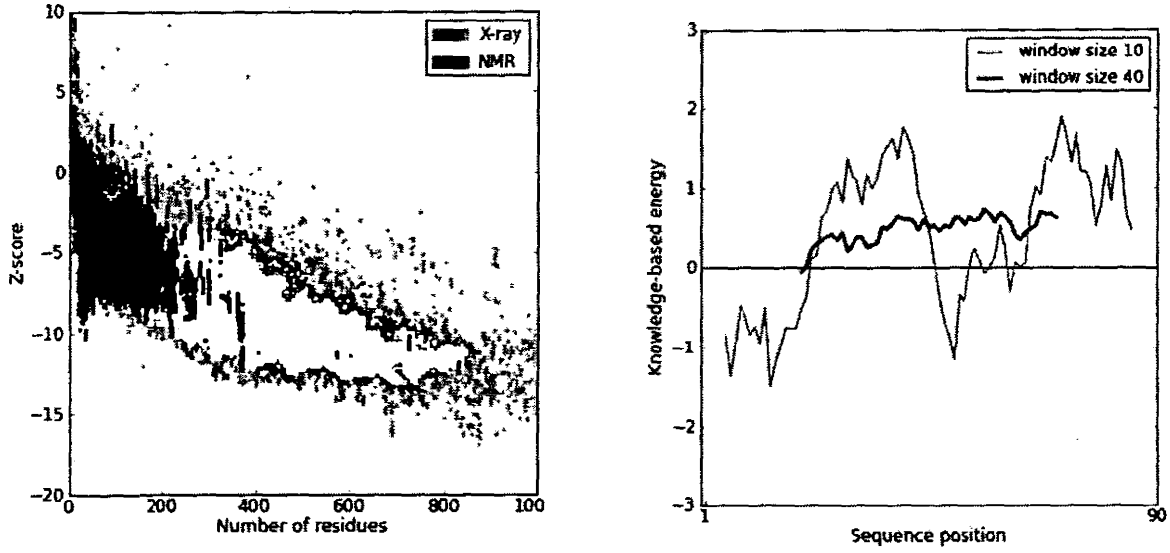


Figure 3.3: ERRAT result of *TNFRSF10B* model showing the 51.85% overall quality of model. X-axis shows the number of resides while the Y-axis represents error values of residues.



(a) Overall Model Quality Plot **(b) Local Quality plot**

Figure 3.4 (a & b): ProSA results of *TNFRSF10B* model **(a)** Plot showing -1.08 Z-score representing the overall quality of the model. **(b)** Local model quality plot of model showing position of sequences against knowledge based energy in both window sizes 10 and 40.

3.1.2. *ING1*

ING1 protein sequence in FASTA format having accession number Q5T9H0 was retrieved from UniprotKB. Top five optimal aligned templates with high query coverage are mentioned in table 3.5. 4AFL template having high query coverage was used for structure prediction by homology modelling. The predicted structure of *ING1* protein is shown in figure 3.5. Figure 3.6 to 3.8 represents an evaluation of the model by Rampage, ERRAT and ProSA respectively.

Table 3.5: Top five prioritized templates of *ING1* proteins are sorted by query coverage.

Accession ID	Total score	Query coverage	E-value	Identity
4AFL	70.1	34%	4e-15	36%
1WES	126	22%	5e-36	79%
2G6Q	124	21%	1e-35	84%
2QIC	143	21%	8e-43	98%
2VNF	116	20%	2e-32	76%

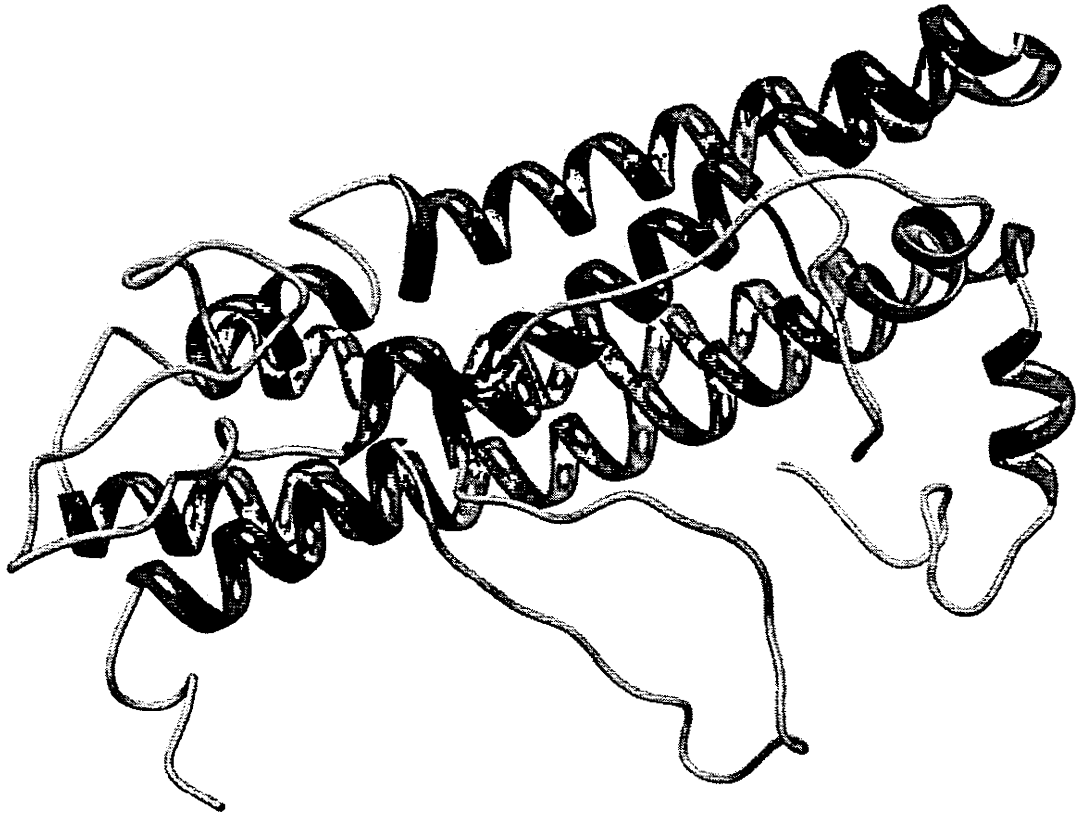


Figure 3.5: 3D model of *ING1* protein built by MODELLER 9v10 and visualized by the UCSF CHIMERA tool. Protein Structure shows helixes and coils in blue and green colour respectively.

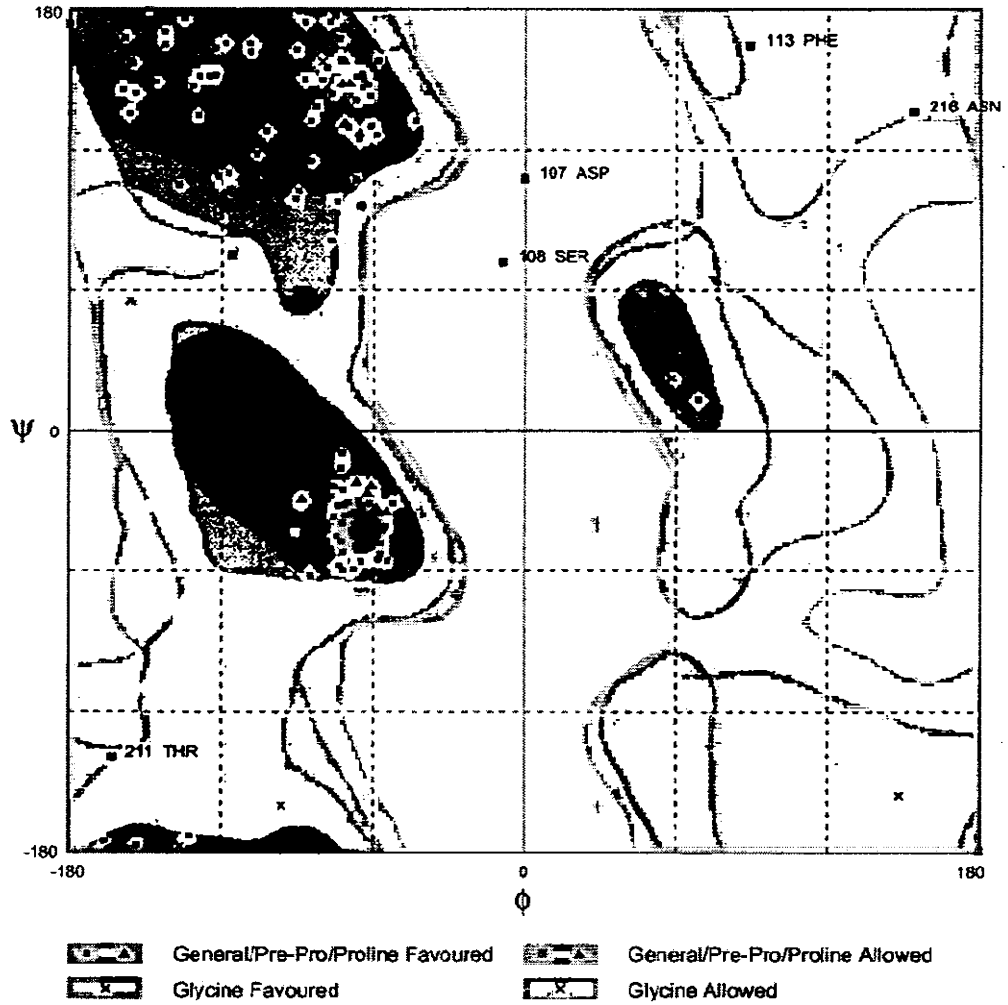


Figure 3.6: Rampage result of *ING1* model showed glycine, proline and general residue distribution in favoured, allowed and outlier regions. 92.8% residues are in a favoured region as expected residues in favoured region are about 98% and outlier region contains 1.8% residues.

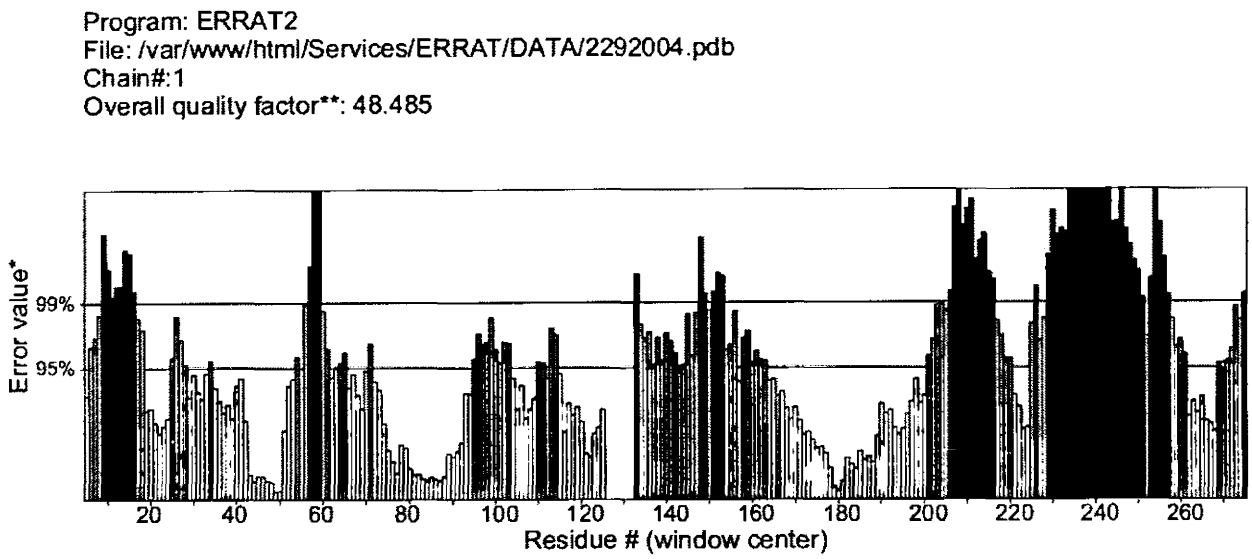
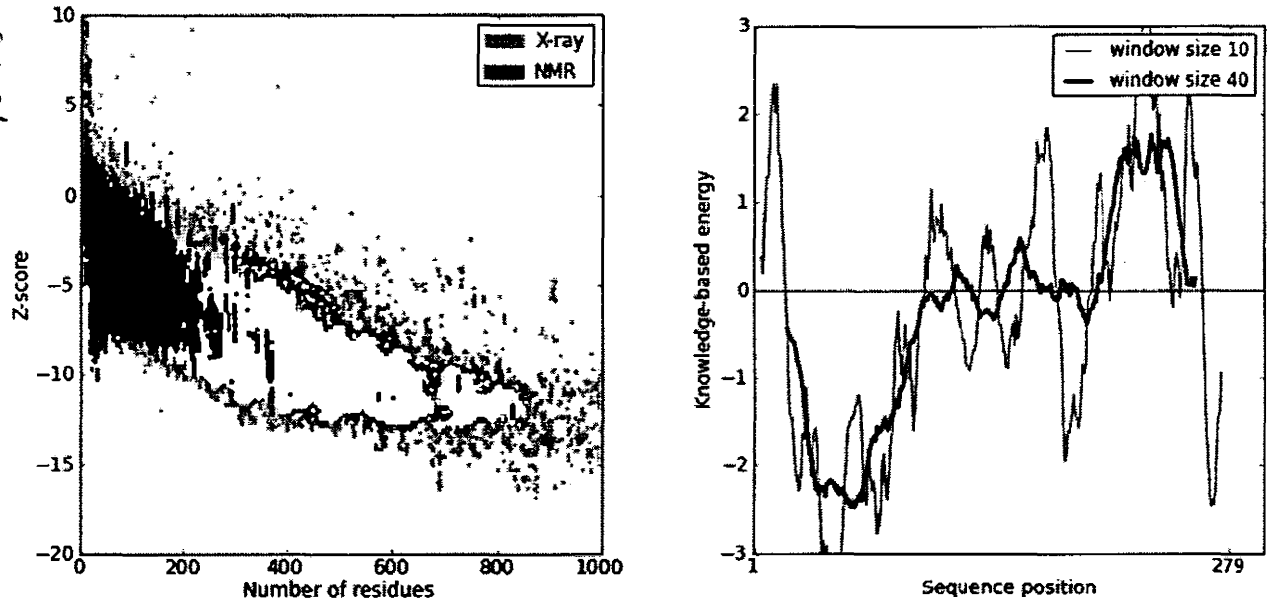


Figure 3.7: ERRAT calculated the percentage of residues that fall below the rejection limit 95% and showed overall quality factor 48.485% of *ING1* structure.



