CAPITAL UNIVERSITY OF SCIENCE AND
TECHNOLOGY, ISLAMABAD

Elastic Network Modeling of Human Papilloma Virus Proteins

by

Rabbiah Manzoor Malik

A thesis submitted in partial fulfillment for the
degree of Doctor of Philosophy

in the

Faculty of Health and Life Sciences
Department of Bioinformatics and Biosciences

2021

Elastic Network Modeling of Human Papilloma Virus Proteins

By

Rabbiah Manzoor Malik

(PI141001)

Dr. M. Saeed Sheikh, Professor

State University of New York, USA

(Foreign Evaluator 1)

Dr. Rozelin Aydin, Assistant Professor

Adana Bilim Ve Teknoloji Univeristesi Adana, Turkey

(Foreign Evaluator 2)

Dr. Sahar Fazal

(Thesis Supervisor)

Dr. Sahar Fazal

(Head, Department of Bioinformatics and Biosciences)

Dr. Muhammad Abdul Qadir

(Dean, Faculty of Health and Life Sciences)

DEPARTMENT OF BIOINFORMATICS AND BIOSCIENCES

CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY

ISLAMABAD

2021

Dedicated To my beloved parents who taught me every aspect of life.

Dedicated to my respected supervisor **Dr. Sahar Fazal**, without his sincere help and support this work could not have been done.

CAPITAL UNIVERSITY OF SCIENCE & TECHNOLOGY
ISLAMABAD
Expressway, Kahuta Road, Zone-V, Islamabad
Phone:+92-51-111-555-666 Fax: +92-51-4486705
Email: info@cust.edu.pk Website: https://www.cust.edu.pk

CERTIFICATE OF APPROVAL

This is to certify that the research work presented in the thesis, entitled "**Elastic Network Modeling of Human Papilloma Virus Proteins**" was conducted under the supervision of **Dr. Sahar Fazal**. No part of this thesis has been submitted anywhere else for any other degree. This thesis is submitted to the **Department of Bioinformatics & Biosciences, Capital University of Science and Technology** in partial fulfillment of the requirements for the degree of Doctor in Philosophy in the field of **Biosciences**. The open defence of the thesis was conducted on **November 15, 2021.**

Student Name : Rabbiah Manzoor Malik
(PI-141001)

The Examination Committee unanimously agrees to award PhD degree in the mentioned field.

Examination Committee :

(a) External Examiner 1: Dr. Ghazala Kaukab
Professor
PMAS-Arid Agriculture University, Rawalpindi

(b) External Examiner 2: Dr. Amjad Ali
Associate Professor
ASAB, NUST, Islamabad

(c) Internal Examiner : Dr. Arshia Amin Butt
Assistant Professor
CUST, Islamabad

Supervisor Name : Dr. Sahar Fazal
Professor
CUST, Islamabad

Name of HoD : Dr. Sahar Fazal
Professor
CUST, Islamabad

Name of Dean : Dr. Muhammad Abdul Qadir
Professor
CUST, Islamabad

AUTHOR'S DECLARATION

I, **Rabbiah Manzoor Malik (Registration No. PI-141001)**, hereby state that my PhD thesis titled, '**Elastic Network Modeling of Human Papilloma Virus Proteins**' is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/ world.

At any time, if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my PhD Degree.

(Rabbiah Manzoor Malik)

Dated: November, 2021 Registration No : PI-141001

PLAGIARISM UNDERTAKING

I solemnly declare that research work presented in the thesis titled "**Elastic Network Modeling of Human Papilloma Virus Proteins**" is solely my research work with no significant contribution from any other person. Small contribution/ help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/ cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of PhD Degree, the University reserves the right to withdraw/ revoke my PhD degree and that HEC and the University have the right to publish my name on the HEC/ University Website on which names of students are placed who submitted plagiarized thesis.

Rabbiah

(Rabbiah Manzoor Malik)

Dated: November, 2021 Registration No : PI-141001

List of Publications

It is certified that following publication(s) have been made out of the research work that has been carried out for this thesis:-

1. **R. Malik.**, S. Fazal. Insights into the Dynamic Fluctuations of the Protein HPV16 E1 and Identification of Motifs by Using Elastic Network Modeling. Protein and Peptide Letters. 2021 Apr 14. doi: 10.2174/0929866528666210415114858. Epub ahead of print. PMID: 33858307.

2. **R. Malik**, S. Fazal, M. A. Kamal. Computational Analysis of Dynamical Fluctuations of Oncoprotein E7 (HPV 16) for the Hot Spot Residue Identification Using Elastic Network Model. Letters in Drug Design & Discovery. Volume 17, 2020. DOI : 10.2174/1570180817999200606225735

Rabbiah Manzoor Malik
(PI141001)

Acknowledgement

With love, First and foremost I would like to dedicate this work to the Almighty ALLAH, Who bestowed me with the courage and faith to complete this work. I have to thank my family for their prayers, love and support throughout this difficult voyage, my mentors for their continuous guidance, my husband for his continuous inspiration, support and prayers and my children, Abdul Hannan Khan and Fatima-tuz-Zahra.

Abstract

Cancers are the main reason for elevating death rate among human population all over the world as well as in Pakistan. Low cancer survival rate and lack of awareness regarding cervical cancer risk is raising significant cancer burden in females of Pakistan therefore; our country is in terrible need of cancer management and prevention measures. It has been reported in literature that more than 30% of cancer deaths could be prevented by modifying or avoiding key risk factors responsible for cancer development. Early detection, accurate diagnosis, and effective treatment, may help increase cancer survival rates and reduce suffering. There are a group of viruses including Hepatitis Virus, Papilloma Virus, Epstein Barr Virus, which in their advance stage of infection cause development of cancers. It has been found that a major cause of cervical cancer is Papilloma viral infection. Human Papillomavirus (HPV) have a specific mechanism by which it causes pathogenesis in humans, which ultimately leads to cervical cancer. The major mechanism involved in this pathogenesis is interaction of specific viral proteins with human proteins resulting in carcinogenesis. In recent years characterization of proteins has been enhanced by the development of elastic network modelling, among which Anisotropic Network Modelling helps in explaining functioning of proteins while Gaussian Network Modelling covers the structural aspect of proteins. In present project the techniques of Anisotropic network modelling and Gaussian network modelling have been applied to have a deep analysis of interactions of HPV viral and human proteins involved in carcinogenesis of cervical cancer. In first step, amino acids sequences of HPV proteins E1, E2, E4, E5, E6 and E7 were mapped on sequences of Eukaryotic Linear Motifs (ELM) in human beings, by using tool ELM. There were retrieved seventy-seven ELMs which had sequences identical to those of human ELMs. On that basis it could be considered that these ELMs are present on HPV proteins. As ELMs interact with their counter domains in other proteins to establish a protein-protein interaction. we predicted protein domains which interact with those conserved linear motifs. Human proteins were predicted which may be targeted by HPV proteins by establishing interactions through a

motif conserved in HPV and the equivalent interacting protein domain. Our predicted proteins were enriched with host proteins known to interact with HPV proteins E1, E2, E4, E5, E6 and E7. We found the Cancer Pathways to be statistically enriched for our predicted proteins. We also analyzed our predictions for enrichment with Gene Ontology molecular function of level 5 categories with both predicted and confirmed HPV targeted proteins. The enriched functions included molecular activities associated with events of phosphorylation, protein kinases and adenyl ribonucleotide binding. We modeled 3D structures of HPV proteins by using FASTA sequences of proteins. These 3D structures were used as input for modeling elastic networks of HPV proteins. Elastic network gave us insights into the HPV proteins, for their flexible regions and their molecular dynamics. On the basis of our findings of the flexible regions shown by elastic network models of HPV proteins, sequence and site of presence of motifs on HPV proteins, predicted proteins, and previous literature about the interactions of HPV with human proteins we predicted new motifs on HPV proteins and domains in human proteins involved in interactions of HPV with human proteins. We also predicted hot spot residues of HPV proteins, for already reported interactions of HPV proteins E2, E6, and E7 with DNA, hDlg and retinoblastoma, respectively. Our study authenticates the part of linearly binding motifs commonly shared between virus and human proteins as a significant part of the crosstalk among virus and host. We also validated the previously reported interactions of HPV proteins with human proteins, by the elastic network models on the basis of flexibility, dynamic fluctuations, correlations analysis and deformation energies, of those specific regions involved in those interactions

Contents

List of Figures

List of Tables

Abbreviations

3D	3 Dimensional
ANM	Anisotropic Network Model
BL	Burkitt's lymphoma
CBP	CREB Binding Protein
CD	Counter Domains
DNA	Deoxy Ribonucleic Acid
E6AP	E6 Associated Protein
EBV	Epstein-Barr virus
ECM	Extracellular Matrix
ELM	Eukaryotic Linear Motifs
EM	Electron Microscopy
ENM	Elastic Network Modeling
GNM	Gaussian Network Model
GO	Gene Ontology
HBV	Hepatitis B Virus
HCV	Hepatitis C
HEP	Human Experimental Proteins
HHV8	Human Herpes Virus 8
HIV	Human Immunodeficiency Virus
HPP	Human Predeicted Proteins
HPV	Human Papillomavirus
HSPG	Heparin Sulphate Proteoglycans
HSV	Herpes Simplex Infection
KSHV	Kaposi's sarcoma herpesvirus

MD	Molecular Dynamics
mRNA	Messenger RNA
MS	Mass Spectrometry
NMA	Nuclear Magnetic Anisotropy
ORF	Open Reading Frames
PDB	Protein Database
PPI	Protein-protein Interactions
PV	Papillomavirus
Rb	Retinoblastoma
RNA	Ribonucleic acid
SLiMs	Short Linear Motifs
US	United States
VLP	Viral Like Particles
WHO	World Health Organization

Symbols

α	Alpha
β	Beta
κ	Kappa
γ	Gamma
ϵ	Epsilon
η	Eta
τ	Tau
ζ	Zeta
Φ	Phi

Chapter 1

Introduction

1.1 Background

Viruses are small obligate intracellular parasites, which are defined as entities containing a virus coded protein, coated and protected genome which could be either a RNA or DNA. Viruses could be considered as versatile hereditary material containing entities, most presumably of cellular origination and described by a long transformative relationship among infecting particle and host [1]. Specific host organisms by which complex metabolic and biosynthetic machinery of eukaryotic or prokaryotic cells is provided, play major role in propagation of these pathogens. A virus with its genome and all other components is called a virion. For synthesis of viral proteins in the host cell virion delivers its genetic material either in the form of DNA or RNA into the host, and this is the major function of virion. Viruses are also armed by a capsid which is composed of symmetric proteins and it plays the role for packaging of the viral genome. The capsid proteins are often associated with basic proteins. Viruses also have the nucleocapsid composed of nucleic acid-associated protein, always present in association with the genome. In enveloped viruses, a lipid bilayer surrounds the nucleocapsid that has been originated through the cell covering in the cell of host, which is modified as well as covered with an external layer of virus envelope glycoproteins [2, 3].

1.1.1 Viruses and Cancers

It has become exceedingly apparent during the past 30 years that in human cancers which develop in more than one stages, several viruses play significant roles;

in approximation less than 30% of malignancies have been found linked with propagation of viruses. The viruses are associated with one or more different steps of the carcinogenesis, which could be contributed by oncogenic viruses [4, 5]. Multiple lines of investigations which have provided the information about the cause of various human cancers can be strengthened by in depth study of mechanism of oncogenesis by oncogenic viruses along with the findings and investigations of major pathways of cell, that are involved in development of carcinogenesis in general.

The history, of the belief that cause of cancer could be infectious in nature, emerged during past have been indicated by description as "cancer houses" where some type of tumor was developed by many occupants. Further support to this concept of an infectious etiology of cancers was provided by observations of occurrence of similar cancers among married couples and appearance of cancer to be transmitted from one generation to other.

Somehow, in previous century, detailed studies were not succeeded in elaborating the functions of cancer-causing parasites, fungus or bacteria resulting in development of concept that none of infectious agent causes cancer. In 1898 a study was presented showing the transmission of oral dog warts without any type of cells, with cell free concentrates, similar transmission studies were published by Ciuffo with human warts in 1907 [6]. Due to being non-malignant and benign hyperplasias in nature, these findings of warts were not fully appreciated despite of significance. In 1908, Ellermann and Bang presented a study in which they demonstrated that concentrates of leukemic cells or blood from birds suffering from leukemia could transmit leukemia in other birds [7]. In fact, the discovery of induction of murine leukemias by viruses lead to the importance and appreciation of this study. During next twenty years study of many cancers took place, which lead to finding of new tumor causing viral particles, the Noble Prize was awarded to Rous for his pioneering work in 1966, as a recognition for significance of the previous research about animal cancer infections [6, 7]. Renewal of interest of scientists in link of viruses with human cancer took place after the Epstein-Barr virus (EBV) finding in cells taken and studied from Burkitt's lymphoma (BL) by

electron microscopy (EM) in 1964 with finding of hepatitis B virus (HBV) in a person having significant surface antigens for hepatitis B in 1970, along with the development of cell culture and systems for animal models. After the long continuation of search for additional human tumor viruses, three major discoveries during early 1980s, despite many setbacks, resulted ultimately to the acceptance of a relation of viruses with cancers in living beings. In next years, isolation of Human Papillomavirus Virus (HPV) type 16 and 18 was done by using samples of cervical tissue in human having cancer. The second major role playing discovery was the development of a convincing connection between persevering HBV disease and liver carcinogenesis, resulting from a large scale epidemiological study [6]. The patients of T-cell lymphoma/leukemia were used to isolate virus of Human T-cell leukemia (HTLV-I) in the third major discovery [7]. All these discoveries show the association of viruses with the pathological lesions. In further studies hepatitis C virus (HCV) and Kaposi's sarcoma herpesvirus (KSHV)/ human herpes-virus 8 (HHV8) [8] have also been found to be associated with cancers. Today, after acceptance of viruses being the cause of human cancers, it has been assessed that around 15-20% of all human cancerous growth may have a viral origin [4, 5, 7]. Unlike the agents which are carcinogenic in nature, the oncogenic viruses are infectious also. Major requirement for understanding the mechanism of cancer caused by viruses completely, there is a need of a detailed investigation for host reaction and viral pathogenesis. Such a comprehension, thus, has expanded our insight into cell pathways included in development and separation and neoplasia in general [9]. There are certain common features shared by all oncogenic viruses, despite of their different groups and different strategies to contribute to cancer pathogenesis. One major property is that they only infect their host cell, but not kill. Oncogenic viruses have a different mechanism of virulence from other viruses that is the ability to develop infections which can persist for longer periods [10].

According to estimate every 15 cancers out of 100 in humans worldwide are due to viruses, carrying a critical segment of burden of the global cancer. Humans experience cancers by both types of viruses, like RNA and DNA. Human Papillomavirus, Epstein Barr virus, Human Herpes virus-8 and Hepatitis B virus are the four DNA viruses that are responsible for the emergence of cancers in humans.

Among RNA viruses Hepatitis C viruses and Human T Lymphotrophic virus type 1 contribute to human cancers [11].

1.1.2 Cervical Cancer

Papilloma viruses (PVs), belonging to the family Papillomaviridae are small-sized, un-capsulated viral particles with double stranded genome. The viral particles invade different organs of host through squamous type of epithelial cells [12]. To date, almost 200 types of Papillomavirus (HPV) have been discovered. Papilloma viral particles result in variety of aggressive growths of cells of epithelial layers and are grouped as viruses of mucosa and cutaneous. The classes are re-grouped into high-risk and low-risk, on the basis of the related propensity for malignant lesion development [13]. The second most commonly occurring cancer of women in the world is cancer of the cervix uteri, with 311,000 deaths and estimated 570,000 new cases in 2018. Cervical tumor is the second most common cause of disease induced mortality in females around the world [14]. Squamous cell carcinoma and adenocarcinoma are among the major outcomes of cervical cancers. HPV and HBV/HCV are responsible for 25% of deaths due to tumors in middle and low income regions of the world [15]. All around, the burden of HPV contamination bringing about cervical neoplasia is high in developing nations. The oncogenic conversion of the genotypes differs in every other human being. A study investigated in 2013 from Brazil identified that 170 of 172 specimens (99%) were observed to be certain for HPV. The most common HPV genotypes found were 77.6% (HPV16), 12.3% (HPV18), 5.9% (HPV35), 8.8% (HPV31), and 7.1% (HPV33). A large portion of these contaminations were created by individual HPV16 genotypes in Brazilian ladies [16]. HPV 16 and 18 are responsible for 70% of the cervical cancer cases in world [17]. In Pakistan on average 57.22 million females, belonging to the age group of 15-20 years, are at the risk of developing cervical cancer. Cervical cancer in Pakistan is ranked being the third common type of cancer among the females, and in the age group of 15 to 44 years old, it comes the second most. In Pakistan PV has been identified in cervix among 5 out of every 100 females. The cervical growths in community have been assessed to harbor HPV-16 and 18, and 88.1

out of 100 obtrusive growths of cervix have been credited with Human Papilloma Viruses 16 and 18 type [18, 19]. There are different epidemiological studies that have recognized as risks associated with cervical disease. The factors, which expand the recurrence of the disease, incorporate low financial status, marriage at a very early age, various sexual accomplices, long term utilization of oral contraceptives, supplement insufficiency, tobacco use, poor individual cleanliness, and viral contaminations, like that of HPV, HIV, Herpes Simplex Infection (HSV) Type II, and bacterial diseases due to Chlamydia trachamatis. Other significant factors that also contribute to the disease are an increase in the population, an increase in the number of old people, and urbanization. The hereditary variables, which contribute in onset of cervical cancer include dynamic oncogenes and tumor silencer proteins [20, 21]. Figure1.1 explains the risk factors causing the cervical cancer [22]. A reliable, effective and inexpensive screening test for cervical cancer is the PAP Smear. The knowledge and practice of the women is inadequate in under developed countries and need to be promoted [23]. Cellular changes before cancer and growths distinguished by tests of Pap smears are being adequately handled by removal of lesions through surgery. Without fruitful screening, infection could not be identified on time. Other helpful choices after surgery are radiotherapy and chemotherapy treatment, that could be connected by numerous complications and don't accomplish an enduring treatment [24]. Although, modulation of cellular pathways that impinge on multiple cervical cancer hallmarks is the basic requirement for the replication strategy of high-risk human papillomaviruses. The infected host has a strict regulatory mechanism to check the expression of viral transforming proteins, due to which infected cells are rarely transformed to malignancy. The infected cell become at a higher risk (HPV 16 and 18) for their transformation into oncogenes after dysregulation of synthesis of proteins of HPV E7 and E6 synthesis due to result of either the insertion of viral genetic material into the genome of the host or due to epigenetic changes of the viral genome. The oncogenic activities of the high-risk HPVs are directly linked with the ability of the E6 and E7 proteins to promote genomic instability, in this manner quickening the foundation and extension of cells with tumor advancing host cell mutations [25–28].

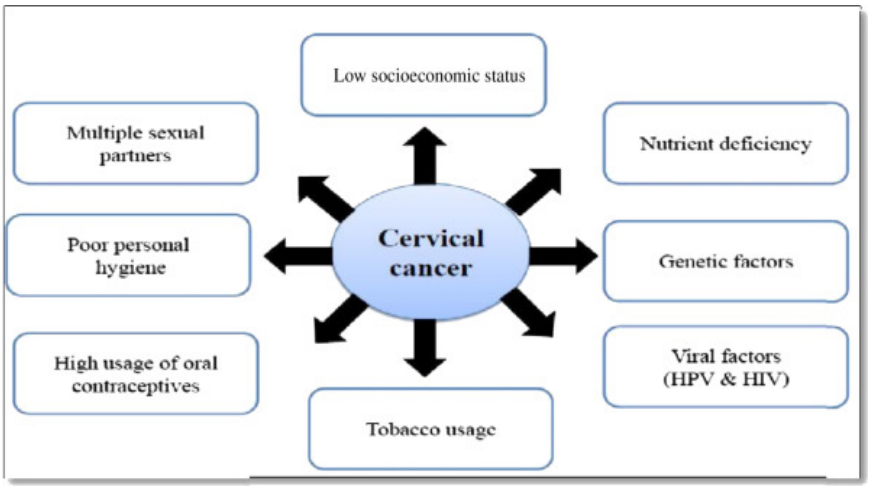


FIGURE 1.1: Multiple risk factors leading to cervical cancer [22]

1.1.3 Protein Interactions

Proteins are the structural organizations of life and every procedure in the cell is directed by mind boggling communications of proteins [29, 30]. These macromolecules are built from 20 different subunits called amino acids. Such sub-atomic combinations are being turned out as an imposing development unit for nature and, despite the fact that being made out of just these couple of fixings, proteins are gifted by an amazing practical assorted qualities over the span of advancement for all targets and purposes each procedure in the cell [29, 31, 32]. Proteome of viral particles play a major role in causing infection Fig. (1.2). Due to smaller genome and less number of proteins expressed viruses use cellular machinery of host to complete its life cycle. Interaction of viral proteins with that of host cells are the key players of malignant transformation of viral particles [33].

1.1.4 Previous Studies on Protein Interactions

Initially the research work on proteins was only based on observations. The principal thoughts regarding the interior working components have been developed by experimental results. Protein dynamics are theoretically described by designing kinetic rate equations. The properties of some proteins, along with the mechanism forecast of structural arrangements and functioning, can be qualitatively described by these kinetic models in combination with experimental data. But drawback of these results is that they lack the structural details explained by resolutions and, accordingly, of no use to help in completing of protein dynamics model [34].

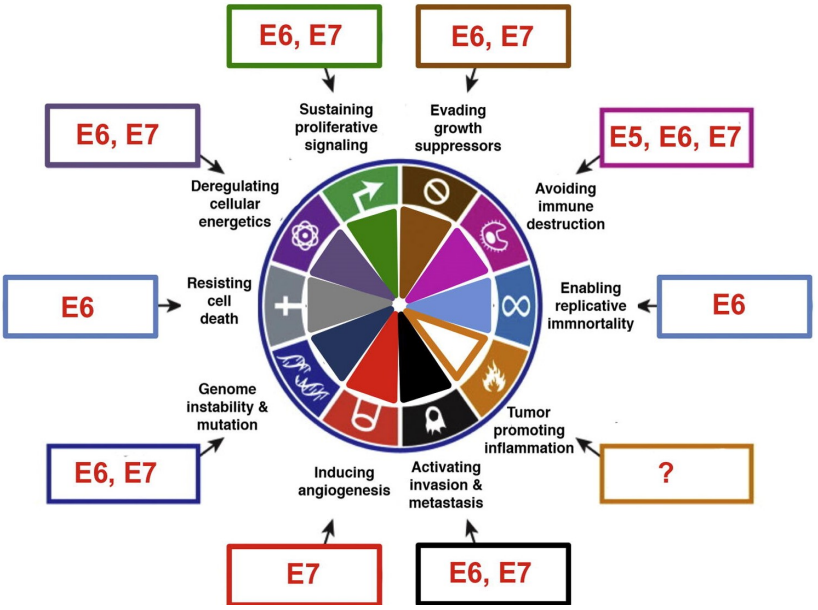


FIGURE 1.2: Hallmarks of Cancer analysis of HPV associated cervical carcinogenesis [27]

The major component of the field of proteomics in today's scientific world, is methodologies for study of protein-protein interactions, which goals to define protein functions, abundances, interactions and modifications. Protein-protein interactions are characteristically dynamic, and are complex combinations of transient and stable interactions, which usually co-exist. Protein associations to establish functional relevancies can be studied by Immuno-affinity purification (IP) techniques along with mass spectrometry (MS). These types of affinity-based techniques have undergone a significant improvement in recent years. These improvements include advances in methods for sample preparation, affinity tools, configurations of mass spectrometry resulting in enhanced accuracy and sensitivity, with additional approaches in bioinformatics enabling a detailed analysis of large-scale, specific and targeted datasets of interactome. By using a wide variety of research techniques diverse biological systems can be studied in detail, by identification of protein interactions. In recent years, studies of proteins have entered the era of techniques of molecular dynamics (MD) simulation [35]. In the last decade application of proteomic approaches has much increased in field of studies of virology, by using different techniques like Advanced mass spectrometry-based proteomics, Antibody-based immunoaffinity purification of protein complexes, Yeast-two hybrid, Cross-linking, and Nucleic Acid Programmable Protein Array [36–39].

Approachable techniques of intermediate complexity have provided the solution [40]. By accurate coarse-graining about protein structural organization with linkage among building blocks like atoms, technical expenditures and costs of molecular dynamics techniques, could be minimized , but we also experience absence of resolution movement in such models. Another approach of Elastic Network Modelling is developed for description of detailed movement of building blocks and is in continuous process of extension and advancement [41–45].

PPI networks have been found useful for annotating proteins of unknown function [46, 47], predicting other interactions types [48], investigating the peptide regions guiding interactions [48–50], and identifying protein complexes [51]. Study of these interactions helps to study host-pathogen networks, which involve relationship of host proteins with pathogenic or viral proteins that results in alteration of host interaction networks [52–54].

1.2 Research Problem

Interactions of Human Papilloma Viral proteins with human proteins have not been analyzed at residual level in the current literature for cervical cancer. Our Research problem is based on residual level exploratory analysis of HPV protein interactions with human proteins associated with cervical cancer. The fluctuation dynamics of proteins can be modelled as elastic networks. The nodes are the amino acids residues of proteins. The linkers represent the inter-residue potentials stabilizing the folded conformations of proteins. This helps to enhance our knowledge on understanding of biological functions of macromolecules based on their structural information. These networks also explain dynamics of functional motions of protein complexes, ligand binding mechanisms, refinement of low resolution structural data, and transition pathways generation.

1.3 Research Objectives

- To predict the human proteins possibly interacted by HPV proteins by using resources of Protein Data Bank, RaptorX, Eukaryotic Linear Motifs (ELMs) and PROSITE.

- To investigate and analyse HPV protein interactions with human proteins at residual level by using Elastic Network Modelling
- To analyze the behavior of individual amino acids residues of HPV proteins with the help of Elastic Network Modeling
- Prediction of motifs and domains on HPV proteins on the basis of data obtained from ELMs, and Elastic Network Modeling.
- Prediction of hot spot residues of HPV proteins.

1.4 Research Philosophy

During viral infection, the host's cellular machinery is interacted by virus proteins, resulting in alteration of the host cellular mechanism to favor replication of virus. The identification and annotation of virus-host interactions are important for discovering the new drug targets as well as for assessment of the efficiency of therapies of antiviral drug on host populations. Additionally, virus-host protein interactions can also be used to discover roles of virus proteins, and inspect common viral approaches for interacting with the hosts. Researchers have collected huge data about virus-host protein interactions, to find out the factors of host required necessarily for completion of viral life cycle. The interactions transform into the form of virus-human interactions of proteins, expression of human genes in response to infection, and genetic alteration of human genes which are necessary for the survival of virus. In development of viral infections some human genes also play a major role in development of pathogenesis by viruses, these roles can be found and analyzed by study of data about viral-host protein interactions and gene expression. So far, these interactions have been studied extensively, for only three types of viruses, influenza A virus, hepatitis C virus (HCV), and human immunodeficiency virus (HIV). There are numerous experimental challenges for identification of virus-host protein-protein interactions. The biggest trial is designing of screens, severe enough for having low false positive rates while guaranteeing that the number of real protein interactions which cannot pass through these rigorous assays is kept back to a minimum. Protein interactions are temporary in nature,

and furthermore these protein-protein binding interactions may be dependent on the presence of cofactors which are not essentially present in these binding assays. These difficulties suggest supplementing experiments with a bioinformatic approach for predicting and understanding virus-host interactions. This has been endeavored for Human Papilloma Virus (HPV) by predicting HPV-human protein-protein interactions, but it was seen that the most predictive feature about the human protein was the number of interacting partners of that particular protein, which made it difficult to find sites for drug targeting. Such results show that more investigation is required to understand the fundamental principles of these interactions, and also to guide new studies.