(a) Overall Model Quality Plot **(b) Local Quality plot**

Figure 3.8: (a & b): ProSA results of *ING1* **(a)** Overall quality plot showing -2.44 Z-score of model. **(b)** Knowledge based energy of the model is represented in the local quality plot.

3.1.3. *PTEN*

Protein sequence of *PTEN* in FASTA format was retrieved from UniprotKB with accession number F2YHV0. Table 3.6 lists the five templates of highest query coverage and good alignment of all the five templates sorted by their query coverage quality. Template with accession number 1DFR was used for homology modelling of *PTEN*. The model is shown in figure 3.9 and evaluation results are shown in figures from 3.10 to 3.12.

Table 3.6: Top five templates of *PTEN* protein showing their total score, query coverage, E-value and identity with the target sequence.

Accession ID	Total score	Query coverage	E-value	Identity
1D5R	291	81%	1e-98	89%
3AWF	139	89%	2e-39	41%
3AWG	137	84%	6e-39	42%
3AWE	136	84%	2e-38	42%
3V0D	135	89%	4e-38	40%

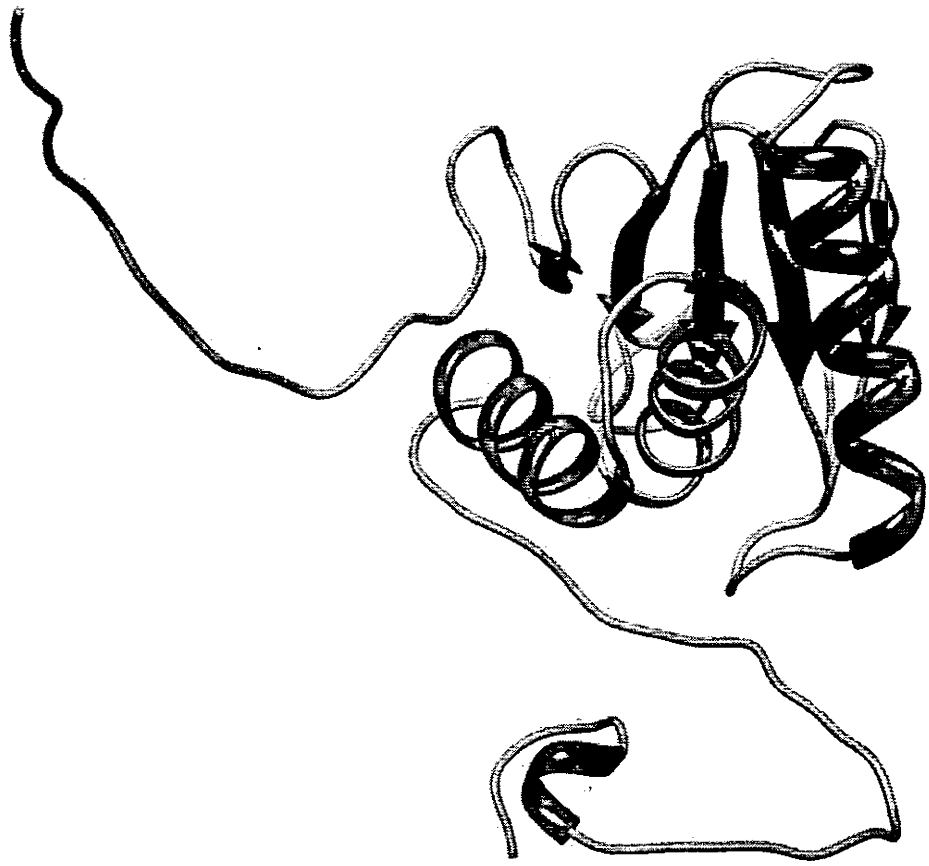


Figure 3.9: UCSF CHIMERA v1.6 was used to visualize MODELLER predicted model of *PTEN* protein. Helixes are displayed in blue colour, strands in red colour and coils are in green colour.

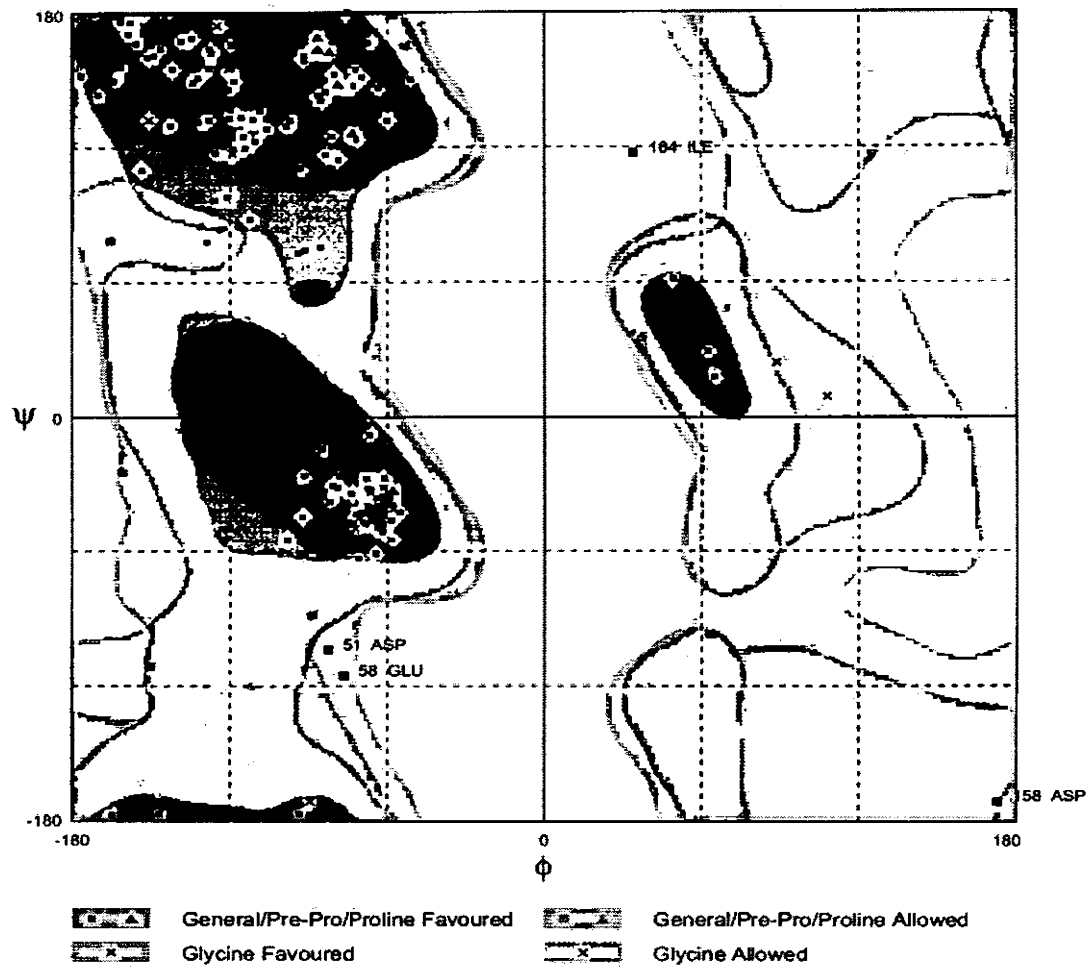


Figure 3.10: RAMPAGE showed 87.9% residues in favoured, 9.8% residues in allowed and 2.3% residues in the outer regions of *PTEN* model.

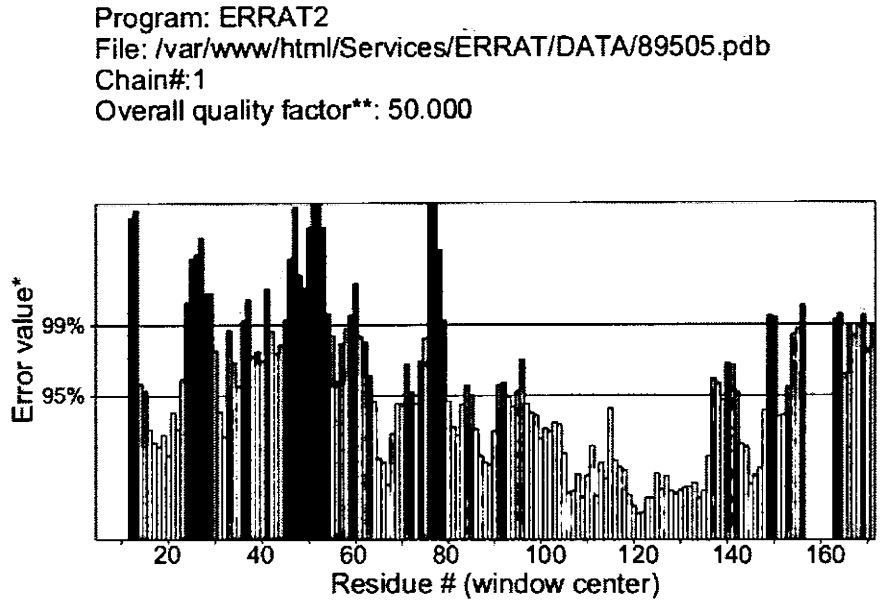
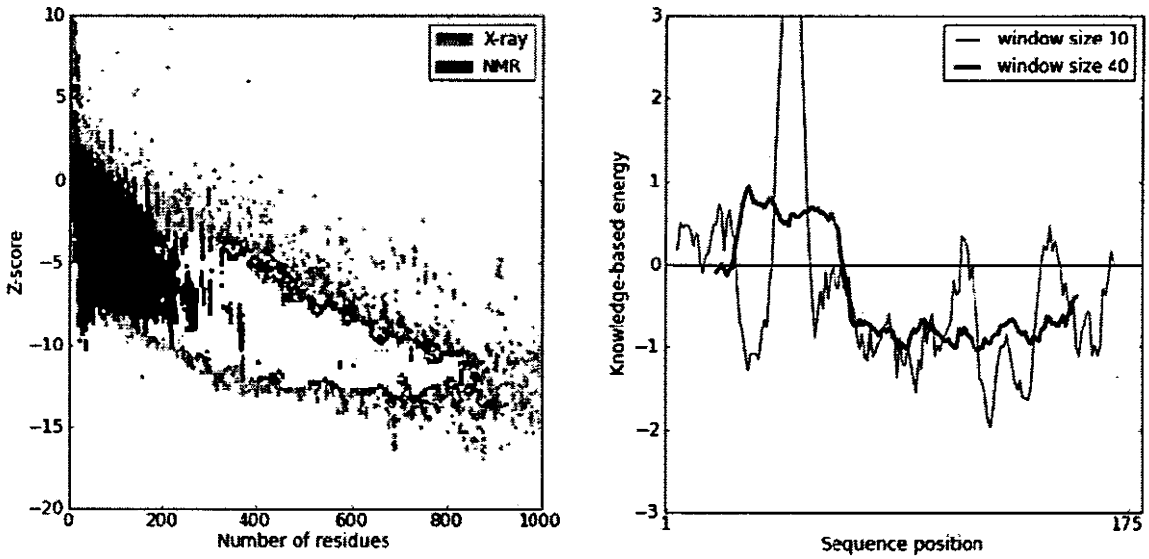


Figure 3.11: ERRAT results of *PTEN* structure showed 50% overall quality of a model. ERRAT determined overall model quality by the calculation of error values below the rejection limit.



(a) Overall Model Quality Plot **(b) Local Quality plot**

Figure 3.12 (a & b): ProSA results of *PTEN*. **(a)** Z-score is -4.24 in overall model quality plot. **(b)** Two lines of window size 10 and 40 are showing the knowledge based energy with respect to the sequence positions.

3.1.4. *ORAOV1*

The *ORAOV1* protein sequence in FASTA format was retrieved from UniprotKB with accession number Q8WV07. Psi-BLAST was applied to find out possible templates of the target protein. 3DOC template showed 28% similarity with query sequence mentioned in table 3.3. Threading based software I-Tasser was employed to predict *ORAOV1* protein shown in figure 3.13. Evaluation results by Rampage, ERRAT and ProSA are displayed in figures from 3.14 to 3.16 respectively.

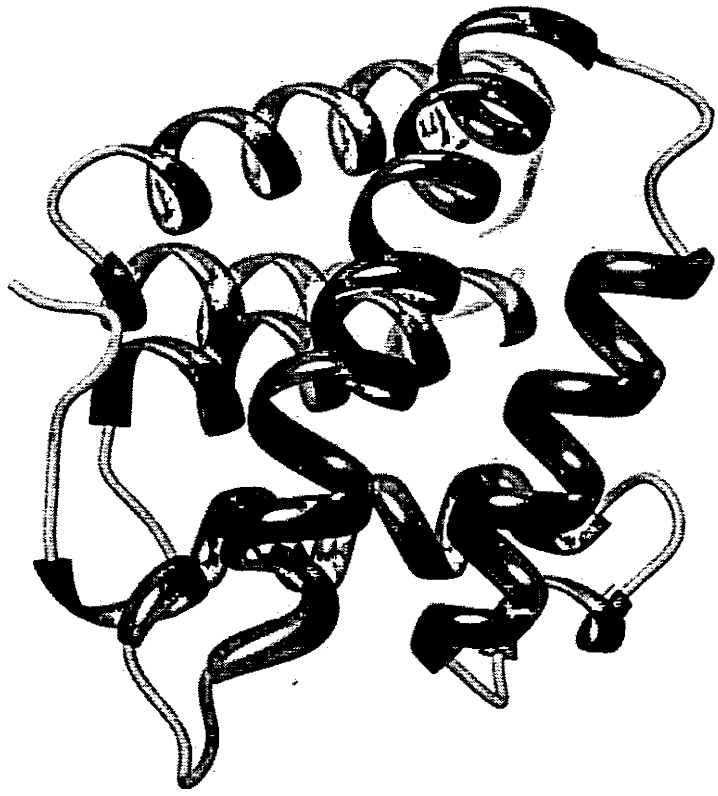


Figure 3.13: *ORAOV1* model was built by threading method through I-Tasser. The structure is presented in smooth ribbons visualized by the UCSF CHIMERA tool. Coils and helixs are represented in green and blue colour respectively.