All biological processes rely on interactions between proteins. For a better understanding of biological processes we should have a deep insight into the proteins interactions. Whenever the proteins interact, there is involvement of the entire interfaces between two interacting proteins. However, it has been shown that the recognition and binding of proteins essentially requires involvement of only few of the interface residues. So the first step toward understanding the function of proteins and studying their interactions is the identification of these residues, commonly referred to as binding "hotspots". One of the experimental ways to find out the hotspots, is to identify them by mutating single residue, but it is a laborious and expensive procedure and cannot be applied on large scale. In this study it has been shown that hotspots can be computationally identified from the information of sequence of amino acids of a protein, without having the knowledge of its interacting protein. An accurate and efficient identification of hotspots from amino acids sequence of proteins helps us to analyze protein-protein interactions in an entire organism and thus may aid in predicting functions of proteins and developing drugs.

The overall goal of this dissertation is to estimate and analyze virus-host interactions, on the basis of identification of hotspots of HPV proteins with the help of elastic network modelling. These interactions have been organized into networks, where proteins are nodes and network edges represent interactions between proteins, and elastic networks where amino acids residues are nodes and their links

are represented by elastic spring projections showing the fluctuating behavior of proteins. We begin this thesis by predicting virus-host protein interactions using short peptide motifs that have been shown to play a role in protein interactions. We demonstrate the validity of our predictions using HPV-human interactions. Next we model the viral proteins through elastic networks explaining the behavior of proteins and role of individual amino acid residues in maintaining the virus host interactions. We conclude this dissertation with a discussion of how the work presented here aids virus-host network analysis and how our findings could help in advancement of drug designing against the pathogenesis of HPV (understanding the mechanism of virulence at residual level).

1.5 Research Hypothesis

Viral proteins cause pathogenesis in human body by interacting with human proteins. Not only the human proteins which are directly interacted by viral proteins in causing pathogenesis are affected by viral proteins, but in addition other proteins which interact with these affected human proteins, are also affected and interacted by viral proteins causing complications of pathogenesis or cancers caused by viruses.

These interactions are stabilized by the specific amino acid residues. These amino acids are hotspots residues which could be used as drug targets for prevention of progression of HPV infection into cervical cancer. Specific regions of these viral proteins and host proteins could be predicted to contain specific motifs or domains playing major role in protein interactions.

1.6 Research Methodology

The methodology that has been adopted to carry out this research work is divided into different phases from identification of motifs in viral proteins and prediction of human proteins having interactions with HPV proteins to the analysis of insights of interactions of HPV proteins with human proteins on the basis of Elastic Network Models of 6 HPV proteins.

1.6.1 Retrieval of Amino Acid Sequence of HPV Proteins E1, E2, E4, E5, E6 and E7:

Amino acid sequences of HPV proteins were retrieved from UniProt. The UniProt Knowledgebase (UniProtKB) is a part of The Universal Protein Resource (UniProt) which is a comprehensive resource for protein sequence and annotation data.

1.6.2 Virus Protein ELM Annotation and Conservation:

Mapping of amino acid sequence of HPV proteins on sequences of motifs in human by using ELMs Resource. The main focus of this computational biology resource is detection and annotation of eukaryotic linear motifs (ELMs) by the help of a repository of an exploratory tool for prediction of motif and annotated motif data.

1.6.3 Human Protein CD Annotation on PROSITE:

PROSITE is a database consisting of protein domains and families of proteins containing these domains.

1.6.4 Prediction of Human Proteins which Interact with HPV Proteins for Developing Cancers:

On the basis of presence of sequences of human motifs in HPV proteins, they can interact with human predicted proteins having counter domains of these motifs.

1.6.5 Validation of Predicted Protein Data:

Mapping of genes of predicted human proteins interacting with viral proteins on GO functional analysis and KEGG pathways.

1.6.6 3D Structure Modeling of HPV Proteins:

RaptorX is a server to be used for prediction of protein 3D structure.

1.6.7 Elastic Network Modelling of Proteins of HPV:

Modeling of Elastic Networks of HPV proteins by using ANM2.1. ANM2.1 is a simple NMA tool for analysis of vibrational motions in molecular systems. It uses Elastic Network methodology by representing the system at the residue level.

1.6.8 Literature Mining & Documents Retrieval from PUBMED

Use of keywords related to HPV proteins, human predicted proteins to explore the literature regarding transformation of HPV infection to Cervical cancer and other clinical manifestations.

1.6.9 Analysis of Interaction of HPV Proteins with Human Proteins:

Insights into the interactions of HPV proteins with human proteins on the basis of presence of human motifs in HPV proteins, ENMs of HPV proteins and literature retrieved from PUBMED.

Chapter 2

Literature Review

2.1 Oncogenesis in Viral Particles and its Mechanism

Oncogenic viruses increase process of transforming cells, prompt wild growth of cells resulting in formation of spreadable cancers as shown in Fig. 2.1 [55]. The tumors which possess ability to metastasize are cancerous [56]. Lymphomas, leukemia in children, and different solid tumors possess cancer causing abilities. The complicated cancer developing mechanism starts from the invasion of viral genome and transforming effects on host cellular machinery [55]. The genes of virus DNA which cause transforming properties to start and alter the mechanism of flourishment of cell resulting in formation of new protein molecules are termed as viral oncogenes (v-oncogenes) [57]. Amplification, deletion, point mutation and chromosomal translocation are the processes by which genes are changed into oncogenes as shown in table 2.1 [58].

Genetic material is protected and cellular processes are controlled by a continuous and constant war between genes which suppress tumors and oncogenes. A number of researches have shown that cancers are developed by victory of oncogenes from tumors suppressing genes [59].

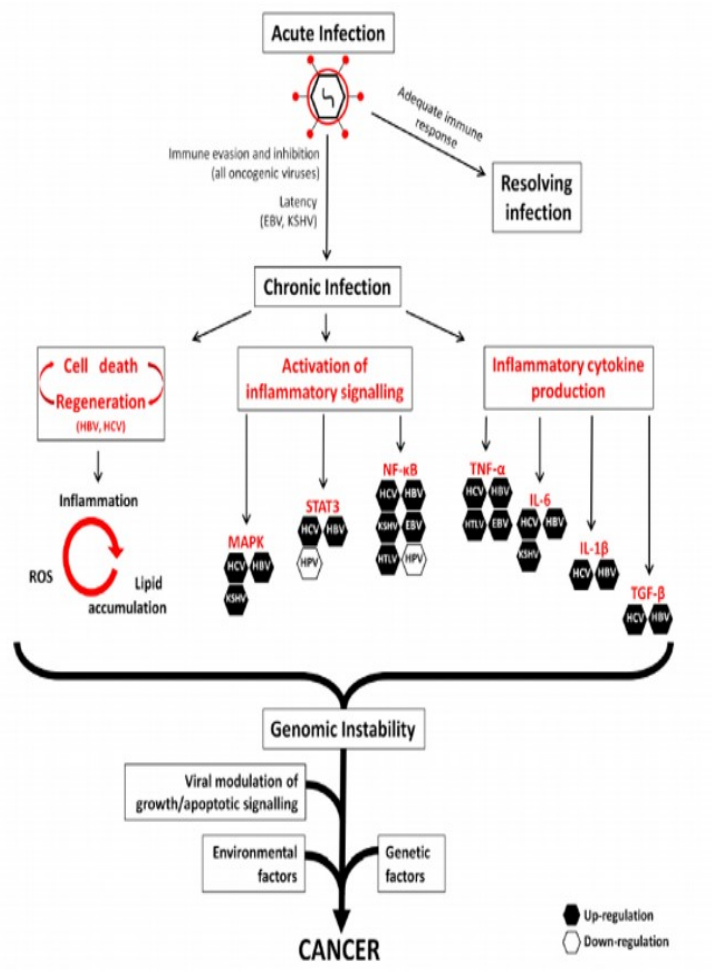


FIGURE 2.1: Mechanism of viral infection developing to cancers modified from Mesri et al, 2014 [27]

2.2 Human Oncogenic Viruses

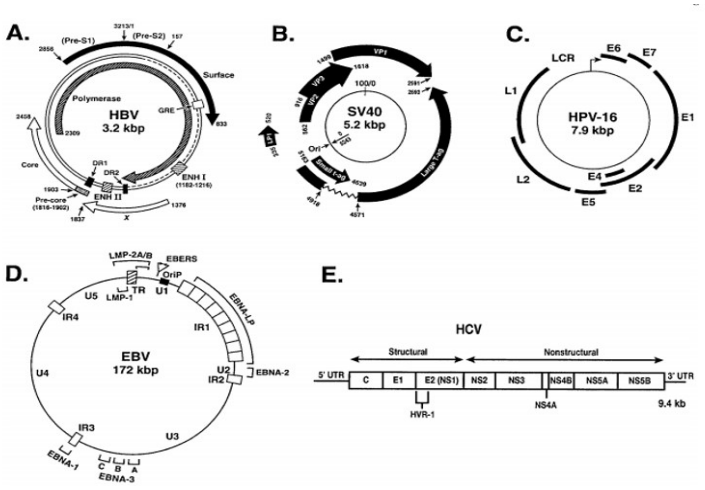


FIGURE 2.2: Genome maps of human tumor viruses modified from Butel 2000 [60]

Tumor causing viruses in human are members of different families, among which Hepadnaviridae, Herpesviridae, and Papillomaviridae are the families of DNA viruses, while the families which have RNA are Flaviviridae and Retroviridae Fig. 2.2. According to latest findings; adult T-cell leukemia (ATL) is caused by HTLV-1 [61], cervical cancer and skin cancer in patients with epidermodysplasia verruciformis (EV), head and neck cancers and other anogenital cancers by HPV [62–65], Kaposi's sarcoma (KS), Castleman's disease and primary effusion lymphoma by HHV-8 [66, 67], Hodgkin's disease, Burkitt's Lymphoma (BL), post-transplant lymphomas and nasopharyngeal carcinoma (NPC) by EBV [68–70] and hepatocellular carcinoma (HCC) by HBV and HCV [71–75]. Viruses which do not directly cause cancer but have potentials to lead their infections towards cancers include Simian Vacuolating Virus 40 (SV40) causing malignancy of nervous organs, and bone cancer [76], BK virus [BKV] causing prostate cancer, JC virus (JCV) causing cancer of nervous organ [77], while ovarian cancer, germ cell tumors, melanoma and breast cancer are caused by Human Endogenous Retroviruses (HERVs) [78, 79], Other examples include Human Mammary Tumor Virus (HMTV) causing cancer of chests in female [80] while myeloma, lung cancer, gastrointestinal cancer and breast cancer can also be caused by Torque Teno Virus (TTV) [81].

2.3 RNA Tumor Viruses

All cancer causing viruses which have RNA genome belong to type retroviruses [82]. RNA tumor viruses were named after the discovery of Rous Sarcoma Virus (RSV) particles which contain genetic material in the form of RNA, in 1961. 30 or more different oncogenes have been defined in retroviruses [60, 68].

The RNA genome of Retroviruses has 3 major genetic sequences named env, gag and pol, from these genes structural proteins are synthesized along with virion-associated enzymes, and glycoproteins of envelope. There also have been found an extra non structural gene termed v-onc in complexed retroviruses like lentiviruses which is responsible for the transformation of the cell [83]. In the RSV this extra

gene is v-scr gene which causes development of sarcoma. This gene of cellular origin is gained by RSV after the infection of cells [84].

A variety of oncogenic mechanisms are used in tumor development, by RNA tumor viruses. Few of them code for oncogenic proteins, identical to the proteins of cell involved in controlling cellular growth. Excessive formation of these oncogenic substances or transformation in their functions triggers cellular proliferation. Increased formation of tumor cells and their spread occurs by these type of viral particle. The second mechanism for initiating cellular transformation involves integration into sequences enhancing virus interaction, and sequences promoting viruses propagation in vicinity of cell proliferation promoting genetic sequence, this mechanism is exhibited by another group of retroviruses. The third mechanism by which RNA tumor viruses cause cancer involves encoding of a protein name tax which causes transactivation of the expression of cellular genes [55, 85]. The RNA genome of retroviruses, either simple or complex, is converted into its DNA copy by the process of reverse transcription, right after entering into the host cell, then the host genetic material is integrated by the DNA of virus and undergoes expression by using the host cellular machinery. After integration the viral genome becomes a permanent part of host genome. This permanent combination of viral and host genome found to be the major event leading to oncogenesis. After the insertion of DNA of virus within chromosome of host cell, regulator genes for cell start controlling it and it becomes a part of host cell without any harmful impact. When cells carrying a viral particle get exposed to some mutation causing or cancer causing factors (irradiation, mutagenic, or cancer causing chemical compounds; triggering of hormones and immune response; etc.), viral particle gets active with its proliferation beginning [86–88]. Recently, Dong et al claimed that they discovered Xenotropic murine leukemia virus-related virus (XMRV) as a new human retrovirus associated with prostate cancer. XMRV was isolated from prostate cancer tissue from patients homozygous for reduced enzyme activity of RNase L due to a single amino acid substitution and is susceptible to inhibition by interferon [89]. However, recent reports from other groups appear controversial for a possible role of XMRV in prostate cancer and in chronic fatigue syndrome prostate cancer [90].

In contrast there is another group of retoviruses which are known as exogenous retroviruses, these viruses infect their host by horizontal spread. These viral particles are found only in infected cells due to some activity like change in DNA sequence or formation of new combinations of genes which could be caused due to some environmental exposure [85, 86]. Whereas endogenous retrovirus genetic material is found in the DNA of all cells [83]. On the basis of time period taken by viruses to form tumors, as seen in experiments on different animals, there are two categories of retroviruses.1. Acute transforming viruses, and 2. Chronic transforming viruses. The former develop tumors in duration of days after their introduction in host, while the latter take time period of months to develop cancers. Acute transforming retroviruses also change growing cell in the neoplastic phenotypes [84].

The advancement of assays to study transformation of viruses outside the living cell opened the door of knowledge about the modern tumor virology, the first virus studied by this technique was Rous Sarcoma Virus (RSV). Life cycle of retroviruses was studied at genetic level along with the information about the changes brought in cell culture by viral particles [91].