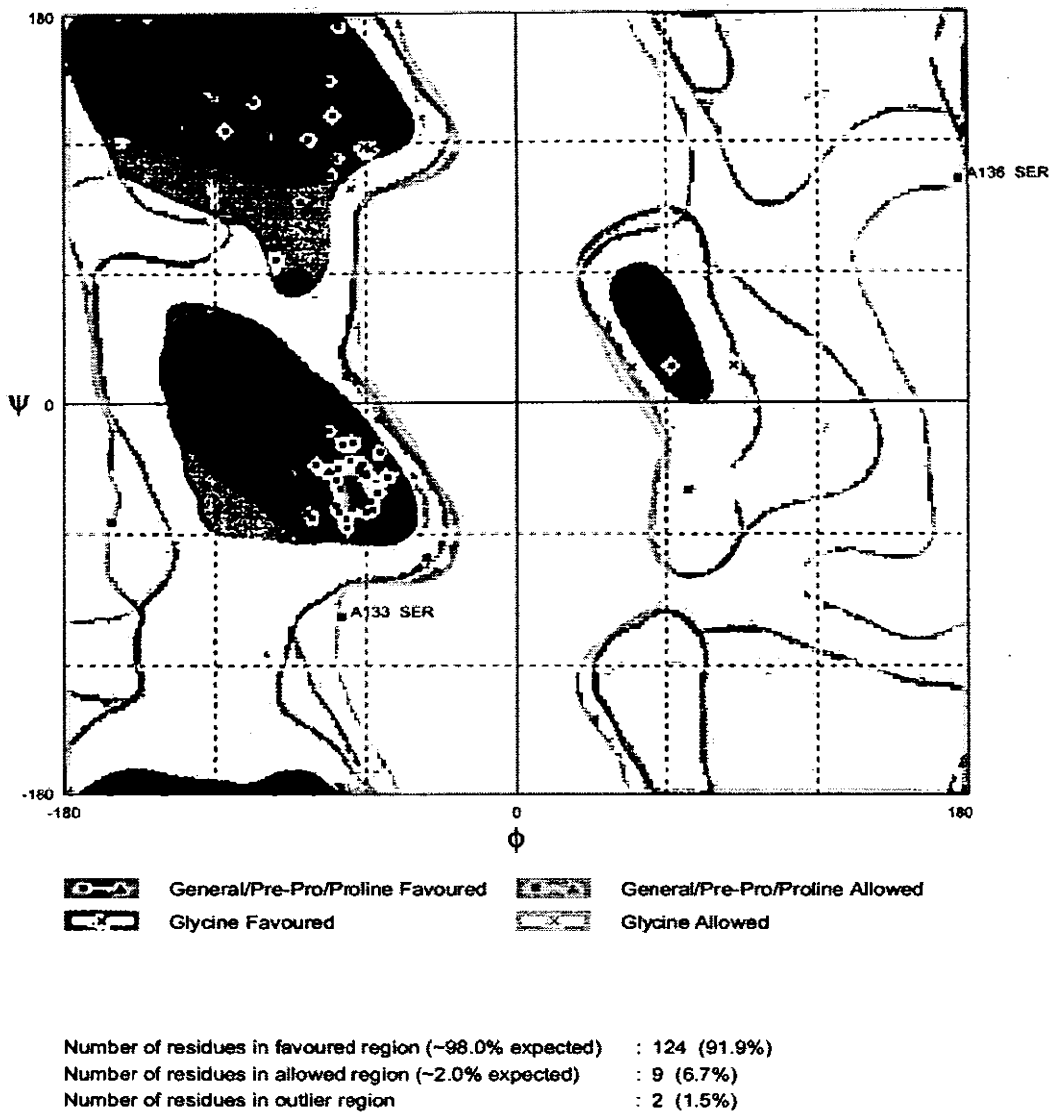


Figure 3.14: *ORAOV1* model was predicted by I-Tasser and evaluated by Rampage that defined results in favoured, allowed and outlier regions. 91.9% residues were in favoured regions, 6.7% were in allowed and 1.5% residues were in outlier regions.

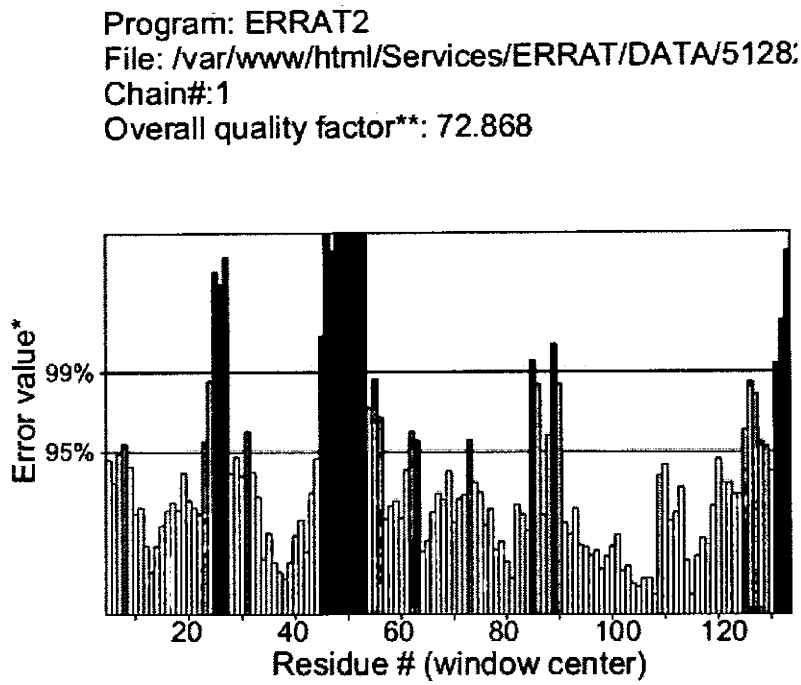
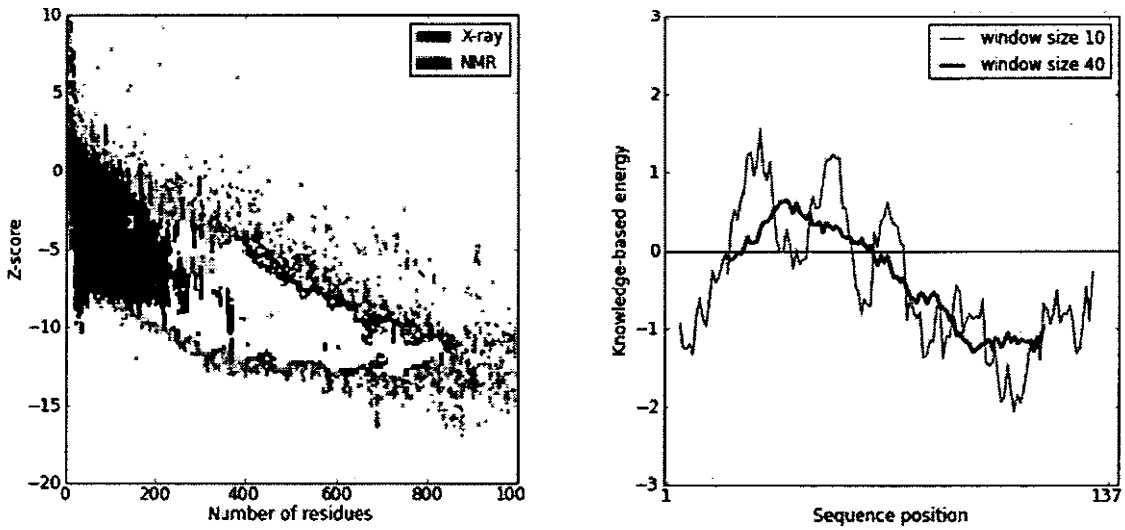


Figure 3.15: ERRAT result of *ORAOV1* model showing 72.868% overall quality factor of the model representing 72.868% residues are below the rejection limit 95%.



(a) Overall Model Quality Plot **(b) Local Quality plot**

Figure 3.16: ProSA results of *ORAOV1*. **(a)** Z-score is -3.25 in overall model quality plot. **(b)** Knowledge based energy with respect to the sequence positions are represented in the local quality plot.

3.2. Phylogenetic Analysis

Distance based approach by using the neighbour-joining method was employed to explore the characteristics of vertebrates and invertebrates of candidate genes. The evolutionary relationship of HNC genes was analysed by using protein sequences of mammals, teleosts, and tetrapods. Amphioxus sequences were also used in phylogenetic analysis as the closest invertebrate relative to vertebrate. Phylogenetic trees of selected genes are shown in figure 3.17 to 3.20.

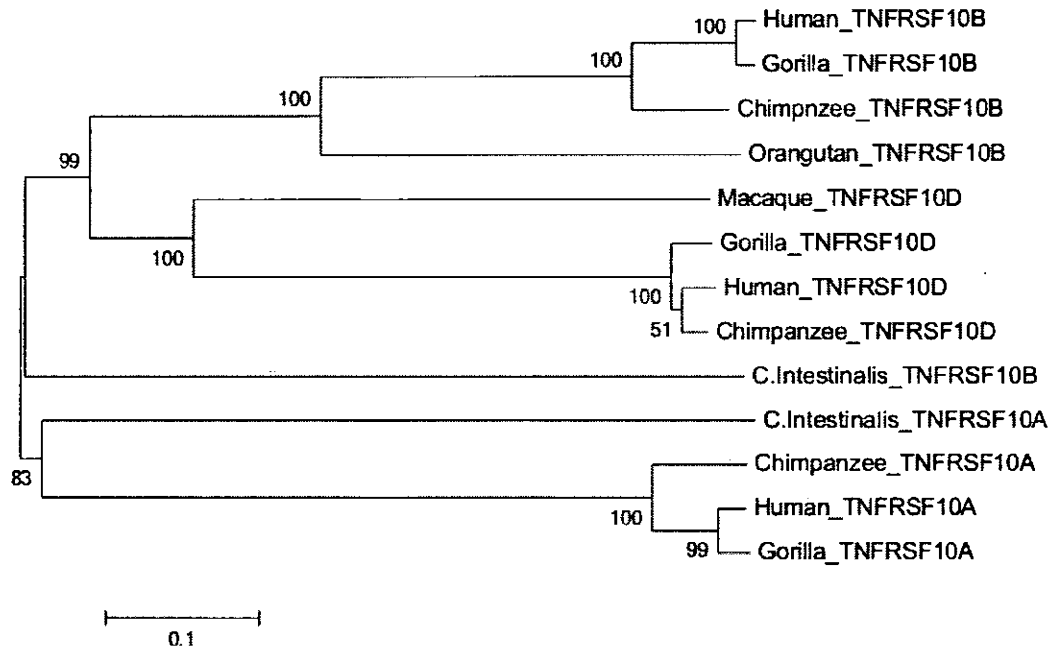


Figure 3.17: Evolutionary tool MEGA 5 was employed to construct a neighbour-joining tree of *TNFRSF10B* gene. Ensembl BLAST was performed to identify paralogos of the target gene. Protein sequences of *TNFRSF10A*, *TNFRSF10D* and *TNFRSF10B* were retrieved to determine the evolutionary relationship between paralogos and orthologos.Numbers of bootstrap replications were 1000 in bootstrap method. P-distance method and complete deletion option were used in the construction of neighbour-Joining tree. Neighbour-joining tree of *TNFRSF10B* gene represented *TNFRSF10A* gene as ancestral gene. *TNFRSF10A* gene gave rise to generation of *TNFRSF10D* and *TNFRSF10B* genes after successive mutations according to the topology of the tree.

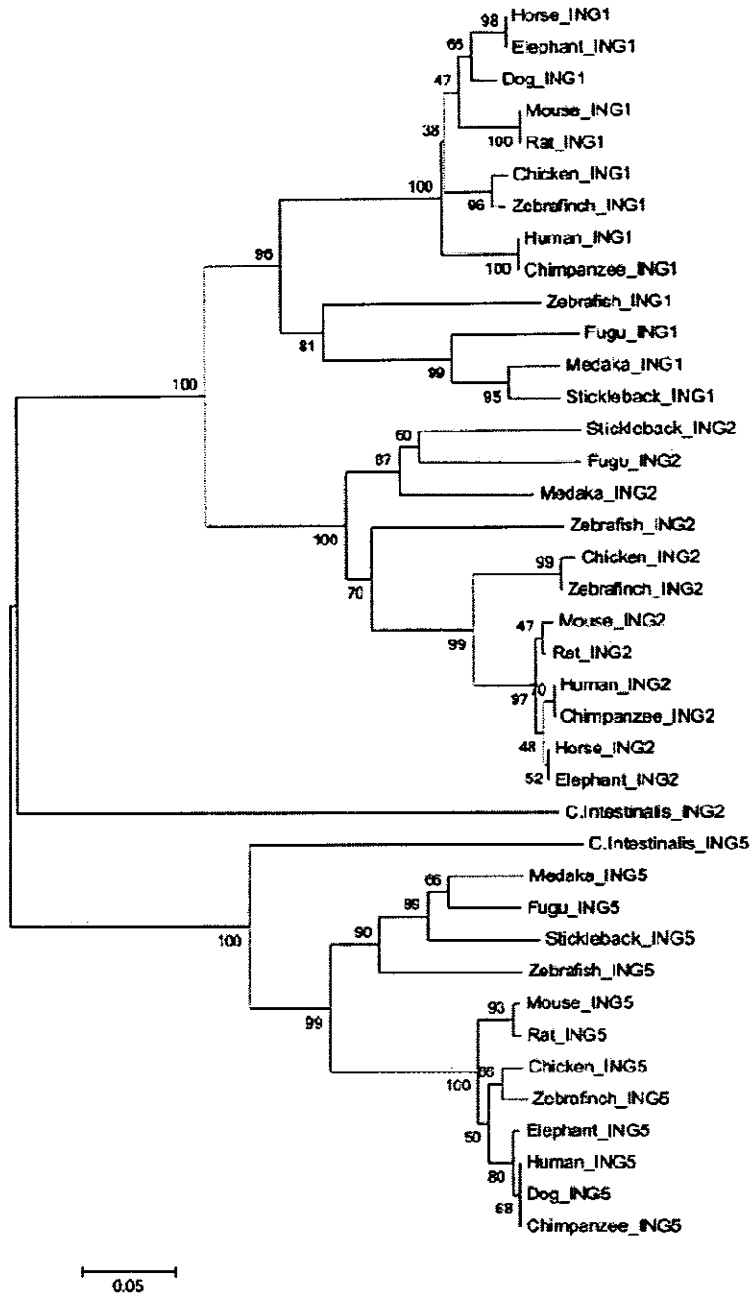


Figure 3.18: Neighbour-Joining tree of *ING1* was constructed by MEGA 5 evolutionary tool. Ensembl BLASTp/Blat was applied to determine the paralogs of *ING1* gene. Two paralogs *ING2* and *ING5* of *ING1* gene were selected to infer evolutionary history. Protein sequences of *ING1*, *ING2* and *ING5* orthologs were retrieved from the Ensembl Genome Browser. Parameters for Phylogenetic tree construction were same as above mentioned figure 17. *ING5* gene is the origin of *ING1* and *ING2* genes. Human is closely related to chimpanzee having more conserved sequences. Teleost (Stickleback, Medaka, Zebrafis and Fugu), birds (Zebrafinc and Chicken), rodents (Mouse and Rat), mammals (Dog, Elephant, Horse, Chimpanzee and Human) are concealing the species tree. The mutation rate is 0.05 per site.