2.4 DNA Tumor Viruses

A variety of structures, highly organized genome, and different techniques of replication are the characteristics which make the human DNA tumor viruses a group of different properties. HPV, KSHV, EBV and HBV are those DNA viruses which cause tumor formation within the cells and organisms in which they are propagating, while in contrast some DNA containing viral particles have been found to cause malignancies in only the experimental living cells, for example adenoviruses [92]. Notwithstanding clarifying the cause of a few human maladies, examination of the DNA tumor infection cancer causing proteins uncovered instruments taking control of cell development in mammals, eventually prompting towards disclosing cell cancerous growth silencer qualities. Despite of this that these viruses have evolved differently, all of them have a common feature of transforming host cells, which reflects their vital requirement of host replication machinery for replication of their own genome efficiently [63, 93, 94].

Cells of host are either permissive to DNA of DNA tumor viruses or non permissive. Cells' break down and ultimately death occurs in permissive cells by viral replication. While the chromosomes of non permissive cells are usually integrated by the viral DNA into the different sites. Different binding proteins are encoded by viral DNA also resulting in inactivation of cell growth. This leads to the expression of viral and cellular DNA synthesis controlling proteins, which causes the cell transformation [55, 95, 96].

TABLE 2.1: The list of different cancers caused by viruses with their respective oncoviral protein [33]

Cancer	**Oncovirus**	**Oncoviral Protein**
Penile cancer	Human papillomavirus types 16 and 18 (HPV16,18)	Early protein 6 (E6), early protein 7 (E7)
Vulvar cancer	Human papillomavirus types 16 and 18 (HPV16,18)	Early protein 6 (E6), early protein 7 (E7)
e Cervical cancer	Human papillomavirus types 16 and 18 (HPV16,18)	Early protein 6 (E6), early protein 7 (E7)
Kaposi's sarcoma	Kaposi's sarcoma-associated herpesvirus (HHV-8)	Viral interferon regulatory factors-1,2,3 (v-IRF-1, 2, 3)
Hodgkin Lymphoma	Epstein-Barr virus (EBV)	Epstein-Barr nuclear antigen (EBNA) latent membrane proteins (LMP)
Burkitt Lymphoma	Epstein-Barr virus (EBV)	Epstein-Barr nuclear antigen (EBNA) latent membran proteins (LMP)
Nasopharyngeal Cancer	Epstein-Barr virus (EBV)	Epstein-Barr nuclear antigen (EBNA) latent membrane proteins (LMP)
Malignant Pleural Mesothelioma	Simian virus 40 (SV40)	SV40 Large T-antigen

Hepatocellular carcinoma	Hepatitis B virus(HBV), Hepatitis C virus (HCV)	Hepatitis B viral protein (HbcAg)
Adult T-cell leukemia	Human T-lymphotropic virus 1 (HTLV-1)	Tax
Squamous cell carcinoma	Human papillomavirus types 8,18,5 (HPV8,18,5)	Not clear yet

2.5 History of Human Papillomaviruses (HPV)

Epithelial cells of mammals and birds are mostly infected by species-specific Papillomaviruses (PV) [13]. Induction of malignant changes and cancerous growth was discovered in the skin of diseased rabbits due to cottontail rabbit papillomavirus (CRPV), by Francis in 1935. The involvement of virus in the development of cancer in mammalian species was proved by this study. The Nobel Prize was awarded to Rous [15].

Despite of failure in finding the herpes simplex virus in samples from cancerous cervix, the affiliation of human papilloma virus with cervical cancer was postulated in 1975, by zur Hausen. After sometime later, HPV 18 was identified in tumor of the cervix and HPV type 16 was identified in carcinoma of the squamous cells by his research group [91, 93].

This was strengthened more by Nordic longitudinal epidemiologic studies which showed that the women infected with HPV developed cervical cancer 10 to 20 years earlier as compared to the women which were not infected [97]. Finally, after the emergence of the concept, it was proved that cervical cancer is protected by prophylactic HPV 16/18 vaccination. In 2006-2007 HPV serology and prophylactic HPV vaccines were licensed for the first time [98]. These developments were made possible after discovering by the scientists that the assembling of viruses occurs majorly by structural L1 protein of HPV, just like the hepatitis B-virus surface [99].

2.6 Papillomaviruses: Classification, Structure and Replication

Genome of PVs consists of two strands of circular DNA, surrounded by a capsid of icosahederal structure. The diameter of the viral particle is about 55 nm with almost 8000 base pairs in genetic material (DNA). There are two structural proteins L1 and L2, which are comprised of 72 capsomers to form capsid or protein shell of the viral particle [13, 99]. The major structural protein of the virus is L1 which forms 80% of the shell coat. Papillomaviruses have been declared as a member of a distinct viruses family from ICTV (International Committee on the Taxonomy of Viruses), named by Papillomaviridae family. The classification of papillomaviruses into types, species and genera is based on comparison of L1 gene. Phylogenetically related species which are included in the genera, are identified by using Greek letters. These species are biologically diverse from each other. Papillomaviruses are grouped at this level in 16 genera. The HPV types affiliated with mucosal and genital infections belong to the genus alpha-papillomaviruses. Papillomaviruses having 60-70% similarity are grouped in identical group [13, 99, 100].

15 species have been reported among genus alpha-papillomaviruses, on the basis of their genomic sequences. A7 and A9 are the groups of species which, are categorized as high risk (hr) HPV types due to their involvement in malignancy of cervix. The most common hrHPV belong to the A9 species that is HPV 16 and A7 species that is HPV 18 of the A7. Complete isolation and sequencing has been done of about 120 different types of HPV. The criteria for grouping of HPV types is on the basis of permanent sequences present in the open reading frame of L1 [13, 99]. In approximation, 40 different HPV infect mucosal epithelial layer from the 120 types discovered. On the basis of the ability of these HPVs to cause malignancy, viruses are grouped into low-risk (lr) and high-risk types. The genome of Human Papilloma Viruses codes for two late structural proteins (L1 and L2) and six early proteins (E1, E2, E4, E5, E6 and E7), having eight open reading frames. Additionally, there is also present a long control region which don't code for some protein [101]. About half of genome is covered by the early genes which codes for the early proteins which are vital for duplication of viral DNA by E1, synthesis of

RNA by E2 conversion by E5, E6 and E7 and virus genome synthesis associated with changes in structure of cells of host by E4. Synthesis and transcription of viral DNA in host cells is not possible without the regulatory proteins E1 and E2. The cells involved in viral genome amplification have been found to show high levels of expression of HPV E4 protein. Start and advancement of cellular conversion, occur due to cancer causing genes, which code for proteins of E5, E6 and E7 proteins in high-risk HPVs. These proteins are called oncoproteins [47].

2.7 HPV Proteins

The procedure of the interpretation and interpretation of HPV proteins is an entangled and a not yet completely comprehended process. Enactment of different promoters and option joining are imperative so as to ensure the right interpretation of viral proteins required amid diverse stages on the infection life cycle. The proteins that are communicated from the eight HPV ORF's can be assembled by the parts that they take part in the viral life cycle [102].

2.7.1 The Replication Proteins

The replicative dimeric origin-binding protein E2 and helicase E1 are the proteins encoded by viral DNA in order to support PV DNA synthesis. HPV borrows all other replication proteins and enzymes from the host cell duplication machinery, which are required for HPV DNA synthesis [103]

2.7.1.1 E1

Essential factor without which synthesis of genetic material of virus cannot occur is the 68kDa E1 protein which makes an important papillomavirus replication factor. Among the ORFs of papilloomaviruses the most conserved sequence is that of the E1 ORF [103]. It is the viral helicase with being a member of class of ATPases associated with diverse cellular activities (AAA+) helicases. There has been found sequence homology between the E1 helicase and identical non structural motifs with other AAA+ helicases like the Large T Antigen (LT) of SV40 virus [104]. There are 4 domains of the E1 protein including an N terminal domain, a DNA binding domain, an oligomerization domain and a C terminal helicase

domain. HPV duplication essentially needs the DNA binding, oligomerization and helicase domains of E1 [105]. As far as, DNA binding activity is concerned, it is minimally possessed by E1 alone however, the rate of formation of E1-DNA complexes increases with the presence of E2. E2 is the viral helicase loading factor. Interaction of E1 with the E1BS is essential requirement for viral origin of replication [106].

2.7.1.2 E2

The E2 proteins of papillomavirus function as dimers within the virus life cycle. The structure of E2 is prototypic of a transcription factor as it is divided into 3 functional domains. The N terminal transactivation domain is well conserved and it is within this domain that the multi-functional E2 protein can regulate viral transcription and replication [102]. The central hinge region varies in sequence size and is not well conserved across varying HPV types. Little is known about the functions of this region but it has been shown as having a role in E2 protein turnover and nuclear localisation. The carboxylic end region is important for dimerization of the E2 protein, DNA binding at 12bp palindromic sites, interaction with the viral helicase E1 and interaction with the transcriptional protein TBP [107].

E2 is involved in the regulation of viral DNA replication through its association with E1. E2 interacts with DNA as a functional dimer at E2 binding sites within the viral LCR. The DNA binding domain of the E2 proteins forms a dimeric structure that positions two alpha helices to recognize and contact the E2 binding site [107].

2.7.1.3 E1 ^ E4

The E4 protein is expressed from a spliced mRNA species as E1 ^ E4. Spliced transcripts are formed between the N terminal of the E1 ORF and an almost complete E4 ORF. E1 ^ E4 accumulates to high levels at mid epithelial cells. The E1 ^ E4 roles in the viral life cycle are a little unclear, however when this protein

is lost it results in severe and adverse effects on multiplication of virus genetic material HPV types 16, 18 and 31 [108–110]. The cottontail rabbit papillomavirus, E1 ^ E4 is essential for completion of the vegetative phase in virus life cycle [111]. Therefore one of the key roles for the E1 ^ E4 protein in the upper epithelial layers appears to support vegetative replication. E1 ^ E4 may be involved in vegetative replication by inhibiting mitotic progression and inducing DNA re-replication [112].

2.7.1.4 E5

HPVs are known to cause long term persistent infections. Viral persistence is caused by a number of factors. The papillomavirus growth and events of propagation occur exclusively in cells of epithelium layer and escaped from the immune system components acting in dermal layer. PVs due to their inability to cause break down of cells, are also unable to activate immune cells against themselves. HPVs are known to cause long term persistent infections. Viral persistence is caused by a number of factors. Proteins of viruses decrease the activity of immune response of host cellular machinery. The hydrophobic membrane bound HPV E5 protein contributes to viral persistence by down-regulating the host cell immune defence. The Major Histocompatibility Complex class I [MHC class I] expression increases presentation of viral antigens to cytotoxic T cells. E5 protein down regulates MHC class I in HPV infected cells [113, 114]. This is achieved by recruiting the MHC complex to golgi bodies as opposed to exterior of cell where it would be recognised by CD8+ T cells [115, 116]. The accumulation of the MHC in the golgi is a consequence of E5 induced alkalinisation of the golgi membrane [117, 118]. Malignant transformation by HPV can also be attributed to the E5 protein. The role of HPV16 E5 has been suggested mostly in early stage of cervical carcinogenesis. Evaluation of E5 expression levels by immunohistochemistry has shown that E5 is detectable in 80% of HPV infected low grade squamous intraepithelial lesions and that this percentage drops to around 60% in cervical carcinoma tissue. Interaction of epidermal growth factor receptor [EGFR] signalling pathway by E5 give another proof of its activity as a protein able to transform cell [119–121].

2.7.1.5 E6

The previously described interaction between E7 and Rb family of proteins may result in restricted flourishment of cells and cell death via the activation of p53 depending pathways [122]. To counter this, HPV expresses the E6 protein. E6 interferes with p53 functions by multiple mechanisms. E6 shifts the E3-ubiquitin ligase E6 associated protein [E6AP] and results in ubiquitination with consequential break down of proteasome by p53 [122]. E6 can also alter p53 function by decreasing its stability. The histone acetylases CREB binding protein [CBP] and p300have been shown to increase p53 stability through acetylation. HPV16 E6 binds to p300 and CBP therefore inhibiting their ability to stabilise p53 [122–124]. E6 contributes to cell immortalisation through the activation of telomerase. Telomerase enzyme is frequently activated in many cancerous diseases and necessary part of synthesis of genetic material on the end of DNA strands. HPV16 E6 can actuate the formation of transcripts of telomerase reverse transcriptase [TERT] and increases its protein levels through associating to nuclear transcripting factor X-123 [125].

2.7.1.6 E7

The 17kDa HPV16 E7 protein is located primarily in the nucleus. Within the amino terminal of E7, there is a conserved LXCXE motif. The motif is the site where E7 interacts to Rb tumourous cell suppressing factor protein alongwith its relative proteins p130 and p107 and targets them for degradation [126]. 'High risk' protein of E7 attaches with the Rb family members through much more potential as compared to 'low risk' HPV's proteins. This may contribute to their oncogenicity [127]. E2F class of factors responsible for transcripting DNA is regulated by the Rb type of proteins controling G1-S phase turnover of the cell cycle [128]. The eight members of the E2F family are used for regulation of transcription of genetic sequences responsible for advancement of cell cycle, differentiation, mitosis and apoptosis. Therefore these functions are controlled by genetic sequences in which promoters have attachment points for E2F. In non HPV infected cells, Rb proteins inhibit process of transcription of E2F depending promoting sequences

through direct mechanism of interacting with the domain of E2F transactivation [129]. In HPV infected cells however, the E7 protein disrupts Rb-E2F interaction by binding to Rb protein. This interaction leads to the synthesis of E2F responsing proteins like cyclin A and cyclin E and consequently facilitates immature progression of S phase and synthesis of genetic material [130].

2.8 The Capsid Proteins: L1 & L2

The 55nm diameter HPV coat is un enveloped, has T=7 icosahedral symmetry and is synthesized from two HPV structural proteins, L1 and L2. Within the capsid is the 7.9kb circular genome that is packed with the histone proteins of cell to form compact packed DNA [131]. There are 72 pentamers of L1 that make up the viral coat and they are associated with L2. The stoichiometry of L1:L2 in purified L1+L2 complexes is 5:1 indicating that a single molecule of L2 interacts with an L1 pentamer [132]. The capsid proteins are essential for viral entry, and are only expressed in terminally differentiated keratinocytes demonstrating their role in capsid formation and egress [99]. The L1 protein is evolutionary conserved. Interaction between the viral capsid and the cell surface receptor relies primarily on L1.

The L1 protein is evolutionary conserved. Interaction between the viral capsid and the cell surface receptor relies primarily on L1. There has been speculation regarding what cell surface receptor L1 interacts with to fulfill capsid attachment to the cell. Glycosaminoglycans (GAG's) are the suggested initial attachment receptors, in particular heparin sulphate [133–136]. Another potential candidate for L1 attachment is syndecan1. Heparin sulphate proteoglycans (HSPG) are frequently located on cell surfaces and in the extracellular matrix. Most dominating HSPG in cells of epithelium layer are syndecans. Syndecan1 may therefore act as first attaching recepting molecule as it is synthesized in cells in large concentration [137, 138].

L1 expressed alone is sufficient to produce synthetic papillomavirus particles. Formation of non-infectious viral like particles is highly immunogenic. This is why L1 is used as a preventive vaccination against HPV types [99].

2.9 Viral Life Cycle

Papillomaviruses are flawlessly adjusted with regular host tissues, skin and mucosa, and phases for the viral life cycle are set by sense of differentiating these [131]. The body openings are lined by columnar cells at internal epithelia, while the proximal epithelia are composed of stratified squamous cells. These structures are seperated by the squamo-columnar junction, composed of keratinocytes which form a metaplastic transformation zone and are in rapid cell synthesizing phase. Such structures can be found in the cervix, larynx, urethra, nasal sinuses, and anal/rectal junction. Due to absence of protective layer of differentiated or desquamating cells metaplastic epithelial cells are particularly sensitive to infection by papillomavirus [139].

Papillomavirus life cycle starts when the infection contaminates the layer of basal cell of epithelial region. For permission to infection for entering the basal cells there is a proposal of requirement of a microwound in the stratified epithelium. Undifferentiated cells and parabasal cells or travel enhancing cells are members of basal cells, belong to group of main multiplying cells in ordinary epithelial layer [140]. For creating a set up of some disease it has likewise been proposed that a papillomavirus must contaminate an epithelial foundational microorganism [131, 141].

Cell has heparan sulphate components of ECM which act as cell surface receptors and that are responsible for the initial attachment by virus to the cell of host [134]. Protein through their FG- and HI loops containing specific amino acids like lysine, are mainly responsible for the contact between extracellular matrix and L1 of HPV16 specifically. In the antibody-mediated neutralization of the viral particle, blocking of this interaction is very important. The mechanisms of a caveolin-mediated pathway or endocytosis of clathrin-coated vesicles are used for entrance of virus into cell slowly, according to the HPV type [142, 143].

Viral genome is established as an extrachromosomal element or episome after migration into the nucleus inside the cell. Right after infection episome of virus genome gets multiplied and makes almost 100 replicas of virus in a cell. In basal

and parabasal cells viral DNA synthesis is accomplished by E1 and E2 proteins. In infection with HPV 16 , E1 is not needed for the maintenance of productive replication but only in initial replication at replication origins [144]. Viral genome cannot be imported to nucleus without mitotic activity of stem. Papilloma viruses for the replication and synthesis of their genome in the host are dependent on host cellular machinery because the HPV genome do not have the genes which could synthesize proteins required for its replication . Different functions of the cell are allowed by activation of different genes in the different epithelial cell layers. The transit-amplifying cells are divided and differentiated. Transit-amplifying cells reach terminal differentiation after dividing for several months as shown in Fig 2.3. Stage is set for HPV duplication by continuous migration of transit amplifying newly born cells, which span the whole layer of epithelial cells undergoing differentiation. Replication of viral genome to a raised copy number takes place after the synthesis of structural viral L1 and L2 proteins and E4 protein occuring in the cells undergoing differentiation in the central and upper layers of epithelium [139].

After the completion of the duplication of viral DNA the structural proteins L1 and L2 are synthesized within the epithelium. In terminally differentiated cells packaging of genomes occurs into self-assembling capsids. Travel of infected cells to epithelia cells and death are the requirements for the cytolysis and release of virions to occur by the infection.

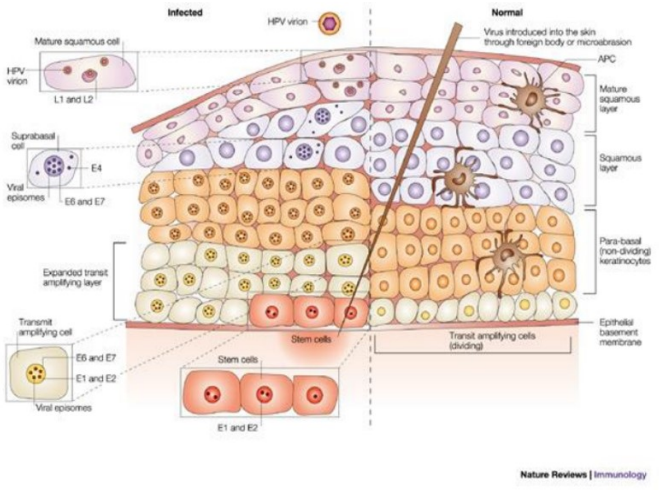


FIGURE 2.3: Steps of life cycle of HPV modified from Frazer 2004 [145]

The duration of the life cycle of papillomaviruses is 2-3 weeks, which starts from a normal epithelial cell undergoing differentiation and ending at keratinized cell from a basal cell [146]. Immune evasion which enables the viral particle to escape for long time, is responsible for persistent HPV infection and the cancer development as shown in Fig. 2.4. Host's immune system is inactivated till the virus enters and causes inflammation. Even after the cure of infection, viral genomes stay there at the place of infection within the basal cells silently like latent phase. Immunosuppression, pregnancy and aging are the conditions in which Reactivation of silent HPV infection occurs [141, 147, 148].

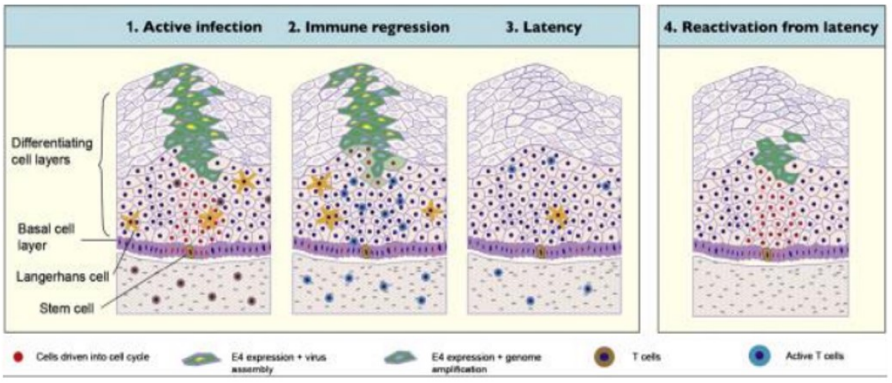


FIGURE 2.4: A model of persistent papillomavirus infection and escape from immune surveillance modified from Meglannon et al, 2011 [141]

2.10 Late-Phase of Papillomavirus Lifecycle

RNA-splicing mechanisms are regulated by cells during keratinocyte differentiation [149–152]. Alternative-splicing mechanisms are utilized by HPV during the whole period of the differentiation program in infected cell [151] and regulation of the polyadenylation machinery promoting the switch from early viral mRNA transcription (ex. E6, E7, E1, E2, E4 & E5 transcripts) to late viral mRNA transcription (ex. E2, E1 ^ E4, L1 & L2 transcripts)is facilitated. The transition from the early-phase to the late-phase of the HPV lifecycle leads to the synthesis of the capsid proteins [L1 and L2], virus formation, and release of infectious virus particles which are at different stages of the advancement from early to late transcription. [152, 153]. Positive immuno-reactivity with antibodies specific for

the presence of the capsid proteins, L1 and L2 can be used to detect the conversion between the early and late HPV infections in the terminally differentiated keratinocytes [154].

2.11 Papillomavirus Capsid Synthesis, Structure and Function

HPV16 encapsidation occurs within the nucleus near PML sites inside the nucleus that immuno-reacts with ND-10 antibodies [155]. L1 and L2 being the major and minor capsid proteins, respectively, are synthesized by alternative-splicing mechanism, through which capsid-ORF transcription is controlled. These are used for packaging of the viral genome. [151, 156–158]. Cellular heparin sulfate-proteoglycan [HSPG] receptors or alpha-6 integrins depending on the type of cell, is utilized by the major capsid protein (L1) for entrance of viral particle, during an initial infection [159, 160]. Endosomal cellular entry is facilitated through the interaction of the L1 with HSPG followed by capsid re-organization to expose the amino-terminus of L2 to cleavage by furin, respectively, as a result of attachment of viral particle at cell surface [159–161]. The entering viral particle is enclosed in a membrane vesicle for transportation to the nucleus along with the degradation of L1, leaving behind L2-to-HPV16 DNA complex [160, 161]. The L2 performs the function of protection of the HPV16 genome from damage and that of guidance of DNA to the nuclear destination, resulting in nuclear-translocation, and ultimately developing the infection [162, 163]. The structures of L1 and L2 show the possibility of involvement of the non-structured terminal-amino acids in non-specific DNA capsulation due to the amino acid sequences which are conserved in these regions in comparison with other non-specific DNA-BPs [164–166].

2.12 Papillomavirus Variable Regions or Surface Loops

There are five conformational immunoreactive epitope regions at the surface of the capsids of HPVs. These epitopes vary among each HPV-type and amongst the PVs capable of infecting other species [for ex. Bovine-PV and Canine Oral-PV]. All PVs

are characterized by the presence of a conformational-dominant immuno-reactive epitope region, which are formed by the oligomerization of five monomer units of L1 protein and is composed of the DE-and FG-loops on L1, also located inside the hollow pentameric-shaft space [99]. There is present a connection between the L1x5 molecules through intra-strand attachments, formed by the terminal amino- and carboxyl- amino acids to create arm like structure protruding and interacting with the adjacent L1 protein, resulting in pantamerization and formation of capsid [167].

2.13 Cervical Cancer: Burden and Etiology

Cervical cancer is the second most common cancer in women all over the world [168]. Annually about half a million women are diagnosed with this type of cancer, among which about 11,300 are found to be the residents of the US. In fact, about 250,000 women die worldwide in a year, including 4,000 Americans [169, 170]. Despite the fact that screening has cut the occurrence and mortality of cervical disease down the middle in the course of recent years, the financial and societal weight stays high. New cases keep on arising because of misdiagnosis, absence of patient follow-up, or particularly forceful tumors that avoid screenings (WHO). It is evaluated that 1.7 billion dollars are being spent yearly in the US for treatment of cervical tumor [171].

The acknowledgment of these weights prompted to the advancement of a cervical growth immunization. Not at all like most different malignancies which are created by the aggregation of irregular changes over numerous years, cervical disease regularly has a one of a kind cause: the human papillomavirus (HPV). HPV is in charge of 99.7% of every cervical disease, making it the objective for most screening and restorative methodologies [172]. One late advancement in the battle against HPV-prompted cervical tumor is the cervical growth immunization. This antibody comprises of the HPV auxiliary external capsid protein L1, which self-collects into infection like particles (VLPs). These VLPs can initiate a high titer of killing antibodies in both people and creatures, in this way forestalling future viral contamination [173].

Be that as it may, the presence of a HPV antibody does not mean a conclusion to cervical tumor or to different genuine HPV pathologies, for example, warts. There are many sorts of human papillomavirus, and the antibody does not ensure against every one of them. Also, the cost of the immunization makes it unattainable for all [173]. At long last, HPV is a sexually transmitted disease (STD). Since the antibody is precaution just, to be viable it must be managed preceding any presentation [174].

Hence, the investigation of HPV and HPV-instigated cervical disease keeps on being vital. Improvement of second era and third era antibodies are in progress. Objectives incorporate building immunizations that are defensive crosswise over many sorts of HPVs and antibodies that can be delivered inexpensively and along these lines conveyed overall [175, 176]. Antibodies went for keeping future malignancies creating from past diseases, and restorative immunizations against flow tumors, would permit inoculation of more established grown-ups and limit the moral concerns right now included [173].

The clinical enthusiasm for HPV and the look for immunizations has uncovered many inquiries yet unanswered about HPV-actuated cervical disease. These points incorporate ranges as endlessly differing as the essential life cycle of HPV, the part of different HPV proteins in subverting the host framework, have invulnerable reaction to viral contamination, have cell reaction to DNA harm, and eventually, the loss of cell cycle control and resulting advancement of tumor.

2.14 Interaction of Viral Proteins with Human Proteins

Fascinating diversity is being shown by viruses in composition, shape, size, tropism, and pathogenesis. Helical or icosahederal capsids form the core of infectious viral particles. They may be non enveloped or enveloped with lipids and proteins. Genetic material of viruses occurs in the form of single stranded or double stranded RNA or DNA. Replication of their genome inside the host cell cytosol or nucleus,

plays key role in viral propagation. For example, alphaviruses like Sindbis virus (SINV) replicate in the cytoplasm. Conversely, herpesviruses, such as herpes simplex virus-1 (HSV-1), and the retrovirus, human immunodeficiency virus (HIV), duplicate in the nucleus. Capacity of viruses to infect new cell types and species has increased due to its continuous coevolution with hosts, despite of having a distinctive tropism for cell types and host types.

The key role in the virus replication and spread is played by the formation of interactions between virus–host protein. One evidence is the remarkable difference between protein banks of host cells and viruses. About 3 million proteins per cubic micron are generated from 20,000 protein encoding genes, in humans [177, 178]. This diversity increases manyfold by different regulatory processes, like the presence of multiple transcription initiation sites [179], alternative splicing [180–183], alternative mRNA polyadenylation [184], pre-mRNA editing [175], and posttranslational modifications [185, 186]. 60% of human genome is regulated by alternative splicing [187, 188]. Proteome is made more complex and highly variable by a variety of posttranslational modifications [189].