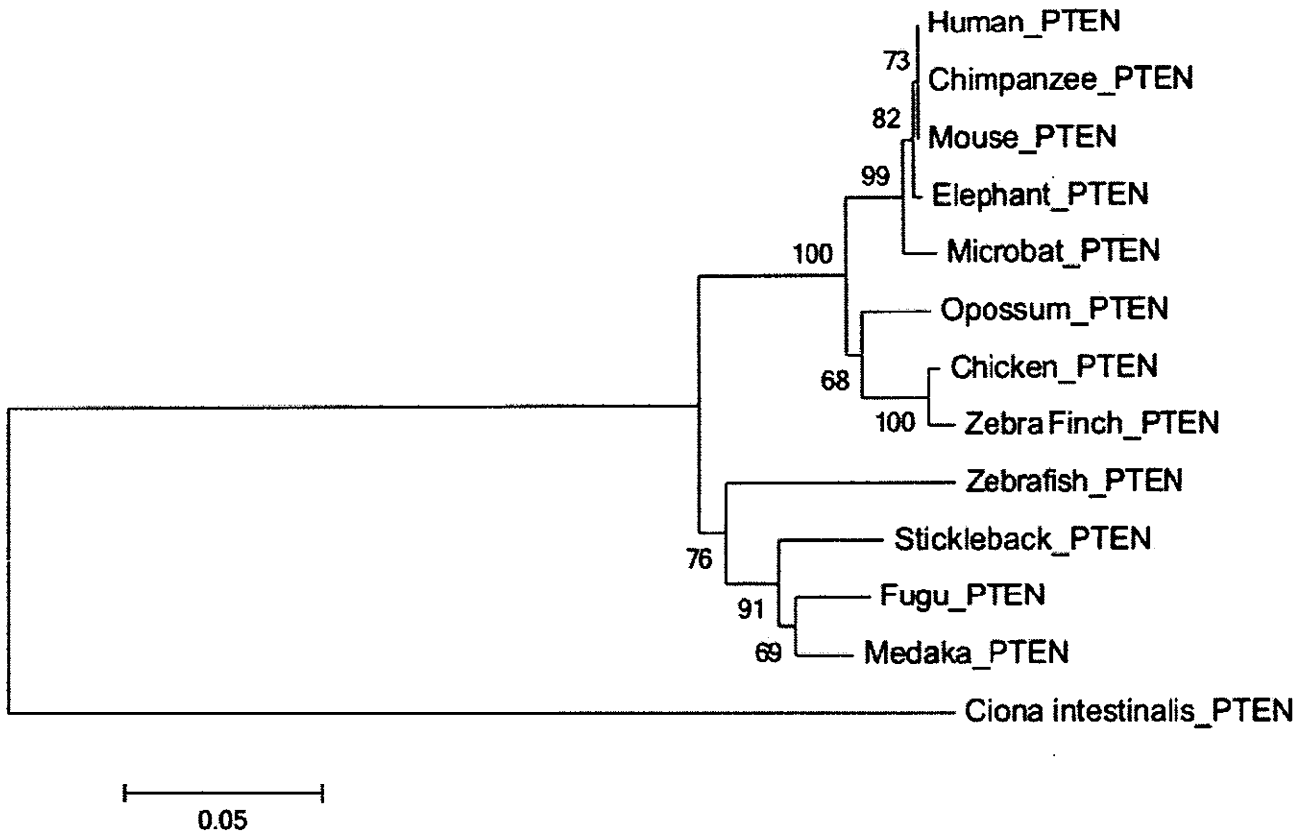


Figure 3.19: MEGA evolutionary tool generated phylogenetic tree of *PTEN* gene by Neighbour-Joining method. Paralogs of *PTEN* gene were not reported in databases and also not determined by Ensembl BLASTP/Blat. Protein sequences of *PTEN* orthologs were used for the construction of a phylogenetic tree. P-distance method and complete deletion option were used in the construction of neighbour-Joining tree. Numbers of bootstrap replications were 1000 in bootstrap method. Ciona intestinalis is lying as an outer group in this tree. Mutations in ciona intestinalis resulted in the generation of teleost and other cluster consists of birds,rodents and mammals. Human is closely related to chimpanzees having high evolutionary similarities. Mammals, rodents, birds, and teleosts are lying on their appropriate place and showing >65 bootstrap values.

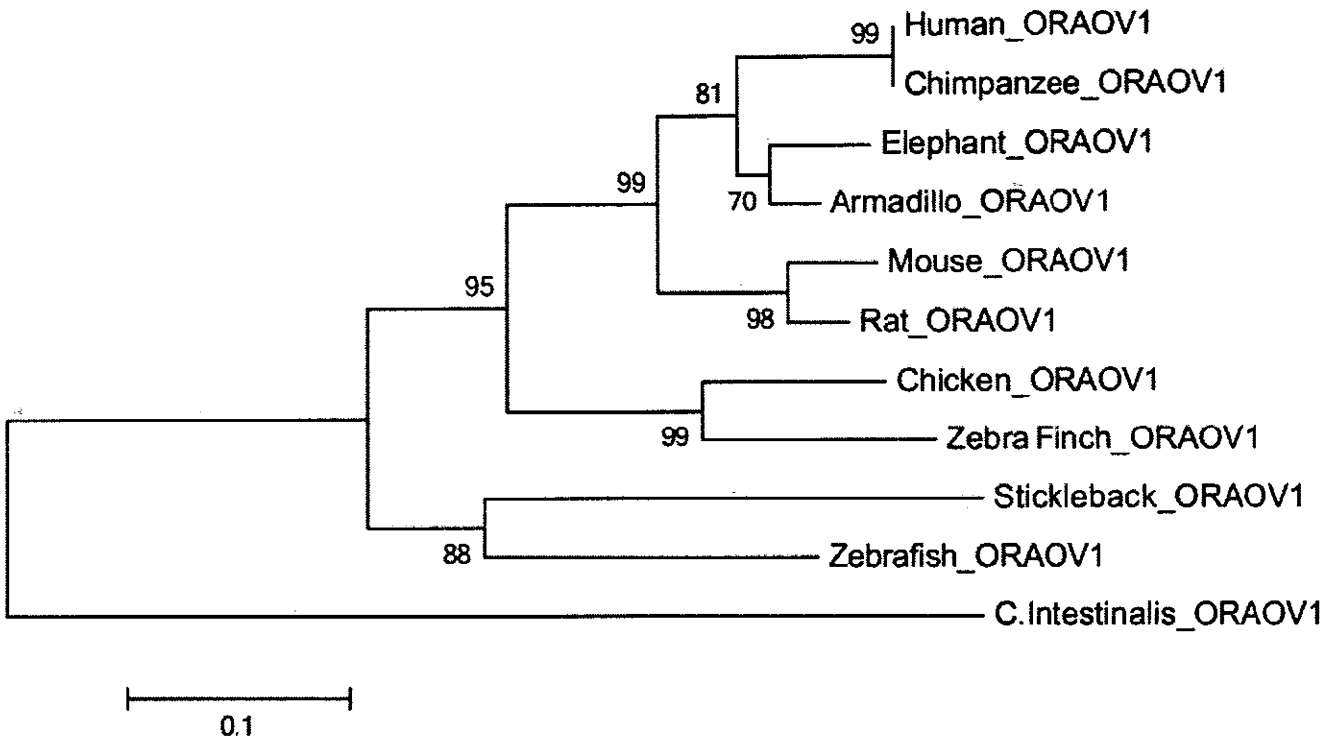


Figure 3.20: MEGA 5 evolutionary tool generated neighbour-Joining tree of *ORAOV1* gene. The Ensembl BLASTp/Blat was used to find out the paralogs of *ORAOV1* gene but significant paralogs were not reported. Amino acid sequences of *ORAOV1* orthologs were retrieved for phylogenetic tree construction. Parameters were same as mentioned in figure 17. Ciona intestinalis is an outer group in this tree representing as a origin of genes in other species. Ciona intestinalis gave rise to the gene in teleosts, birds, rodents, and in mammals successively after mutations. Teleosts are in separate cluster and other species are also in their appropriate place. A tree showing the >65 bootstrap values of all the species.

3.3. Mutational Analysis

Mutational analysis was employed on candidate proteins mutations to determine structural variations and reveal the functions of mutations reported in SCCHN. One missense mutation was found in *PTEN* protein and three in *ING1* protein results in SCCHN mentioned in table 3.7. Mutational analysis of proteins was performed to find out structural and functional changes by HOPE server.

Table 3.7: Missense mutations reported in candidate proteins with amino acid substitution.

Protein Name	Mutations	Mutation Type
PTEN	Alanine121-to-Glycine	Missense
ING1	Alanine192-to-Aspartic acid	Missense
	Cysteine215-to-Serine	Missense
	Asparagine216-to-Serine	Missense

3.3.1. *PTEN*

Ala121-to-Gly mutation in *PTEN* protein involved in SCCHN. 3D-structure of target *PTEN* protein was unknown. However, HOPE built structure of wild and mutated protein based on homologous structure. Alanine of the wild type protein is mutated into glycine at 121 position. Structure of wild type and mutated protein showing same backbone shown in red colour while different side chain shown in black colour is shown in figure 3.21.

Original and newly introduced mutant residue differ from each other by hydrophobicity values, charge and its specific size. In protein complex, hydrophobic interactions will be lost after mutation. The wild type is larger than mutant residue. The mutation is located within a domain, annotated as: "Phosphatase tensin-type". Mutation introduced glycine residue having different properties which alternately disturbed the domain and rigidity of protein at this position. 3D Structure of wild type and mutated protein is shown in figure 3.22.

Figure 3.21: Backbone and side chains of wild type (shown on left) and mutated protein (shown on Right) (Ala121-Gly). Backbone and side chains are coloured red and black respectively.

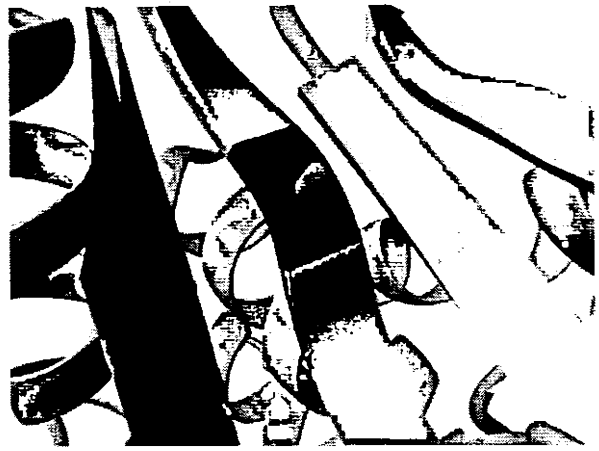


Figure 3.22: *PTEN* Protein is shown in grey colour and wild type and mutated (Ala121-Gly) side chains are shown and coloured green and red respectively.

3.3.2 *ING1*

Three missense mutations in *ING1* protein were reported involved in SCCHN. Ala192-to-Asp mutation in *ING1* protein results results in SCCHN. There is no protein structural information, modelling template and resolved 3D structure of this protein. Backbone and side chains of wild type and mutated protein are coloured in red and black respectively in figure 3.23.

Figure 3.23: Backbone and side chain of wild type is shown in left while mutant (Ala192-Asp) backbone and side chain is shown at right, coloured with red and black respectively.

Mutant residue is negatively charged, bigger and less hydrophobic than wild type residue. Mutant converted the original residue alanine into a aspartic acid that does not prefer α-helices as secondary structures. This residue is part of domain named zinger finger, PHD-type and mutant

residue might disturb the core structure of the domain. Mutant residue introduces a charge in a buried residue which can lead to the folding problems and mutation also cause loss of hydrophobic interactions in the core of the protein complex. Cysteine215-to-Serine mutation in *ING1* protein involved in SCCHN. Schematic structure of the original (left) and mutant protein (right) is shown in figure 3.24.

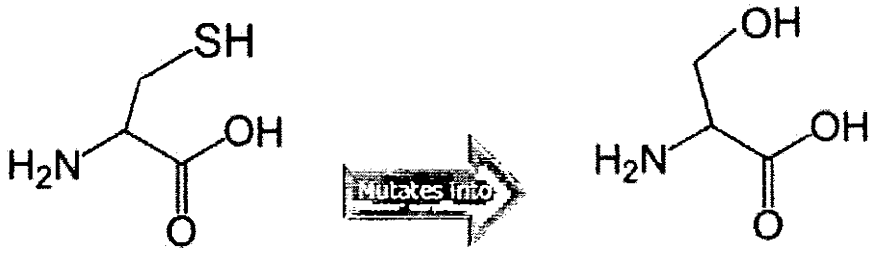


Figure 3.24: Schematic structure showing the wild type residue on the left and the mutated (Cys215-Ser) residue at right. Backbone and side chains are coloured red and black respectively.

Wild type residue is more hydrophobic than mutant residue. Zinc-finger domain are known to bind DNA which are disturbed by this mutation. Mutated residue is located on the surface of a domain and this mutation disturbs the interaction with other molecules. Close up of the mutated and the wild type protein are shown in figure 3.25.

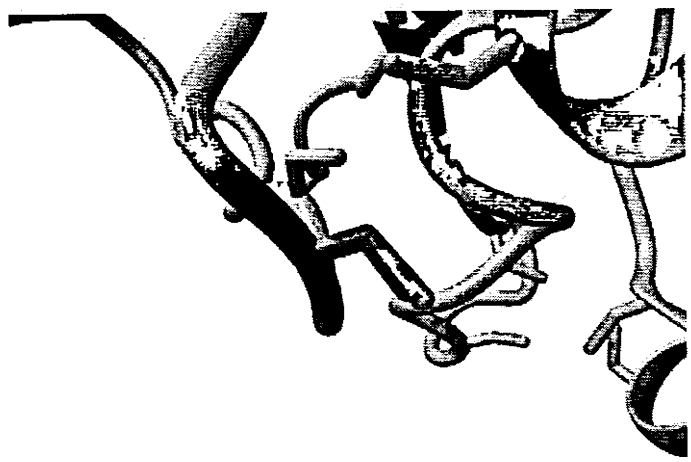


Figure 3.25: *ING1* Protein is shown in grey colour. The side chain of wild type and mutated (Cys215-Ser) residue is shown in green and red colour respectively.

Asparagine-to-Serine mutation at position 216 is reported that is involved in SCCHN. Wild type residue is larger than mutant residue differs from each other by charge and hydrophobicity values. The size difference between mutant and wild type residues disturbed the interaction with metal ions. Schematic structure of wild type and mutated residue is shown in figure 3.26.

Mutated serine residue is located on the surface of a domain and this mutation disturbed the interaction with other molecules. Mutated and wild type residue is shown in close up figure 3.27.