Some of these regulatory processes are also used by viruses to expand viral proteome. Contrary to human proteins viral DNA code for less proteins, that is upto 2500 in approximation, usually present in the form of hundreds to thousands of copies in a single viral particle [190, 191]. Despite of coding for a lesser number of proteins virus can undergo successful replication by interacting with host proteins. Protein-protein interactions mediate the virus host relationship, on which evolutionary conserved life cycles of viruses, mainly rely. The attachment and entry of the virion particle into the host cell, viral translation of mRNA by host ribosomes, viral genome replication, assembly of viral particles enclosing the genome, and release of infectious particles from the cell are the stages of a complete infectious life cycle, independent of the virus classification. For example, HSV-1 is a DNA virus that undergoes a series of protein synthesis and associations with the host. Blockage of host defence mechanism and successful replication are the aims of these virus–host protein interactions. The process initiates by the attachment of viral particle with the host cell through interaction between a host cell surface protein

receptors and glycoproteins of virus. Viral capsid is released alongwith the DNA and matrix proteins inside the cell as a result of the attachment of virion with plasma membrane of the host cell. viral capsids navigate the cytoplasm, stay at nuclear pores, and expel the twofold stranded viral DNA genome into the nucleus through relationship with cellular motor machinery [192].

Regulation of viral genome expression alongwith the inhibition of cellular intrinsic and innate immune response occurs by transportation of many incoming viral proteins to different subcellular destinations [193, 194]. Expression of immediate-early, early, and late gene is organized by initiation of a highly regulated viral gene expression cascade, inside the nucleus. Immediate-early mRNAs of virus is transcribed by co-operation of host cell RNA polymerase. After their synthesis in cytosol, immediate-early proteins are recruited to nucleus for stimulation of transcription of early gene, resulting ultimately in synthesis of late gene products, in order to complete viral assembly. For transforming into a new completely infectious agent, the immature viral particle travels through nuclear membranes, Golgi compartments, endosomes, and cell membranes to acquire matrix and envelope components, after packaging of viral DNA with capsids in the nucleus. Progression of viral life cycle alongwith the modulation of host antiviral defenses, are mainly accomplished by the establishment of temporally and spatially regulated virus–host and host–host protein interactions. Notwithstanding giving unthinking bits of knowledge into the science of a contamination, information of transient virus–host associations can likewise uncover viral or host components that can be focused in antiviral therapeutics. Viral pathogens, for example, HIV, Ebola, flu, and hepatitis C infection speak to critical dangers to human wellbeing, yet useful antiviral medications are not promptly accessible for a large portion of them. Immunizations or viable medications are additionally missing against other critical human pathogens that trigger long lasting diseases and are worldwide human medical problems, for example, human cytomegalovirus (HCMV) and Dengue infection. The ID and portrayal of virus–host protein associations can indicate fundamental occasions required for viral passage, replication, or spread, which can be utilized as new roads for antiviral therapeutics to foresee, anticipate, or treat infection instigated distresses [33, 195].

2.15 Identification and Prediction of Human Proteins Involved in Carcinogenesis

Finding the human proteins essentially like viral oncoproteins prompts to a classification of the malignancy related pathways that are as of now not obviously known. Distinctive sorts of infection brought on growths have been investigated in view of their comparability keeping in mind the end goal to clear up the obscure tumor instruments. Thus, a few potential tumor pathways that might be huge and basic in oncogenic malignancy process were discovered, which will be useful for further review on disease instruments and the advancement of new medication targets [33].

The structure and progression of infections are an entrancing examination subject from an organic, as well as from a physical viewpoint [196]. Specifically, they are an exceptionally informational model framework to study self-gathering of extensive protein buildings with a moderately clear natural capacity. One key stride amid the replication procedure is the arrangement of the protein shell containing the viral genome. For some infections, the capsid arrangement is adequately self-sufficient that it happens even in vitro. This heartiness of the procedure ensures effective replication inside the dynamic and heterogeneous environment of a living cell. Despite the fact that infection shell development is considered as a worldview for the self gathering of protein edifices [117], its fundamental standards are a long way from being completely caught on. Advances in our comprehension of infection get together would build our insight into a procedure of extensive organic and therapeutic significance and in addition help to progress new self get together systems in materials science applications [197].

An extensive assortment of scientific models and recreation approaches has been created to pick up understanding into the flow of capsid arrangement from a hypothetical point of view. In these methodologies the attributes of protein affiliation and separation procedures were broken down contingent upon parameters like connection quality, subunit geometry or temperature. The utilized systems run from expensive scale Molecular Dynamics (MD) reproductions with just a humble

measure of coarse-graining of the nuclear subtle elements [198] through different plans of coarse-grained MD [199–204] to inconsistent molecule reenactments with cooperation possibilities [205, 206]. A thermodynamic system for assembling of icosahedral components of capsid has been built up by Zlotnick and associates [207–212]. As a rule, these reviews have uncovered that the development of mature capsids requires highly directed and organized assembling of capsid proteins. On the off chance that collaboration quality is too high (or, proportionately, temperature excessively low), the framework turns out to be dynamically caught in intermediates which can't be reversed. In the event that association quality is too low (or, equally, temperature excessively high), the objective structure is not adequately steady. Another system which can forestall complete capsid arrangement is the event of mavericks, prompting to basic polymorphism as frequently examined with molecular dynamics [213–215]. Because of the vast number of single building pieces collecting amid infection development (the most straightforward icosahedral capsid, T1, has as of now 60 protein components), there is a huge number of topologically conceivable interacting pathways. Like protein collapsing, the predominance of few key structures is accepted to restrain the quantity of pathways and to accelerate the procedure [207]. In this regard it has been observed that some infections have created components to coordinate and interact with the host proteins and preventing them to interact with their counter parts [216].]. This exchanging builds up a chain of events in the arrangement of transient intermediates amid the normal interaction and metabolic process. In various studies, it has been demonstrated that intermediates of pentameric and hexameric symmetry are of unique significance for the assembling procedure of icosahedral capsid proteins [117, 217–221]. Investigations from Brome Mosaic Virus [217], Cowpea Chlorotic Mottle Virus [219], Human Papillomavirus [220] and Simian Virus 40 in vivo and in vitro [221] have reported that capsid protein components assemble from pentameric capsomers. Assembling of the capsid proteins have been studied by molecular dynamics and kinetics techniques. These studies are limited to the patterns of icosahederal symmetry of capsid proteins. The hierarchies of these assembled capsid proteins have not been investigated at residual level with respect to a specific protein among the capsid component proteins [213, 222].

2.16 Elastic Networking Modelling

Biological functions of macromolecules have been shown to be understood in a comprehensive manure successfully, by using ENM-based simulation methods with the use of information regarding basic structure of these macromolecules. The major philosophy and methodology of ENM is that the topological characteristics of a molecule play an important and dominant role in explaining the global as well as collective movements of proteins. A variety of biochemical problems associated with proteins have been solved by the use of coarse-grained ENMs. These problems include generation of transition pathways by proteins, protein complexes functional motions, mechanism of bind with ligand molecules, and low-resolution structural refinement [223].

Elastic Network Models give us physical bits of knowledge: The general topology of the protein assumes a noteworthy part in the mechanical conduct of the protein, suggesting that proteins related by development are relied upon to demonstrate comparable quantitative conduct. Essential research in science and natural chemistry alongside factual mechanical and logical strategies prompts to enhanced transferability and consistency of conduct of proteins. Polles *et al.,* in 2013, made an endeavor to build up a general and effective strategy for distinguishing the fundamental, mechanically stable protein units beginning from the sole contribution of the completely gathered protein capsid. The technique depended on the portrayal of the interior progression of the capsid by method for flexible system models and utilized it to ideally break down the protein shell into hinders that have the qualities expected for certifiable capsid practical units, for example, mechanical security, auxiliary honesty of the constitutive proteins, or little quantities of inequivalent piece sorts and so forth.

The ENMs have proven themselves to be highly useful in representing the global motions for a wide variety of diverse protein structures. Multiple scales can be used to apply ENMs for simulating and modeling the proteins dynamics. In ENM models all atoms of a protein provide finer description of dynamics of proteins. In coarse graining models, every single residue is represented by a single-site, identified by the Cα atoms and bonding is represented by the uniform springs connecting

these atoms or sites. ANM or GNM could be used to describe the dynamics of that interconnected model. In GNM information about the magnitudes of the proteins' structural fluctuations, has been yielded, very successfully but it neither provides any directional information about the protein nor its 3-dimensional (3D) nature of motion. However, in actual framework protein fluctuations are generally anisotropic and directional. On other hand, ANM deliberates the anisotropy of the structure of protein while modeling protein dynamics. Thus, biological functions and the mechanisms lying behind their functionality can be studies more relevantly by computational modeling of proteins through ANM.

Chapter 3

Materials and Methods

Our study comprised of following major steps.

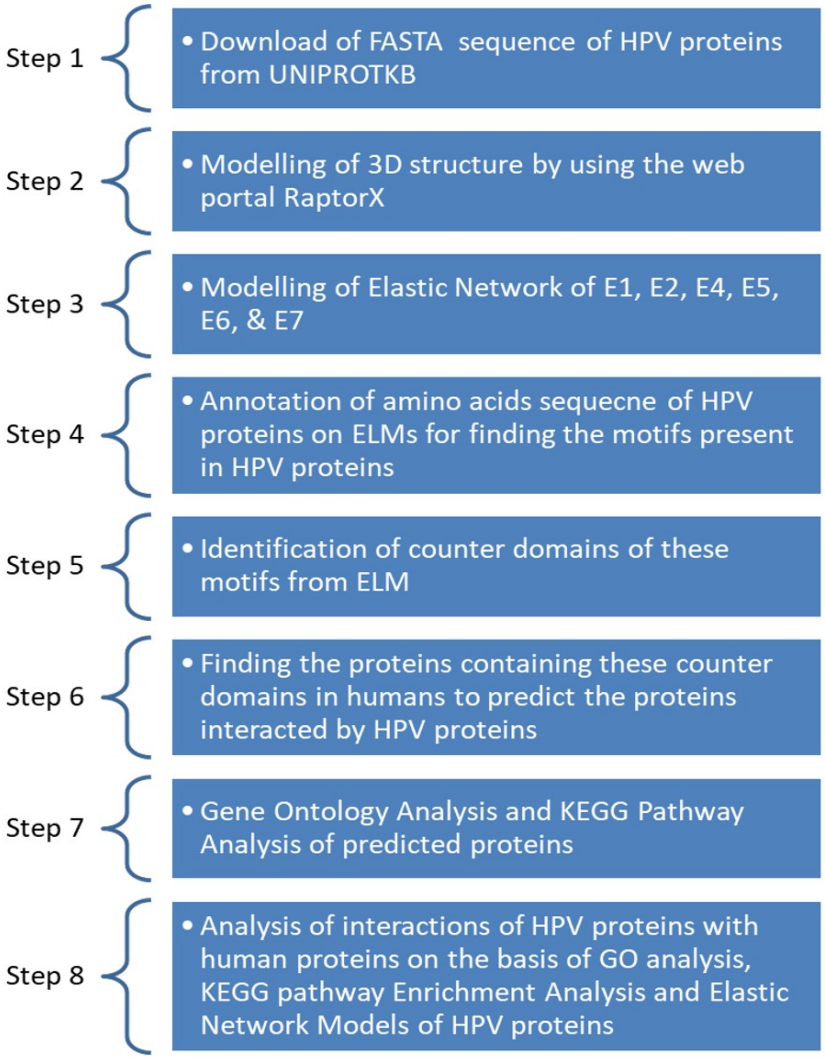


FIGURE 3.1: Steps of Methodology

Interactions between proteins underlie all biological processes. Hence, to fully understand or to control biological processes we need to unravel the principles of protein interactions. The quest for these principles has focused predominantly on the entire interfaces between two interacting proteins. However, it has been shown that only few of the interface residues are essential for the recognition and binding to other proteins. The identification of these residues, commonly referred to as binding "hotspots," is a first step toward understanding the function of proteins and studying their interactions. Experimentally, hotspots could be identified by mutating single residues—an expensive and laborious procedure that is not applicable on a large scale. Here, we show that it is possible to identify protein interaction hotspots computationally on a large scale based on the amino acid sequence of a single protein, without requiring the knowledge of its interaction partner. Our results suggest that most protein complexes are stabilized by similar basic principles. The ability to accurately and efficiently identify hotspots from sequence enables the annotation and analysis of protein–protein interaction hotspots in an entire organism and thus may benefit function prediction and drug development.

3D structures built from FASTA sequences of HPV proteins were used as an input to model the elastic networks of protein. The purpose of modelling was to identify the regions of protein which show maximum flexibility to ensure the involvement of those specific hot spot residues in PPIs. The behavior of these specific regions, comprised of hot spot residues, was analyzed in establishing the interactions and involvement of the PPIs to trigger the progression of viral infection towards cervical cancer.

Different tools have been reported to be used for building 3D structure of proteins like TASSER, MULTICOM, Mu Fold and Pcomb. RaptorX was selected in this study, as this is one of the very few tools for 3D modelling that use a single server.

There are 2 major approaches for elastic network modelling. One is GNM, among the 2 major approaches for elastic network modelling. In GNMs amino acid residues of proteins experience Gaussian distributed fluctuations about their mean positions, also accompanied by harmonic potentials. Different types of amino acids

are not distinguishable, due to which all residues are to be considered to adopt a generic force constant for the interaction potential between all residues pairs close enough. The magnitudes of fluctuations have been verified to be provided by GNM. In the GNM there is no consideration of directions or 3-D characteristics of movements of residues, all fluctuations are un-expressively assumed to be isotropic.

Each residue is represented as N and protein molecule as a cluster of N sites, forming an ensemble of N having 1 independent mode. No involvement of 3D description which would show 3 N modes instead of 1. It is necessary to keep in view the directions of collective movements of residues as they play a major role in biological functioning and mechanisms of interactions of proteins, so regarding the proteins, these fluctuations are anisotropic instead of being isotropic [224–226].

It is not indeed possible to acquire an understanding of the mechanism of motion unless the fluctuation vectors, in addition to their magnitudes, are elucidated. This issue has been addressed by an advanced extension of the GNM, called the anisotropic network model (ANM).

The development of the ANM has occurred as an extension of a recent comparison of the results from MD simulations with the GNM for treating the anisotropy of fluctuations [227].

The tools designed for elastic network modelling include elNemo, ProDy, and ANM2.1. Among these tools ANM2.1 is the tool which alongwith the other attributes of elastic networks also builds animations of models, at the same server, while the other tools only provide the data, which has to be modelled by using some other tool.