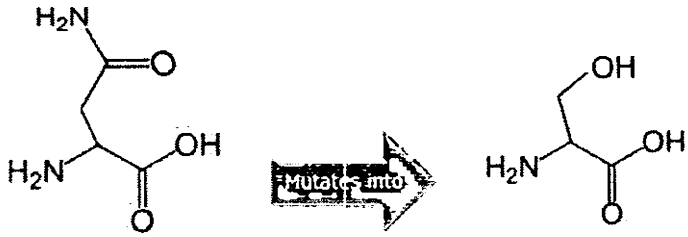


Figure 3.26: Schematic structure showing the wild type residue on the left and the mutated (Asn216-Ser) residue on the right. Backbone and side chains are coloured red and black respectively.

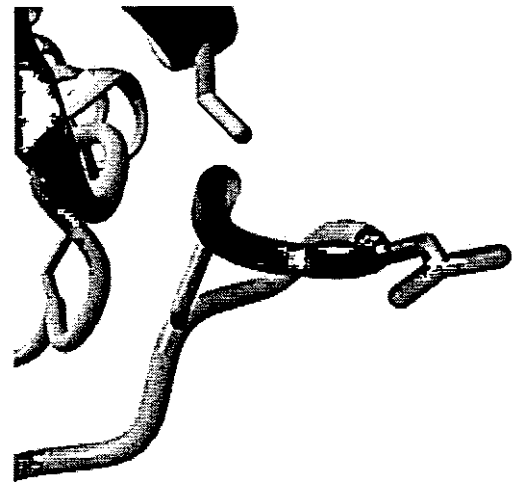


Figure 3.27: *ING1* protein is shown in grey colour and green and red colour representing the side chain of wild type and mutated residue (Asn216-Ser) respectively.

3.4. Molecular Docking

Molecular docking studies were performed to explore the interactions of candidate genes by protein-ligand docking and protein-protein docking.

3.4.1. Protein-Ligand Docking

Docking program Autodock Vina was employed to determine the binding sites of Ligand with receptor protein and the interactions of residues. Top two prioritized Ligands for *PTEN* gene were retrieved from KEGG Ligand database mentioned in table 3.8. Autodock Vina generated docked complexes separately of both ligands with receptor protein. Docked complex of Prostaglandin F2alpha Ligand with *PTEN* receptor protein having lowest binding energy was selected for further interaction analysis.

LigPlot and VMD were applied to identify interacting residues and type of interactions present in protein ligand docked complex. Docked complex analysed by LigPlot and VMD are shown in figure 3.28 and 3.29 respectively. Table 3.9 shows ionic, hydrogen and hydrophobic interactions of protein-ligand docked complex.

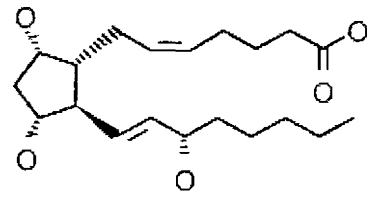
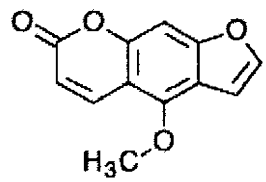
3.4.2. Protein-Protein Docking

Protein-protein docking of *TNFRSF10B*, *ING1* and *ORAOV1* proteins was performed by GRAMM-X and HEX servers. Ligands of these proteins were not reported in databases/ literature therefore interactor proteins were determined by the STRING database for protein-protein docking. Interactions of protein-protein docked complexes residues were determined by visualizing tool PyMOL.

3.4.1.1. *PTEN*

Ligands of *PTEN* were retrieved from KEGG Ligand database. A docked complex of protein-ligand was generated by an Autodock Vina tool to determine how ligand binds with receptor protein and their interactions. Complex having lowest binding energy was selected for analysis. Docked complex was visualized and analysed by LigPlot and VMD tools to determine the interacting residues in complex. Top two ligands of *PTEN* protein are shown in table 3.8. Interacting residues showing Ionic, hydrogen and hydrophobic interactions are mentioned in table 3.9.

Table 3.8: Top two prioritized ligands of *PTEN* protein were retrieved from KEGG Ligand database. Names, formulas and structures of Ligands are mentioned.

Receptor Protein	Ligand Database	Accession Number	Ligand Name	Formula	Structure
PTEN	KEGG	C00639	Prostaglandin F2alpha; (5Z,13E)-(15S)-9alpha,11alpha,15-Trihydroxyprost a-5,13-dienoate;	C20H34 O5	
	KEGG	C01557	Berga*PTEN*; 5-Methoxypsorale n; O-Methylbergaptol ; 5-Methoxyfuranoc oumarin	C12H8O 4	 C01557

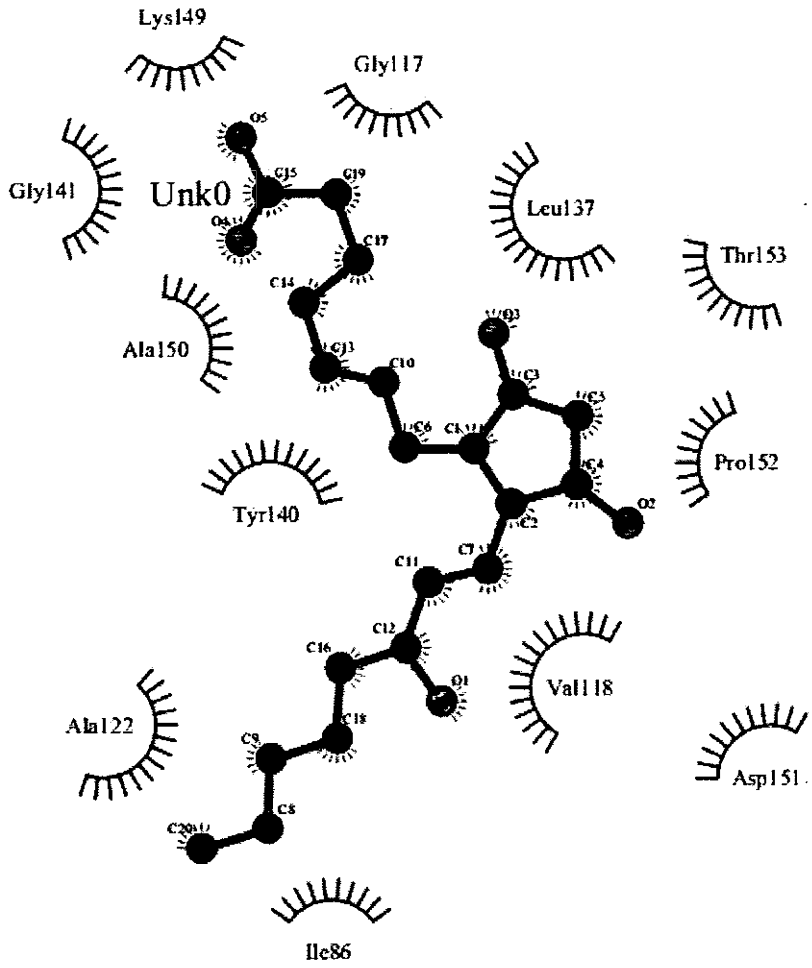


Figure 3.28: LigPlot tool was employed to determine the interactions between receptor and ligand protein. Residues of receptor protein are shown that interacted with ligand. No ionic interactions were identified in these interactions.

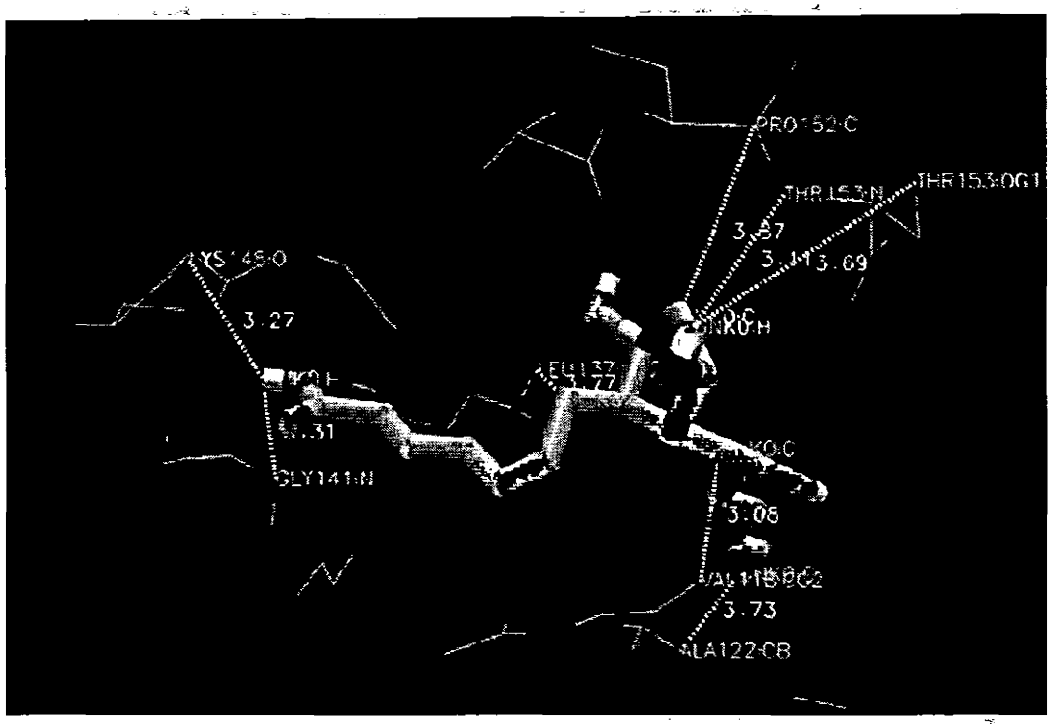


Figure 3.29: Prostaglandin F2alpha Ligand was used in protein-ligand docking of *PTEN* protein by Autodock Vina. VMD visualizing tool was applied to determine and analyse the active sites and residue interactions. The figure showing interactions of residues with less than 4 bond distance.

Table 3.9: Autodock Vina generated docked complex of protein and Ligand. Interactions of residues were determined by the VMD visualizing tool. Hydrophobic and hydrogen bonding in residues is shown in the table. Ionic bonding is not present in these interacting residues.

Receptor Protein	**Interactions Type**	**Residues Bonding Receptor ⟶ ligand**	**Bond Distance**
PTEN	Hydrogen Bonding (H-O) (H-N)	LYS148:O UNK0:H	3.27
		THR153:OG UNK0:H	3.69
		THR153:N UNK0:H	3.11
		GLY141:N UNK0:H	3.31
	Hydrophobic interactions (C-C)	ALA122:CB UNK0:C	3.73
		LEU137:CD2 UNK0:C	3.77
		VAL118:CG2 UNK0:C	3.01
		PRO152:C UNK0:C	3.88
	Ionic Bonding (N-O)	No ionic bonds	

3.4.2.1. *TNFRSF10B*

A functional partner network of *TNFRSF10B* protein was generated by the STRING database to explore the highly interacting proteins of the target protein. *TNFSF10* protein having highest interaction score 0.999 with receptor protein was used as a ligand protein in protein-protein docking by GRAMM-X. Interaction network and protein-protein docked complex is shown in figure 3.30 and 3.31 respectively. Interactions of interacting residues were determined and analysed by PyMOL, mentioned in table 3.10.

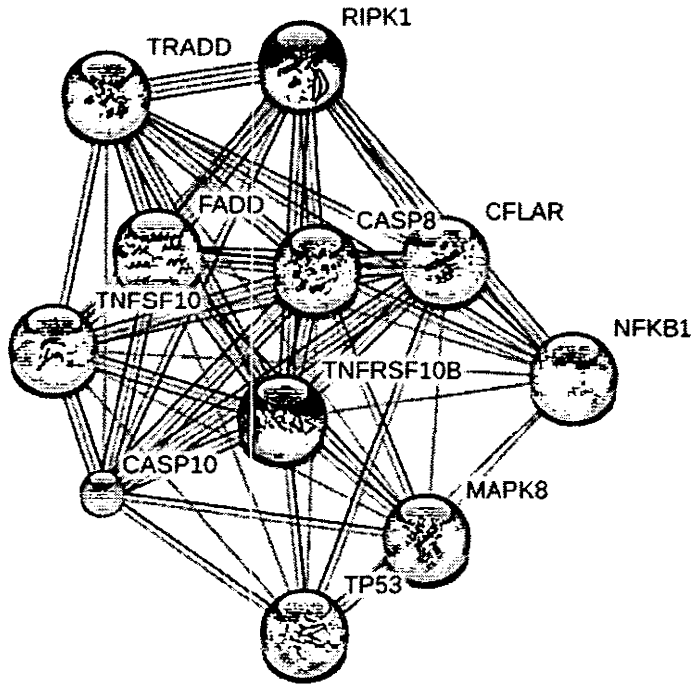


Figure 3.30: Interaction network of *TNFRSF10B* was generated by STRING database. In this network *TNFSF10B* protein showed the highest interaction score 0.999 with *TNFSF10* protein. *TNFSF10* protein is used in protein-protein docking of *TNFRSF10B*.

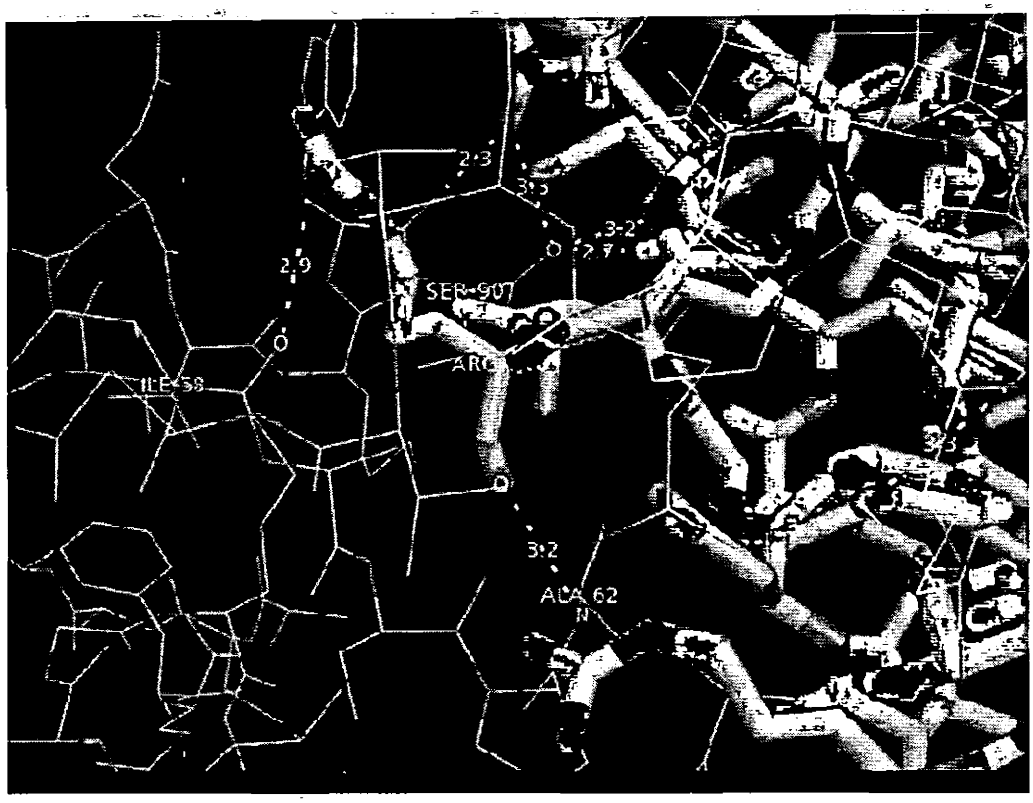


Figure 3.31: Interaction analysis of *TNFRSF10B* and *TNFSF10* visualized by PyMOL tool

Table 3.10: Interacting residues showed only ionic bonding between *TNFRSF10B* and *TNFSF10* protein.