3.1 Amino Acid Sequences of HPV proteins

Interactions between proteins underlie all biological processes. The amino acids sequences of HPV proteins were downloaded from UniProtKB (www.uniprote.org) (version 2018) [228], as shown in Table 3.1

TABLE 3.1: Amino Acids Sequences of HPV Proteins

Protein	ID No.	Amino Acids Sequence/Number
E1	UniProtKB P03114 (VE1_HPV16)	sp—P03114—VE1_HPV16 Replication protein E1 OS=Human papillomavirus type 16 OX=333760 GN=E1 PE=1 SV=2 MADPAGTNGEEGTGCNGWFYVEAVVEKKTGDAISDDENENDSDTGE DLVDFIVNDNDYLTQAETETAHALFTAQEAKQHRDAVQVLKRKYLVS PLSDISGCVDNNISPRLKAICIEKQSRAAKRRLFESEDSGYGNTEVETQ QMLQVEGRHETETPCSQYSGGSGGGCSQYSSGSGGEGVSERHTICQT PLTNILNVLKTSNAKAAMLAKFKELYGVSFSELVRPFKSNKSTCCDW CIAAFGLTPSIADSIKTLLQQYCLYLHIQSLACSWGMVVLLLVRYKCG KNRETIEKLLSKLLCVSPMCMMIEPPKLRSTAAALYWYKTGISNISEV YGDTPEWIQRQTVLQHSFNDCTFELSQMVQWAYDNDIVDDSEIAYKY AQLADTNSNASAFLKSNSQAKIVKDCATMCRHYKRAEKKQMSMSQW IKYRCDRVDDGGDWKQIVMFLRYQGVEFMSFLTALKRFLQGIPKKNCI LLYGAANTGKSLFGMSLMKFLQGSVICFVNSKSHFWLQPLADAKIGM LDDATVPCWNYIDDNLRNALDGNLVSMDVKHRPLVQLKCPPLLITSNI NAGTDSRWPYLHNRLVVFTFPNEFPFDENGNPVYELNDKNWKSFFSRT WSRLSLHEDEDKENDGDSLPTFKCVSGQNTNTL (649 Amino Acids)
E2	UniProtKB P03120 (VE2_HPV16)	sp—P03120—VE2_HPV16 Regulatory protein E2 OS=Human papillomavirus type 16 OX=333760 GN=E2 PE=1 SV=1 METLCQRLNVCQDKILTHYENDSTDLRDHIDYWKHMRLECAIYYKA REMGFKHINHQVVPTLAVSKNKALQAIELQLTLETIYNSQYSNEKWT LQDVSLEVYLTAPTGCIKKHGYTVEVQFDGDICNTMHYTNWTHIYIC EEASVTVVEGQVDYYGLYYVHEGIRTYFVQFKDDAEKYSKNKVWE VHAGGQVILCPTSVFSSNEVSSPEIIRQHLANHPAATHTKAVALGTEET QTTIQRPRSEPDTGNPCHTTKLLHRDSVDSAPILTAFNSSHKGRINCNS NTTPIVHLKGDANTLKCLRYRFKKHCTLYTAVSSTWHWTGHNVKHK SAIVTLTYDSEWQRDQFLSQVKIPKTITVSTGFMSI (365 Amino Acids)
E4	UniProtKB P06922 (VE4_HPV16)	sp—P06922—VE4_HPV16 Protein E4 OS=Human papillomavirus type 16 OX=333760 GN=E4 PE=1 SV=2 MADPAAATKYPLLKLLGSTWPTTPPRPIPKPSPWAPKKHRRLSSDQD QSQTPETPATPLSCCTETQWTVLQSSLHLTAHTKDGLTVIVTLHP (92 Amino Acids)
E5	UniProtKB - P06927 (VE5_HPV16)	sp—P06927—VE5_HPV16 Probable protein E5 OS=Human papillomavirus type 16 OX=333760 GN=E5 PE=1 SV=2 MTNLDTASTTLLACFLLCFCVLLCVCLLIRPLLLSVSTYTSLIILVLLL WITAASAFRCFIVYIIFVYIPLFLIHTHARFLIT (83 Amino Acids)
E6	UniProtKB P03126 (VE6_HPV16)	sp—P03126—VE6_HPV16 Protein E6 OS=Human papillomavirus type 16 OX=333760 GN=E6 PE=1 SV=1 MHQKRTAMFQDPQERPRKLPQLCTELQTTIHDIILECVYCKQQLLR REVYDFAFRDLCIVYRDGNPYAVCDKCLKFYSKISEYRHYCYSLYG TTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLDKKQRFHNIRGR WTGRCMSCCRSSRTRRETQL (158 Amino Acids)
E7	UniProtKB P03129 (VE7_HPV16)	sp—P03129—VE7_HPV16 Protein E7 OS=Human papillomavirus type 16 OX=333760 GN=E7 PE=1 SV=1 MHGDTPTLHEYMLDLQPETTDLYCYEQLNDSSEEEDEIDGPAGQA EPDRAHYNIVTFCCKCDSTLRLCVQSTHVDIRTLEDLLMGTLGIVC PICSQKP (98 Amino Acids)

3.2 3D Modelling of HPV E1, E2, E4, E5, E6 & E7 Proteins

3D structures built from FASTA sequences of HPV proteins were used as an input to model the elastic networks of protein. 3D structures of HPV proteins coded by early genes were also constructed by using the web portal RaptorX (http://raptorx.uchicago.edu/) [229], by giving FASTA sequences of HPV proteins downloaded from UniProtKB. These 3D structures were used as input to ANM2.1 [230], for elastic network modelling.

3.3 Elastic Network Modelling of E1, E2, E4, E5, E6 and E7

By using version 2016 of Normal Mode Analysis (NMA) tool ANM 2.1 (anm.csb.pitt) the elastic networks of the HPV protein were modelled. PDB files 3D structure of these proteins were used as input for ANM 2.1.

3.4 Virus Protein ELM Annotation and Conservation

FASTA sequence of E1, E2, E4, E5, E6 and E7 were annotated with sequences of motifs of human by using server ELMs using the ELM resource [231], using default settings except selecting human for the species field.

3.5 Human Protein ELM and CD Annotation

The ELM resource lists Counter Domains (CDs) or proteins known to interact with ELMs. For each ELM conserved on HPV proteins, we found the appropriate CDs. and mapped them to PROSITE domains (https://prosite.expasy.org/). When the ELM resource (http://elm.eu.org/) listed a set of interacting proteins instead of CDs, we assumed that all proteins had a common unknown CD, and annotated them with that. We constructed a list of CDs and interacting proteins for each HPV protein conserved ELM.

3.6 Prediction of Human Proteins Interacting with HPV Proteins

The prediction of HHP, the set of human proteins that might interact with HPV proteins, was based on interactions mediated by ELMs and CDs. By using PROSITE we found the human proteins reported to be containing those CDs.

3.7 Validation using the HPV Databases

2211 human proteins have been retrieved from different databases of Human Papilloma Virus which have been reported to interact with HPV proteins. These

databases include PaVE, HPV Sequence Database (pave.niaid.nih.gov), HPVDB (cvc.dfci.harvard.edu/hpv), Virus Host (www.genome.jp/virushostdb) and HPRD (www.hprd.org) which we called HHE, and used to investigate the usefulness of HHP. We restricted the human proteins interacting with HPV proteins to those belonging to the set of proteins that have PROSITE domains.

The statistics in this research focused on the comparison of our predicted set HHP and the experimental dataset HHE based on the overlap between the two sets, GO molecular function enrichment, and KEGG pathway enrichment. P-values for the overlap between HHP and HHE and their various subsets were calculated using the hypergeometric test in the R Project for Statistical Computing. P-values for GO and KEGG enrichment for a given protein set compared to a background set of proteins were found using Bonferroni corrected p-values from Gene Ontology Resource (http://geneontology.org/) and KEGG Pathway (https://www.genome.jp/kegg/pathway.html) [232, 233].

3.8 Gene Ontology Enrichment Analysis

Gene set enrichment analysis is a common approach to interpret gene expression data that is based on the functional annotation of the genes expressed deferentially. This helps to find the association of differentially expressed genes with a certain molecular function or biological process. The basis for GO annotations is the scientific literature. This analysis is done on the basis of three criteria, description of cellular locations, molecular functions, and biological processes of any gene product. For this purpose we used database The Gene Ontology Resource. The list of genes we found was used as input and we got our output in form of lists of molecular functions, biological processes, in which these genes were involved. These annotations are done on the basis of determinations of direct experimentation and original information is cited by every individual assertion in the Gene Ontology knowledge base. Presently, experimental findings from over 140,000 published papers are included in the GO, represented by more than 600,000 annotations supported by experiments. Over 6 million functional annotations are being provided by this central dataset for additional inference for a diverse set of organisms [232].

3.9 KEGG Pathway Enrichment Analysis

Pathway enrichment analysis helps researchers gain mechanistic insight into gene lists generated from genome-scale (omics) experiments. This method identifies biological pathways that are enriched in a gene list more than would be expected by chance [233]. We used database KEGG for this analysis.

3.10 Tools

The main focus of this computational biology analysis was for prediction of motif and annotated motif data.

3.10.1 UniProtKB

The UniProt Knowledgebase (UniProtKB) is a part of The Universal Protein Resource (UniProt) which is a comprehensive resource for protein sequence and annotation data. The other UniProt databases are the UniProt Archive (UniParc) and the UniProt Reference Clusters (UniRef). It is collaboration between the SIB Swiss Institute of Bioinformatics, the European Bioinformatics Institute (EMBL-EBI), and the Protein Information Resource (PIR).

3.10.2 ELM

The main focus of this computational biology resource is detection and annotation of eukaryotic linear motifs (ELMs) by the help of a repository of an exploratory tool for prediction of motif and annotated motif data. ELMs, or short linear motifs (SLiMs), are compact sites of protein interaction comprised of short regions of amino acids present adjacently. These specific regions are responsible for a variety and versatility of proteins functionality. They play critical roles in regulation of cellular activities and thus are clinically very important, because abnormal functions of SLiMs have been found to be associated with many diseases. SLiMs

have also been reported as targets used by pathogens for manipulation of cellular machinery of their hosts [231].

3.10.3 PROSITE

PROSITE [234] is a database consisting of protein domains and families. The basis of the structure of this database is that, proteins can be grouped into a limited number of families on the basis of their sequences similarities, despite of their existence in nature in an enormous number. Protein domain or proteins having similar properties and belonging to a specific family also have same functional characteristics and usually possess an common ancestor.

Proteins belonging to different families possess some specific regions which remain conserved during evolution and these specific regions play a vital role in functionality of proteins. These regions are also important for the maintenance of three-dimensional structure of proteins. There is a possibility of deriving a signature for a protein domain or family, by analyzing the variable and constant properties of such groups of similar sequences, which make a protein or members of a particular family distinguishable from all other unrelated proteins. An applicable likeness is that the police use fingerprints for identification purposes. A fingerprint is generally sufficient for identification of a specific individual. In the same way s specific protein family can be assigned to a newly sequenced protein by using its protein signature, and consequently its function can also be hypothesized.

Profiles and patterns specific for more than a thousand protein domains or families are present in PROSITE currently. Documentation can also be found about each of these signatures which provides background information on the structure and function of these proteins.

3.10.4 RaptorX

RaptorX is a server for prediction of protein 3D structure developed by Xu group, without any close homology in the Protein Data Bank (PDB). RaptorX predicts

secondary and tertiary structures of a protein, its points of contacts, solvent accessibility, staggered regions and its binding sites, by using amino acid sequence of the protein as input. RaptorX also assigns some confidence scores for indication of the quality of a predicted 3D model: GDT (global distance test), P-value for the relative global quality, and uGDT (un-normalized GDT) for the absolute global quality, along with the modeling error at each amino acid residue.

3.10.5 ANM2.1

ANM (Anisotropic network model) is a simple NMA tool for analysis of vibrational motions in molecular systems introduced in 2000. It uses Elastic Network methodology by representing the system at the residue level. A network or graph of the macromolecule is generated. In the original model each protein node is represented as the Cα atom of an amino acid residue and the harmonic potentials among the interacting nodes are added to give the overall potential of interacting nodes. All interactions are included in the network within a cutoff distance which is the only predetermined parameter of the model. Anisotropic motions are predicted on the basis of the information about the orientation of each interaction with respect to the global coordinates system that is considered within the Force constant matrix. The model is being successfully applied for analyzing the relation between dynamics and function for proteins. Hessian matrix is applied to describe the force constant of the system.

The minimal required input is a coordinates file in a PDB format. There are 2 ways to specify the file:

1. Specification by a 4 letter code PDB ID. The user can choose by a checkbox if to use the putative biological unit supplied by the PDB or the original PDB coordinates.

2. Uploading of a coordinates file to the server from the user computer.

3. There are two parameters in the model which could be controlled by user. The first is the cutoff for defining the maximum interaction range between

Cα atoms. The cutoffs of 15-21 Ågive the best results for globular proteins on average. For specific proteins, like membrane proteins or extended proteins, smaller cutoffs should be preferred. Better results of anisotropy resembling more to the experimental values are obtained by using smaller cutoffs, in the range of 12-15Å[230]. The calculations with larger cutoffs are slower (the Hessian matrix is less sparse). Combining all those reasons the default value of 15Å is chosen for distance cutoff in the ANM server.

4. The distance weighting factor is the second parameter. 2.5 Hz has been indicated the best value for this parameter by large scale statistical analysis. The correlation between experimental and theoretical B-factors are better by 4% comparing to the original ANM from about 55% to about 59% by using this value. The "number of modes" parameter determines how many dominant normal modes could be calculated. More complicated picture is obtained by using greater number of modes on the dynamics of the system at the cost of computational time. Slow modes are usually the functional ones so 20 modes are usually sufficient to describe the important motions of small-medium systems.

Output is obtained in two types of files, text files include slow eigenvalues, slow eigenvectors, fluctuations of residues, components of eigenvectors along all three axis, correlation between B factors, theoretical and experimental B factors, hessian matrix values, and estimated values of spring constant. While coordinate files include coordinates files for anisotropic temperature factors and individual mode fluctuations. Elastic network models and graphs showing the fluctuations of amino acids in 20 different modes for E1, E2, E4, E5, E6 and E7 were used for analysis of interactions of HPV proteins with different proteins.