Receptor Protein	**Interacting Protein**	**Interactions Type**	**Interactions (Receptor residue → Interacting protein residue**	**Bond Distance**
TNFRSF10B	*TNFSF10*	Ionic Bonding (N-O)	ILE-58/O → ARG-130/NH1	2.9
			SER-90/O → SER-156/N	3.2
			ALA-62/N → ARG-130/O	3.2

3.4.2.2. *ING1*

Ligand of *ING1* protein was not reported in literature and databases so that protein-protein docking approach was employed for *ING1* protein. STRING database was used to determine interactor proteins of *ING1* protein and *TP53* protein showed highest interaction score 0.998. The structure of *TP53* protein was built by MODELLER and used in *ING1-TP53* docked complex. Interaction network is displayed in figure 3.32 and docked complex visualized by PyMOL is shown in figure 3.33. Interacting residues of protein-protein docked complex was analysed by PyMOL shown in table 3.11.

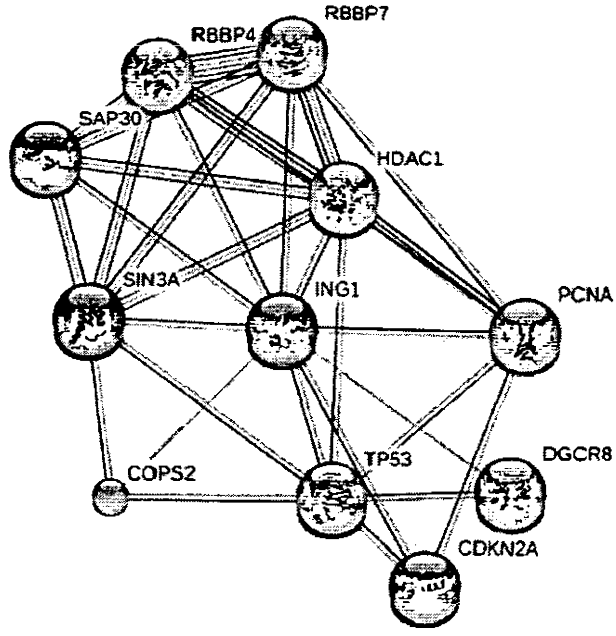


Figure 3.32: Functional partners of *ING1* protein through STRING database. Interaction network showed highly interacting proteins. *TP53* having highest score 0.998 was selected as a ligand protein in protein-protein docking of *ING1* protein.

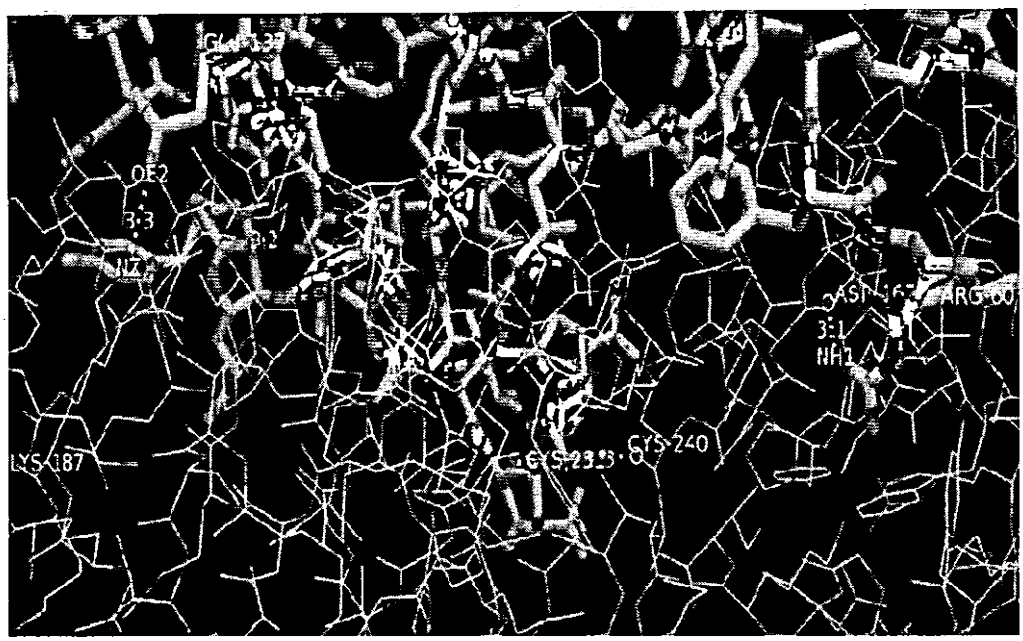


Figure 3.33: *ING1* receptor protein shows interactions with ligand protein *TP53*.

Table 3.11: *ING1-TP53* docked complex showed ionic interactions between interacting residues.

Receptor Protein	**Interacting Protein**	**Interactions Type**	**Interactions (Receptor residue → Interacting protein residue**	**Bond Distance**
ING1	*TP53*	Ionic Bonding (N-O)	CYS-240/O→ CYS-231/N	3.3
			ASP-167/O→ ARG-60/NH1	3.1
			LYS-187/NZ→ GLU137/OE2	3.3

3.4.2.3. *ORAOV1*

STRING database was used to find out the functional partners of *ORAOV1* protein for protein-protein docking. *FGF3* protein showed highest similarity score 0.917 among other protein in the interaction network shown in figure 3.34. *FGF3* protein utilized as a ligand protein to dock with *ORAOV1* receptor protein by GRAMM-X. Interactions between docked complex were analysed by PyMOL. Docked complex is presented in figure 3.35 and interactions are mentioned in table 3.12.

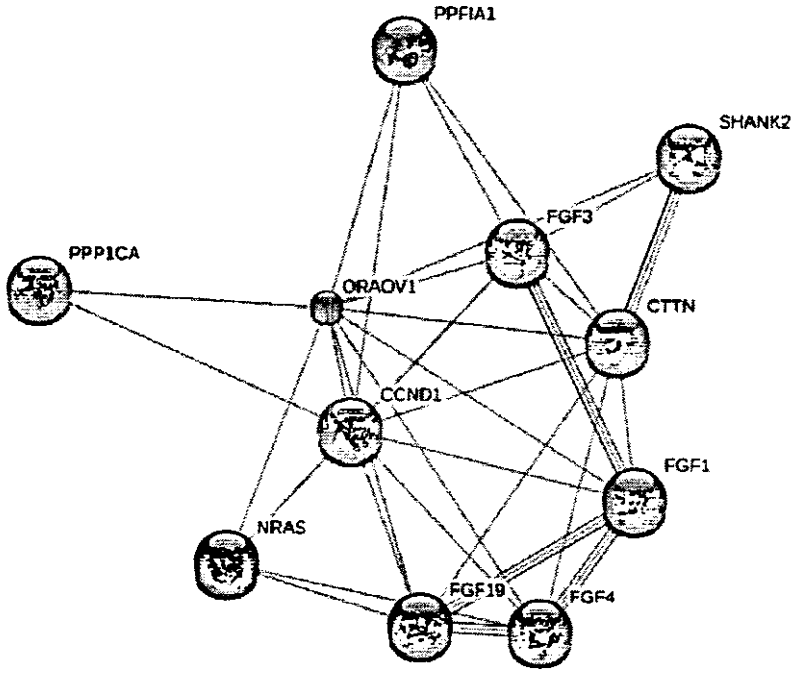


Figure 3.34: STRING database generated network of closely related proteins of *ORAOV1* protein. The network showed the highest interaction score 0.917 with *FGF3* protein. *FGF3* protein structure was retrieved from the Swiss model database as a ligand protein for docking with *ORAOV1* receptor protein.

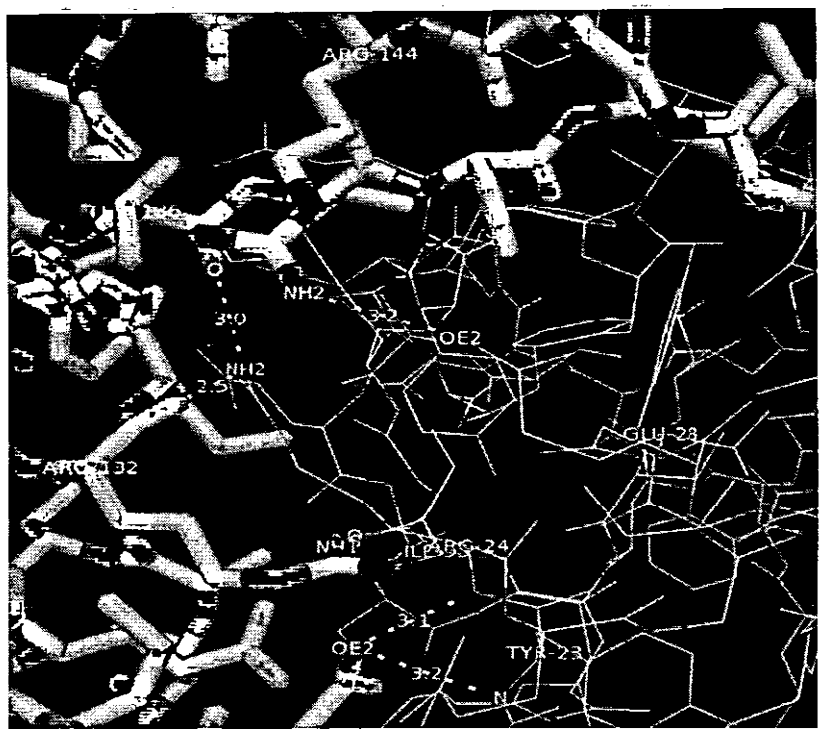


Figure 3.35: Docked complex of *ORAOV1* and *FGF3* protein visualized by PyMOL tool.

Table 3.12: Interaction results of *ORAOV1* receptor with *FGF3* ligand protein.

Receptor Protein	**Interacting Protein**	**Interactions**	**Interactions (Receptor residue → Interacting protein residue**	**Bond Distance**
ORAOV1	*FGF3*	Ionic Bonding (N-O)	GLU-28/OE2→ ARG-144/NH2	3.2
			ARG-24/N→GLU-82/OE2	3.1
			TYR-23/N→ GLU-82/OE2	3.2
			ILE-53/O → ARG-132/NH1	2.8
			ARG-24/NH2 → ARG-132/O	2.5
			ARG-24/NH2→ THR-136/O	3.0

CHAPTER 04

DISCUSSION

Discussion

Head and Neck Cancer (HNC) is a disease linked with major mortality and morbidity (Jemal *et al.,* 2011). Potential environmental agents and the multigenic nature of HNC made complex epidemiology of disease (Vokes *et al.,* 1993). Worldwide, HNC is the sixth most occurring cancers affecting the oral cavity, larynx and pharynx (Devasena *et al.,* 2007) and second most prevalent cancer in Pakistani population (Hanif *et al.,* 2009).

Literature survey and database searching approaches were applied in the screening of SCCHN candidate genes. Four genes were shortlisted for further analysis that are directly involved in SCCHN. In the present work, computational analysis was performed on four prioritized candidate genes i.e. *TNFRSF10B*, *ING1*, *PTEN* and *ORAOV1*.

In the present work, 3D structures of *TNFRSF10B*, *ING1*, *PTEN* and *ORAOV1* proteins were not resolved by X-ray crystallography and NMR techniques and also not available in PDB. The comparative modelling approach was used to predict 3D structures of proteins by applying structure prediction tool i.e. MODELLER. The most authentic templates of *TNFRSF10B*, *ING1*, *PTEN* and *ORAOV1* proteins were retrieved based on extensive literature survey and E-value for reliable structure prediction.

For *TNFRSF10B* gene, three templates namely 2ZB9, 3NKE and 3NKD templates were retrieved having 3d structures in PDB. Among these templates, 2ZB9 having high alignment score and query coverage showed better evaluation results with respect to above mentioned templates. RAMPAGE showed the 93.2% residues in favoured region and 5.7% in allowed region whereas only 1 residue was in outlier region. 51.852% quality factor and -1.08 Z-score were shown by ERRAT and ProSA evaluation tools.

For *ING1* gene, five suitable templates namely 4AFL, 1WES, 2G6Q, 2QIC and 2VNF were retrieved by BLASTp. 4AFL template having 34% query coverage and 36% identity against target sequence showed good results with respect to other templates. 92.8% residues in favoured, 5.4% in allowed and 1.8% residues were in outlier regions determined by RAMPAGE tool. ERRAT showed 48.485% quality factor and -2.44 z-score of predicted model by ProSA.

The *PTEN* protein sequence was used in BLASTp to find out appropriate template and top five suitable templates (1D5R, 3AWF, 3AWG, 3AWE, 3V0D) against the target protein were retrieved.

1DFR having 81% query coverage and 89% identity was employed for homology modelling of the target protein. Residues of predicted model were 87.9% in favoured, 9.8% in allowed and 2.3% in outlier regions. Model quality as 50% quality factor and -4.24 z-score were determined by ERRAT and ProSA evaluation tools respectively. The *ORAOV1* protein sequence was applied to determine suitable templates by BLASTp for comparative modelling. Homology modelling program i.e. MODELLER showed bad results by using template having less than 30% identity. I-Tasser online tool based on the threading approach was utilized for structure prediction of *ORAOV1* protein. Rampage showed 91.9% residues in favoured and 6.7% residues in allowed region. Only 2 residues were lying in outlier region. Predicted model has 72.868% quality factor and -3.25 z-score assessed by ERRAT and ProSA respectively.

Evolutionary tool MEGA 5 was employed for neighbour-joining trees of selected genes to determine the evolutionary relationship of genes among teleosts, rodents, birds, primates and mammals. In literature, *TNFRSF10B* gene and its paralogs genes are predicted in primates and human. *TNFRSF10D* and *TNFRSF10A* are the paralogs of *TNFRSF10B* gene that are used in the construction of a phylogenetic tree. *TNFRSF10A* gene is an outgroup in *TNFRSF10B* tree that presents *TNFRSF10A* gene as an origin of other genes. *TNFRSF10A* gene gave rise to *TNFRSF10B* gene in ciona intestinalis and other cluster is further diverged into *TNFRSF10D* and *TNFRSF10B* genes. Human and Gorilla are closely related in *TNFRSF10B* and *TNFRSF10A* genes while in *TNFRSF10D*, Human is closely related with Chimpanzee. Bootstrap replication values >50 are presented in this tree representing the reliability of topology.

ING2 and *ING5* are paralogs of *ING1* gene that are used in phylogenetic tree to determine the ancestral and evolutionary history of genes. *ING5* is an outgroup in the tree and inferred as an ancestor of other genes in the tree. Reliability of tree topology was tested by the bootstrap value that is mostly >50. Topology of tree showing that *ING1* and *ING2* are diverged from *ING5* gene. Zebrafish is more diverged in teleosts. Human is closely related to chimpanzee having similar evolutionary history. Primates (Chimpanzee), mammals (Dog, Elephant, Horse, and Human)

rodents (Mouse and Rat), birds (Zebra Finch and Chicken) and Teleost (Stickleback, Medaka, Zebrafish and Fugu) are according to their place and concealing with the specie tree.

Significant paralogs of *PTEN* and *ORAOV1* genes were not identified by Blast and Ensembl. The thirteen orthologs of both *PTEN* and *ORAOV1* genes were retrieved for the construction of the neighbour joining tree. A tree showing that the *PTEN* gene in teleost, rodents, primates and mammals is evolved from ciona intesinalis. Opossum and zebrafish are more diverged in birds and teleosts respectively. All orthologs are concealing with the specie tree and >70 bootstrap value showing the reliability of topology.

Ciona Intestinalis is an outgroup in *an ORAOV1* orthologs tree. Human is closely related to chimpanzee having more conserved sequences. Teleosts, birds, rodents, primates and mammals are in their respective cluster and diverged from ciona intestinalis. Reliability of topology was tested by number of replication values that are □> 70. Mutation rate was 0.1 per site in the tree of *ORAOV1* gene.

Mutations of candidate proteins involved in SCCHN were identified. Mutational analysis was performed by the HOPE server to determine the structural changes in proteins. Ala-to-Gly missense mutation was reported in *PTEN* protein at 121 location. Mutated residue is smaller than wild type residue and mutation led to loss of hydrophobic interactions in a protein complex. Mutated residue is located within the phosphate tensin type domain. Mutation introduced Glycine residue having different properties that disturbed the domain and stability of protein.

Three missense mutations Ala192-to-Asp, Cys215-to-Ser and Asn216-to-Ser are reported in *ING1* protein that cause SCCHN. These mutated residues having their own charge, size and hydrophobic properties alter the integrity of original protein. These residues are located within Zinc-finger domain. Mutation in these residues disturbs the domain and interaction with other molecules.

Molecular docking studies were carried out on HNC candidate genes. Protein-Ligand docking was performed for *PTEN* gene using AutoDock software and protein-protein docking for *TNFRSF10B*, *ING1*, and *ORAOV1* genes using Gramm-X and PatchDock servers. Ligand for protein ligand docking was retrieved from KEGG ligand database.

Prostaglandin F2alpha and Berga*PTEN* 5-Methoxypsoralen ligands were retrieved to explore the binding interactions of *PTEN* protein. Docked complex with Prostaglandin F2alpha ligand having lowest binding energy was selected for post docking analysis. Total Eight interactions were observed between receptor protein and ligand by VMD. No ionic interactions were found in this post-dock analysis. Amino acids residues located within the vicinity of 4Å from the ligand were analysed. Four hydrogen bondings were observed between oxygen atoms of LYS148 and THR153 and nitrogen atoms of THR153 and GLY141 of receptor protein with hydrogen atoms of ligand having 3.27Å, 3.69Å, 3.11Å, and 3.31Å bond distances respectively. Four hydrophobic interactions were observed between carbon atoms of receptor protein and ligand. The carbon atom of ALAI22 showed hydrophobic interactions with the carbon atom of ligand with 3.73 Å bond distance. The carbon atom of LEU137 interacted with the carbon atom of ligand with 3.77Å bond distance. PRO152 and VAL118 of receptor protein also showed hydrophobic interactions with carbon atoms of ligand compound with a bond distance 3.88Å and 3.01Å respectively.

The STRING online database was utilized for finding the interacting partners of *TNFRSF10B*, *ING1* and *ORAOV1* proteins. GRAMM-X and PatchDock softwares were utilized for *TNFRSF10B*, *ING1* and *ORAOV1* protein-protein docking. *TNFSF10* protein showed highest interacting score 0.999 with *TNFRSF10B* target protein belonging to the same family. *TNFSF10* protein was retrieved as a ligand for protein-protein docking with *TNFRSF10B* receptor protein. Post docking analysis was performed by PyMOL and hydrogen, ionic and hydrophobic interactions were analysed. Only ionic interactions were found in docked complex. Isoleucine-58 of receptor protein *TNFRSF10B* showed ionic interactions with Arginine-130 of ligand protein *TNFSF10* with the distance of 2.9 Å. The nitrogen atom of arginine showed interaction with oxygen atom of isoleucine. Serine-90 of *TNFRSF10B* receptor protein showed ionic interactions with Serine-156 of ligand protein with a bond distance of 3.2 Å. Nitrogen of serine of ligand protein *TNFSF10* interacted with the oxygen of serine of receptor protein. Alanine-62 nitrogen of

receptor protein *TNFRSF10B* interacted with arginine-130 oxygen of ligand protein *TNFSF10* with 3.2Å bond distance.

For *ING1* protein-protein docking, *TP53* having the highest score 0.998 was selected as a ligand protein. Three ionic interactions were determined by PyMOL in psot dock analysis. Oxygen of Cysteine-240 of receptor protein *ING1* interacted with nitrogen of cysteine-231 of ligand *TP53* protein with a 3.3Å bond distance. Aspartic Acid-167 of receptor *ING1* protein showed ionic interactions with arginine-60 of ligand protein *TP53* with a 3.1Å bond distance. An Oxygen atom of receptor protein interacted with the nitrogen atom of ligand protein. Lysine-187 of receptor protein showed ionic interactions with Glutamic Acid-137 of ligand protein with a 3.3Å bond distance.

For *ORAOV1* protein, *FGF3* (Fibroblast Growth factor 3) protein having the highest interacting score was selected for protein-protein docking with *ORAOV1* protein. Post docking analysis by PyMOL showed six ionic interactions between docked complex. An Oxygen atom of Glutamic Acid-28 and Isoleucine-53 of receptor protein *ORAOV1* showed ionic interactions with nitrogen of arginie-144 and arginine-132 of ligand protein FGF3 with a 3.2Å and 2.8Å bond distances respectively. A nitrogen atom of Tyrosine-23 of receptor protein *ORAOV1* interacted with oxygen of glutamic acid-82 of ligand protein *FGF3* with a 3.2Å bond distance. Nitrogen of Arginine-24 showed interactions with oxygen of three residues i.e. glutamic acid-82 , arginine-132 and threonine-136 of ligand protein FGF3 with a bond distance 3.1Å, 2.5Å and 3.0Å.

Conclusion and Future Prospects

The present work focused on computational analysis of Head and Neck Cancer (HNC) candidate genes. Highly subjective candidate approach was applied to identify the most prioritized genes in pathogenesis of HNC and four plausible genes (*TNFRSF10B*, *PTEN*, *ING1* and *ORAOV1*) were shortlisted based on their direct involvement in the disease. Three dimensional (3D) structures of candidate genes, mutational analysis, structure based receptor-ligand and protein-protein interactions are may be used in HNC treatment. Interactions of candidate genes were identified that need to be tested for further analysis.

Prostaglandin F2alpha is involved in metabolic and neuroactive receptor-ligand interaction pathway. Mutated *PTEN* gene negatively regulate the signalling pathway while Prostaglandin F2alpha may act as an activator that inhibits negatively Regulation of AKT/PKB signalling pathway. Prostaglandin F2alpha ligand could be used *in-vivo* experiment in mouse to test the effect and interactions of ligand. Analysed ligand may be used in drug designing of HNC. Humans and primates are sharing >70% sequence similarities that may be used in the prediction of novel gene families and their functions and ancestors. Our research suggested a baseline for functional insights into the structure and development of drugs to cure the HNC.

In future prospects, there is an urgent need to develop a biological database on HNC contain all the information associated with disease including pathogenesis, genetics, interacting proteins and drugs. There should be tools and applications for computational analysis of HNC candidate genes, pathways, interactions with proteins and ligands.There is an also need to analyse the reported ligands by *in-vitro* and *in-vivo* experiments for drug development of HNC in less time.

CHAPTER 05

REFERENCES

References

1. Altschul et al., (1990) Basic local alignment search tool, J. Mol. Biol., 215, p. 403-410.
2. Ang K.K., Haris J., Wheeler R., And Weber R., (2010) Human papilomavirus and survival of patients with oropharyngeal cancer, N. ENG. J. MED., 363, p. 24-35.
3. Argiris A., Brockstein B.E., And Haraf D.J., (2004) Competing causes of death and second primary tumors in patients with locoregionally advanced head and neck cancer treated with chemoradiotherapy, Clin. Cancer. Res., 10, p.1956–1962.
4. Argiris A., And Eng C., (2003) Epidemiology, staging, and screening of head and neck cancer. Cancer Treat Res., 114, p. 15–60.
5. Argiris A., Karamouzis M.V., Raben D., And Ferris R.L., (2008) Head and neck cancer seminar, Lancet, 371, p. 1695–1709.
6. Balz V., Scheckenbach K., Gotte K., Bockmuhl U., Petersen I., And Bier H., (2003) Is the p53 inactivation frequency in squamous cell carcinomas of the head and neck underestimated? Analysis of p53 exons 2-11 and human papillomavirus 16/18 E6 transcripts in 123 unselected tumor specimens, Cancer Res., 63, p. 1188–1191.
7. Baras A., Yu Y., Filtz M., Kim B., And Moskaluk C.A., (2009) Combined genomic and gene expression microarray profiling identifies ECOP as an upregulated gene in squamous cell carcinomas independent of DNA amplification, Oncogene, 28, p. 2919-2924.
8. Barnes L., Eveson J.W., And Reichart P., (2005) World Health Organization classification of tumours, Pathology and genetics of head and neck tumours, Lyon, France, IARC Press.
9. Birle D., Bottini N., Williams S., Huynh H., deBelle I., Adamson E., and Mustelin T., (2002) Negative feedback regulation of the tumor suppressor PTEN by phosphoinositide-induced serine phosphorylation, J. Immunol., 169, p. 286–291.
10. Blot W.J., McLaughlin J.K., Winn D.M. et al., (1988) Smoking and drinking in relation to oral and pharyngeal cancer, Cancer Res., 48, p. 3282–3287.
11. Bonneau D., And Longy M., (2000) Mutations of the human PTEN gene, Hum. Mutat., 16, p.109 –122.

12. Branstetter B.F., Blodgett T.M., And Zimmer L.A., (2005) Head and neck malignancy: is PET/CT more accurate than PET or CT alone?, Radiology, 235, p. 580–586.
13. Campos E.I., Martinka M., Mitchell D.L., Dai D.L., And Li G., (2004) Mutations of the ING1 tumour suppressor gene detected in human melanoma abrogate nucleotide excision repair. Int. J. Oncol., 25, p. 73–80.
14. Capaccio P., Pruneri G., And Carboni N., (2000) Cyclin D1 expression is predictive of occult metastases in head and neck cancer patients with clinically negative cervical lymph nodes, Head Neck, 22, p. 234–240.
15. Chen H., Gou Y., Zhang K., And Fang B., (2011) Homology modeling and molecular docking xylitol dehydrogenase from Aspergillus Oryzae, Wei Sheng Wu Xue Bao, 51, p. 948-955.
16. Cheung K.J., Mitchell D., Lin P., And Li G., (2001) The tumor suppressor candidate p33 (ING1) mediates repair of UV-damaged DNA. Cancer Res., 61, p. 4974–4977.
17. Choi Y.W., Bae S.M., Kim Y.W., Lee H.N., Kim Y.W., Park T.C., Ro D.Y., Shin J.C., Shin S.J., Seo J.S., Ahn W.S., (2007) Gene expression profiles in squamous cell cervical carcinoma using array-based comparative genomic hybridization analysis, Int. J. Gynecol. Cancer, 17, p. 687-696.
18. Chou K.C., (2004). Structural Bioinformatics and its Impact to Biomedical Science, Current Medicinal Chemistry, 11, p. 2105-2134.
19. D'Souza G., Kreimer A.R., And Viscidi R., (2007) Case-control study of human papillomavirus and oropharyngeal cancer, N. Engl. J. Med., 356, p. 1944–1956.
20. Devasena A., Pranay M.C., Sadhana K., Rajani A.B., And Manoj B.M., (2007) Susceptibility to oral cancer by genetic polymorphisms at CYP1A1, GSTM1 and GSTT1 loci among Indians: Tobacco exposure as a risk modulator, Carcinogenesis, 28, p.1455–1462.
21. Doyon Y., Selleck W., Lane W.S., Tan S., And Cote J., (2004) Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast to humans, Mol. Cell Biol., 24, p. 1884–1896.
22. Eswar N., Eramian D., Webb B., Shen MY., And Sali A., (2008) Protein Structure modeling with MODELLER, Methods of Molecular Biology, 426, p. 145-159

23. Faheem A., (2007) Cancer registration in Pakistan: contemporary state of affairs, Asia. Pac. J. Cancer Prev., 8, p. 452-456.
24. Ferris R.L., Hunt J.L., And Ferrone S., (2005) Human leukocyte antigen (HLA) class I defects in head and neck cancer: molecular mechanisms and clinical significance, Immunol Res., 33, p. 113–133.
25. Foulkes W.D., Brunet J.S., Sieh W., Black M.J., Shenouda G., And Narod S.A., (1996) Familial risks of squamous cell carcinoma of the head and neck: retrospective case-control study, B.M.J., 313, p. 716–721.
26. French C.A., Kutok J.L., And Faquin W.C., (2004) Midline carcinoma of children and young adults with NUT rearrangement, J. CLIN. Oncol., 22, p. 4135-4139.
27. Fuchs R., (2002) From sequence to biology: the impact on Bioinformatics, Bioinformatics, 4, p. 505-506.
28. Garkavtsev I., Kazarov A., Gudkov A., And Riabowol K., (1996) Suppression of the novel growth inhibitor p33ING1 promotes neoplastic transformation, Nat. Genet., 14, p. 415–420.
29. Garkavtsev I., And Riabowol K., (1997) Extension of the replicative life span of human diploid fibroblasts by inhibition of the p33ING1 candidate tumor suppressor, Mol. Cell Biol., 17, p. 2014–2019.
30. Garnis C., Coe B.P., Ishkanian A., Zhang L., Rosin M.P., And Lam W.L., (2004) Novel regions of amplification on 8q distinct from the MYC locus and frequently altered in oral dysplasia and cancer, Genes Chromosom Cancer, 39, p. 93–98.
31. Gunduz M., Ouchida M., Fukushima K., Hanafusa H., Etani T., Nishioka S., Nishizaki K., And Shimizu K., (2000) Genomic structure of the human ING1 gene and tumor-specific mutations detected in head and neck squamous cell carcinomas, Cancer Res., 60, p. 3143-3146
32. Ha P.K., And Califano J.A., (2006) Promoter methylation and inactivation of tumour-suppressor genes in oral squamous-cell carcinoma, Lancet Oncol., 7, p. 77–82.
33. Hanif M., Zaidi P., Kamal S., And Hameed A., (2009) Institution-based cancer incidence in a local population in Pakistan: nine year data analysis, Asia. Pac. J. Cancer Prev., 10, p. 227-230.

34. Hashibe M., Boff etta P., And Zaridze D., (2006) Evidence for an important role of alcohol- and aldehyde-metabolizing genes in cancers of the upper aerodigestive tract, Cancer Epidemiol Biomarkers Prev., 15, p. 696–703.
35. He G.H., Helbing C.C., Wagner M.J., Sensen C.W., And Riabowol K., (2005) Phylogenetic analysis of the ING family of PHD finger proteins, Mol. Biol. Evol., 22, p. 104–116.
36. Hobbs C.G., Sterne J.A., Bailey M., Heyderman R.S., Birchall M.A., Thomas S.J., (2006) Human papillomavirus and head and neck cancer: a systematic review and meta-analysis, Clin. Otolaryngol, 31, p. 259–266.
37. Huang X., Gollin S.M., Raja S., And Godfrey T.E., (2002) High-resolution mapping of the 11q13 amplicon and identification of a gene, TAOS1, that is amplified and overexpressed in oral cancer cells, Proc. Natl. Acad. Sci. U.S.A., 99, p. 11369–11374.
38. Irwin J.J., And Shoichet K.B., (2005) ZINC-A Free Database of Commercially Available Compounds for Virtual Screening, J. Chem. Inf. Model., 45, p. 177–182.
39. Jarvinen A.K., Autio R., And Kilpinen S., (2008) High-resolution copy number and gene expression microarray analyses of head and neck squamous cell carcinoma cell lines of tongue and larynx, Genes Chromosom Cancer, 47, p. 500–509.
40. Jemal A., Bray F., Center M., Ferlay J., Ward E., And Forman D., (2011) Global cancer statistics, CA Cancer J. Clin., 61, p. 69–90.
41. Jiang L., Zeng X., Yang H., Wang Z., Shen J., Bai J., Zhang Y., Gao F., Zhou M., And Chen Q., (2008) Oral cancer overexpressed 1 (ORAOV1): a regulator for the cell growth and tumor angiogenesis in oral squamous cell carcinoma, Int. J. Cancer, 123, p. 1779-1786.
42. Karamouzis M.V., Grandis J.R., And Argiris A., (2007) Therapies directed against epidermal growth factor receptor in aerodigestive carcinomas, JAMA, 298, p. 70–82.
43. Khuri F.R., Lee J.J., And Lippman S.M., (2006) Randomized phase III trial of low-dose isotretinoin for prevention of second primary tumors in stage I and II head and neck cancer patients, J. Natl. Cancer Inst., 98, p. 441–450.
44. Kuss I., Hathaway B., Ferris R.L., Gooding W., And Whiteside T.L., (2004) Decreased absolute counts of T lymphocyte subsets and their relation to disease in squamous cell carcinoma of the head and neck, Clin. Cancer Res., 10, p. 3755–3762.

45. Kuzmichev A., Zhang Y., Erdjument-Bromage H., Tempst P., And Reinberg D., (2002) Role of the Sin3-histone deacetylase complex in growth regulation by the candidate tumor suppressor p33(ING1), Mol. Cell Biol., 22, p. 835–848.
46. Li D.M., And Sun H., (1997) TEP1, encoded by a candidate tumor suppressor locus, is a novel protein tyrosine phosphatase regulated by transforming growth factor, Cancer Res., 57, p. 2124 –2129.
47. Li J., Yen C., Liaw D., Podsypanina K., Bose S., Wang S.I., Puc J., Miliaresis C., Rodgers L., McCombie R., Bigner S.H., Giovanella B.C., Ittmann M., Tycko B., Hibshoosh H., Wigler M.H., And Parsons R., (1997) PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer, Science, 28, p. 1943–1947.
48. Lopez-Albaitero A., Nayak J.V., And Ogino T., (2006) Role of antigen-processing machinery in the in vitro resistance of squamous cell carcinoma of the head and neck cells to recognition by CTL, J. Immunol., 176, p. 3402–3409.
49. Maehama T., And Dixon J.E., (1998) The Tumor Suppressor, PTEN/MMAC1, Dephosphorylates the Lipid Second Messenger, Phosphatidylinositol 3,4,5-Trisphosphate, The Journal of Biological Chemistry, 273, p. 13375-13378.
50. Mao L., Lee J.S., And Fan Y.H., (1996) Frequent microsatellite alterations at chromosomes 9p21 and 3p14 in oral premalignant lesions and their value in cancer risk assessment, Nat. Med., 2, p. 682–685.
51. Martin C.L., Reshmi S.C., And Ried T., (2008) Chromosomal imbalances in oral squamous cell carcinoma: examination of 31 cell lines and review of the literature, Oral Oncol., 44, p. 369–382.
52. Schabath M.B., Giuliano A.R., Thompson Z., Fenstermacher D., Jonathan K., Sellers TA., And Haura E., (2012) TNFRSF10B polymorphisms and haplotypes predicts survival in non-small cell lung cancer patients, Cancer Research, 72, p. 4506-4520.
53. McCaul J.A., Gordon K.E., Clark L.J., And Parkinson E.K., (2002) Telomerase inhibition and the future management of head-and-neck cancer, Lancet Oncol., 3, p. 280–288.
54. Mendelsohn L.D., (2004) ChemDraw 8 Ultra: Windows and Macintosh Versions, J. Chem. Inf. Comput. Sci., 44, p. 2225–2226

55. Miller S.J., Lou D.Y., Seldin D.C., Lane W.S., And Neel B.G., (2002) Direct identification of PTEN phosphorylation sites, F.E.B.S.Lett., 528, p. 145–153.
56. Neville B.W., And Day T.A., (2002) Oral cancer and precancerous lesions. CA Cancer J. Clin., 52, p. 195–215.
57. Pai S.I., Wu G.S., Ozoren N., Wu L., Jen J., Sidransky D., And El-Deiry W.S., (1998) Rare loss-of-function mutation of a death receptor gene in head and neck cancer, Cancer Res., 58, p. 3513-3518.
58. Parkin D.M., Whelan S.L., Ferlay J., Teppo L., And Thomas D.B., (2002) Cancer incidence in five continents, IARC scientific publication, 8, p. 7-12.
59. Patel S.G., And Shah J.P., (2005) TNM staging of cancers of the head and neck: striving for uniformity among diversity, CA Cancer J. Clin., 55, p. 242–58.
60. Perez-Ordonez B., Beauchemin M., And Jordan R.C., (2006) Molecular biology of squamous cell carcinoma of the head and neck, J. Clin. Pathol., 59, p. 445–453.
61. Poeta M.L., Manola J., And Goldwasser M.A., (2007) TP53 mutations and survival in squamous-cell carcinoma of the head and neck, N. Engl. J. Med., 357, p. 2552–2561.
62. Poetsch M., Lorenz G., And Kleist B., (2002) Detection of new PTEN/MMAC1 mutations in head and neck squamous cell carcinomas with loss of chromosome 10, Cancer Genet. Cytogenet., 132, p. 20-24.
63. Raghava G.P., Singh H., Chauhanjs., And Gromiha M.M., (2011) Open source drug discovery consortium, Nucleic Acids Res., 40, p. 486-489.
64. Rashid S., (2006) Bioinformatics resource development in Pakistan: a review, Proc Pakistan Acad. Sci., 43, p. 295-307.
65. Ries L.A.G., Melbert D., And Krapcho M., (2006) SEER Cancer Statistics Review, 1975–2004, Bethesda, MD, National Cancer Institute, 67, p. 1175–1180.
66. Rocco J.W., And Sidransky D., (2001) p16 (MTS-1/CDKN2/INK4a) in cancer progression, Exp. Cell Res., 264, p. 42–55.
67. Sankaranarayanan R., Ramadas K., And Thomas G., (2005) Effect of screening on oral cancer mortality in Kerala, India: a cluster-randomised controlled trial, Lancet, 365, p. 1927–1933.
68. Schwab M., (1999) Oncogene amplification in solid tumors, Semin, Cancer Biol., 9, 319-325.

69. Scott M., Boisvert F.M., Vieyra D., Johnston R.N., Bazett-Jones D.P., And Riabowol K., (2001) UV induces nucleolar translocation of ING1 through two distinct nucleolar targeting sequences, Nucleic Acids Res., 29, p. 2052–2058.
70. Sian A., Brancali A., And Simens C., (2011) Comparative modeling of 25-hydroxycholesterol-7α hydroxylase (CYP7B1) ligand binding and analysis of hereditary spatic paraplegia type 5 CYP7B1 mutations, Journal of Molecular Model, 13, p. 134-138.
71. Skowyra D., Zeremski M., Neznanov N., Li M., Choi Y., Uesugi M., Hauser C.A., Gu W., Gudkov A.V., And Qin J., (2001) Differential association of products of alternative transcripts of the candidate tumor suppressor ING1 with the mSin3/HDAC1 transcriptional corepressor complex, J. Biol. Chem. 276, p. 8734-8739.
72. Stambolic V., Suzuki A., de-la-Pompa J.L., Brothers G.M., Mirtsos C., Sasaki T., Rulans J., Penninger J.M., Siderovski D.P., And Mak T.K., (1998) Negative regulation of PKB/Akt-dependent cell survival by the tumor suppressor PTEN, Cell, 95, p. 29-39.
73. Steck P.A., Pershouse M.A., Jasser S.A., Yung W.K., Lin H., Ligon A.H., Langford L.A., Baumgard M.L., Hattier T., Davis T., Frye C., Hu R., Swedlund B., Teng D.H., And Tavtigian S.V., (1997) Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers, Nat. Genet., 15, p. 356-362.
74. Sturgis E.M., And Wei Q., (2002) Genetic susceptibility—molecular epidemiology of head and neck cancer. Curr. Opin. Oncol., 14, p. 310-317.
75. Suarez C., Rodrigo J.P., Ferlito A., Cabanillas R., Shaha A.R., And Rinaldo A., (2006) Tumours of familial origin in the head and neck, Oral Oncol., 42, p. 965–978.
76. Szklarczyk D., Franceschini A., Kuhn M., Simonovic M., Roth A., Minguez P., Doerks T., Stark M., Muller J., Bork P., Jensen J.L., And Mering V.C., (2011). The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored, Nucleic Acids Research, 39, p. 561-568
77. Tabor H.K., Risch N.J., And Myers R.M., (2002) Candidate-gene approaches for studying complex genetic traits: practical considerations, Nat. Rev. Genet., 3, p. 391-397.
78. Tamura K., Peterson D., Peterson N., Stecher G., Nei M., And Kumar S., (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood,

Evolutionary Distance, and Maximum Parsimony Methods, Molecular Biology and Evolution, 28, p. 2731-2739.

79. Tovchigrechko A. And Vakser L.A., (2006) GRAMM-X public web server for protein–protein docking, Nucleic Acids Research, 34, p. 310-314.
80. Tovchigrechko A., Wells A.C., And Vakser A.I., (2002) Docking of protein models, Protein Science, 11, p. 1888–1896.
81. Trizna Z., And Schantz S.P., (1992) Hereditary and environmental factors associated with risk and progression of head and neck cancer, Otolaryngol Clin North Am., 25, p. 1089–1103.
82. Tsugawa K., Jones M.K., Sugimachi K., Sarfeh I.J., And Tarnawski A.S., (2002) Biological role of phosphatase PTEN in cancer and tissue injury healing, Front. Biosci., 7, p. 245-251
83. Vazquez F., Grossman S.R., Takahashi Y., Rokas M.V., Nakamura N., And Sellers W.R., (2001) Phosphorylation of the PTEN Tail Acts as an Inhibitory Switch by Preventing Its Recruitment into a Protein Complex, J. Biol. Chem., 276, p. 48627-48630.
84. Vineis P., Alavanja M., And Buffler P., (2004) Tobacco and cancer: recent epidemiological evidence, J. Natl. Cancer Inst., 96, p. 99-106.
85. Vokes E.E., Weichselbaum R.R., Lippman S.M., And Hong W.K., (1993) Head and neck cancer, N. Engl. J. Med., 328, p. 184-194.
86. Wang X., Gjorloff-Wingren A., Saxena M., Pathan N., Reed J.C., And Mustelin T., (2000) The tumor suppressor PTEN regulates T-cell survival and antigen receptor signaling by acting as a phosphatidylinositol 3-phosphatase, J. Immunol., 164, p. 1934-1939.
87. Weng L., Brown J., And Eng C., (2001) PTEN induces apoptosis and cell cycle arrest through phosphoinositol-3-kinase/Akt-dependent and -independent path-ways, Hum. Mol. Genet., 10, p. 237–242.
88. Westra W.H., (2009) The Changing Face of Head and Neck Cancer in the 21st Century: The Impact of HPV on the Epidemiology and Pathology of Oral Cancer, Head Neck Pathol., 3, p. 78–81.

89. Wiederstein M., And Sippl M.J., (2005) Protein sequence randomization: efficient estimation of protein stability using knowledge-based potentials, J. MOL. BIOL, 345, p. 1199-1212.
90. Witcher T.P., Williams M.D., And Howlett D.C., (2007) "One-stop" clinics in the investigation and diagnosis of head and neck lumps, Br. J. Oral Maxillofac. Surg., 45, p. 19-22.
91. Zeremski M., Horrigan S.K., Grigorian I.A., Westbrook C.A., Gudkov A.V., (1997) Localization of the candidate tumor suppressor gene ING1 to human chromosome 13q34. Somat, Cell Molec. Genet., 23, p. 233-236.
92. Zhu M., And Zhao S., (2007) Candidate Gene Identification Approach: Progress and Challenges, Int. J. Biol. Sci., 3, p. 420-427.