3.10.6 KEGG (Kyoto Encyclopedia of Genes and Genomes)

KEGG is a database resource for providing information about the biological system's utilities and their functionalities at higher levels. Information could be from molecular-level, obtained by sequencing of genome and other advanced experimental technologies. One of the domain of KEGG is KEGG PATHWAY, that is

a collection of manually drawn pathway maps representing our knowledge of the molecular interaction, reaction and relation networks for Metabolism, processing of Genetic Information, Organismal Systems, Cellular Processes, Drug Development and Human Diseases.

Chapter 4

Results and Analysis

In this chapter, the results are explained stepwise which have been obtained after applying methodologies followed in chapter 3 described above. It is mainly comprised of different sections of research methodology and the summary concluding all the findings. The results of each phase of methodology were evaluated and validated using VirusHost, Gene Ontology Resource and KEGG pathways and from literature source.

This work has been planned according to three specific objectives: (1) the prediction of virus host interactions using sequence motifs, (2) Elastic Network Modeling of HPV proteins and (3) the analysis of interactions between HPV proteins and their targeted host human proteins, on the basis of Elastic Network Models.

4.1 Prediction of Human Proteins Having Interactions with Proteins E1, E2, E4, E5, E6 and E7 of HPV

4.1.1 Eukaryotic Linear Motifs (ELMs)

ELMs or short linear motifs (SLiMs), are compact sites on proteins, involved in interaction, formed by short continuous sequences of amino acids. They are enriched with intrinsically flexible regions of the proteome and provide a broad range

of functionality to the functional proteins. They play important roles in regulation of cellular activities, and are also clinically important, because exceptional SLiM functions have been found to be associated with many diseases and SLiM imitations are often used by pathogens to manipulate cellular machinery of their hosts [235, 236]. Table 4.1 shows the conserved ELMs for each HPV protein, which were obtained by annotation of amino acid sequence of HPV proteins on 133 peptide motifs (ELMs) by using the ELM Resource, which was accessed on March 2018. Overall, 77 of these ELMs present in the ELM resource were conserved on HPV proteins. Some of the ELMs, like LIG_PDZ_Class_1 and LIG_Rb_LxCxE_1 [237, 238] have been experimentally proved as active binding sites for human proteins with HPV proteins E6 and E7, respectively. E1 was having the highest number (28) of ELMs, contrary to .E5 having only 1 motif.

TABLE 4.1: Names of ELMs conserved on HPV proteins. Green color shows the presence of ELM on the respective HPV protein

S. No.	ELMs	E1	E2	E4	E5	E6	E7
1.	CLV_C14_Caspase3-7	■	■				■
2.	CLV_NRD_NRD_1	■		■		■	
3.	CLV_PCSK_FUR_1					■	
4.	CLV_PCSK_PC1ET2_1	■				■	
5.	CLV_PCSK_PC7_1	■					
6.	CLV_PCSK_SKI1_1	■	■			■	
7.	CLV_Separin_Metazoa	■					
8.	DEG_APCC_KENBOX_2	■					
9.	DEG_Kelch_Keap1_1	■					
10.	DEG_Nend_UBRbox_2		■				
11.	DEG-MDM2_SWIB_1	■					
12.	DOC_CKS1_1			■			
13.	DOC_CYCLIN_1	■				■	
14.	DOC_CYCLIN_RxL_1						
15.	DOC_MAPK_FxFP_2	■					
16.	DOC_MAPK_gen_1	■					
17.	DOC_MAPK_MEF2A_6	■		■			■

18.	DOC_PP1_RVXF_1	
19.	DOC_PP2A_B56_1	
20.	DOC_PP2B-LxvP_1	
21.	DOC_USP7_MATH_1	
22.	DOC_USP7_UBL2_3	
23.	DOC_WW_Pin1_4	
24.	LIG_14-3-3_CanoR_1	
25.	LIG_Actin_WH2_2	
26.	LIG_APCC_ABBA_1	
27.	LIG_BIR_II_1	
28.	LIG_BIR_III_1	
29	LIG_Clathr_ClatBox_1	
30.	LIG_eIF4E_1	
31.	LIG_FHA_1	
32.	LIG_FHA_2	
33.	LIG_LIR_Gen_1	
34.	LIG_MYND_2	
35.	LIG_NRBOX	
36.	LIG_PDZ_Class_1	
37.	LIG_Pex14_1	
38.	LIG_Pex14_2	
39.	LIG_PTB_Apo_2	
40.	LIG_PTB_Phospho_1	
41.	LIG_Rb_LxCxE_1	
42.	LIG_SH2_GRB2	
43.	LIG_SH2_SRC	
44.	LIG_SH2_STAT3	
45.	LIG_SH2_STAT5	
46.	LIG_SH3_1	
47.	LIG_SH3_2	
48.	LIG_SH3_3	

49.	LIG_SUMO_SIM_anti_1
50.	LIG_SUMO_SIM_par_1
51.	LIG_TRAF2_1
52.	LIG_TRFH_1
53.	LIG_TYR_ITIM
54.	LIG_TYR_ITSM
55.	LIG_UBA3_1
56.	MOD_CDK_SPK_2
57.	MOD_CDK_SPxK_1
58.	MOD_CDK_SPxxK_3
59.	MOD_CK1_1
60.	MOD_CK2_1
61.	MOD_GlcNHglycan
62.	MOD_GSK3_1
63.	MOD_LATS_1
64.	MOD_NEK2_1
65.	MOD_NEK2_2
66.	MOD_N-GLC_1
67.	MOD_N-GLC_2
68.	MOD_PIKK_1
69.	MOD_PKA_1
70.	MOD_PKA_2
71.	MOD_PLK
72.	MOD_ProDKin_1
73.	MOD_SUMO_rev_2
74.	TRG_ENDOCYTIC_2
75.	TRG_ER_diArg_1
76.	TRG_NLS_MonoExtC_3
77.	TRG_NLS_MonoExtC_4

ELMs of Proteins interact with their specific counter domains (CDs) present in other proteins. CDs are larger stretches of conserved sequences of amino acids on

proteins. Binding of ELMs with CDs play major role in stabilizing the protein-protein interactions. 46 counter domains were found in human proteins which could interact with these ELMs, enabling HPV proteins to interact with human proteins on the basis of presence of their respective ELMs on them (table 4.2). On the basis of presence of these domains in human proteins, over all 9468 proteins were predicted to have possibility of interactions with HPV proteins. Individual number of proteins predicted for HPV proteins E1, E2, E4, E5, E6 and E7 are given in table 4.3 Domain PKinase has been found on more than 3000 human proteins, while only 9 proteins have shown the presence of domain Adap_comp_sub. These proteins were grouped as Human Predicted Proteins (HPP), containing human proteins which might have interactions with proteins of HPV. This prediction was based on the mechanism of protein-protein interactions mediated through ELMs and their CDs. FHA and PKinase domains show interaction with ELMs of all HPV proteins.

TABLE 4.2: Counter Domains of ELMs found on HPV proteins, The blue color indicates the possibility of interaction of respective HPV protein with CD.

S. No.	CDs	E1	E2	E4	E5	E6	E7
1	14-3-3						
2	Actin						
3	Adap_comp_sub						
4	Arm						
5	Armadillo type fold						
6	Atg 8						
7	BIR						
8	Branch						
9	CKS						
10	Clathrin_propel						
11	Cyclin_N						
12	FHA						
13	Hormone_recep						
14	ICAP-1_int_bdg						
15	IF4E						

16	IRS
17	Kelch_1
18	MATH
19	Metallophos
20	EF_hand_7
21	PDZ
22	Peptidase _M16
23	Peptidase _S8
24	Peptidase_C14
25	Peptidase_C50
26	PI3_PI4_kinase
27	PID
28	Pkinase
29	Pkinase_C
30	PNP_UDP_1
31	PTB
32	Pex14_N
33	Rad60-SLD
34	RB_B
35	SH2
36	SH3_1
37	SH3_9
38	STT3
39	SWIB
40	TRF
41	UBA3
42	UQ_con
43	USP7_ICP0_bdg
44	WD40
45	WW
46	Zf_MYND

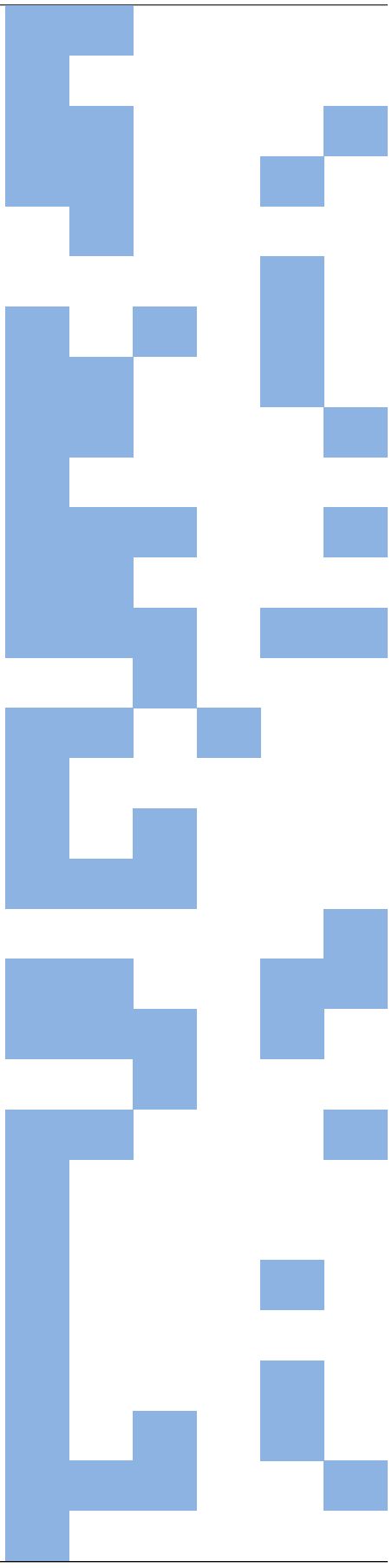


TABLE 4.3: Number of Human Proteins predicted to have interactions with individual HPV Proteins

S. No.	HPV Proteins	Number of Human Proteins
1	E1	4006
2	E2	395
3	E4	1230
4	E5	280
5	E6	3657
6	E7	3390

4.1.2 Prediction of Human Proteins having Interactions with HPV Proteins

This work has been verified by finding significant overlap between predicted proteins and validated proteins, which are experimentally proven for interacting with HPV proteins. All the experimentally reported proteins having interactions with HPV proteins were grouped into a Human Experimented Proteins (HEP). For every HPV protein, significance of the overlap has been evaluated between human proteins and HPV proteins. In Table 4.4. The HPP column shows the number of human proteins which have been predicted in this work for having interaction with individual HPV proteins (Appendix A-1 to A-6), while HEP are the number of humans proteins which have been verified experimentally as proteins targeted by HPV in the NCBI database, and VirusHost database. The lists of HEP proteins is provided in Appendix A-7. The column of Overlap gives the counts of the number of proteins present in both sets. P-values for this overlap between validated and predicted proteins were calculated by using a hypergeometric test in R. p-value gives the significance of results. P-value < 0.01 gives statistically highly significant results.

This work has been verified by finding significant overlap between predicted proteins and validated proteins, which are experimentally proven for interacting with HPV proteins.

TABLE 4.4: Overlap between HEP and HPP with p-values for all individual HPV proteins

HPV protein	HPP	HEP	Overlap	p-value
E1	4006	853	150	1.74E-02
E2	395	534	151	1.38E-02
E4	1230	437	51	4.19E-02
E5	280	790	30	1.42E-19
E6	3657	688	222	1.27E-11
E7	3390	877	204	1.63E-02

We retrieved two attributes from ELM Resource, one was sequences of ELMs on HPV proteins and other was counter domains of human proteins which are known to interact with ELMs. PROSITE database was used to predict the human proteins containing these counter domains of human, that comprises HPP (Human Predicted Proteins) data set. PROSITE predicted HPP dataset of 12958 proteins. This HPP dataset, when compared with experimentally validated proteins, resulted in 808 human proteins out of 4179 proteins, which contain our predicted CDs and that have experimentally proved interactions with HPV proteins. Virus-Host data that contains experimentally validated proteins of human interacting with HPV proteins. This step was performed to check the validity of predicted domains in experimentally tested proteins [48, 227, 228]. For example, in our study, E1 protein of HPV was predicted to interact with 4006 proteins that contain the CD of human of ELMs on E1 of HPV. We found interactivity with 853 proteins in total, among experimentally proved human proteins from database that have interactivity with HPV. Out of 853 experimentally reported human proteins interacting with E1 of HPV, 150 proteins were common among both databases and this also confirms the presence of predicted domains. P-value also has confirmed the validity of results.

Further validation of these results has been done by using Gene Ontology and KEGG Pathways [48, 227, 228]. In Virus-Host database human proteins interacting with HPV proteins are mentioned but interactions at residual level, of CDs

and ELMs is not available. PROSITE can tell the proteins by giving input of predicted CDs retrieved from ELM Resource. The objective of this study was to model the HPV proteins to find the fluctuating ELMs and to compare with our predicted ELMs. Predicted ELMs were mapped on fluctuating regions of HPV proteins exhibited by Elastic Network Models of HPV proteins. The occurrence and interactivity of ELMs with CDs of human proteins was cross verified from literature. The values of B-factors of individual amino acid residues were used to predict hot spot residues of HPV proteins. The prediction and validation of protein has only been done to just identify the proteins carrying counter domains that were our preliminary data to find the interactivity of ELMs of HPV proteins. Wet lab techniques used to analyze PPIs at residual level are based on X-ray crystallography and NMR. But each residual behavior for a single PPI involves a complete separate project. These techniques provide information about the basic molecular structure of proteins. Additionally, these techniques have certain limitations. They analyze the protein in fixed and rigid state. NMR analyzes the structure in magnetic field. While Elastic network modeling provides information about the flexibility and fluctuations of individual amino acids, in cellular environment.

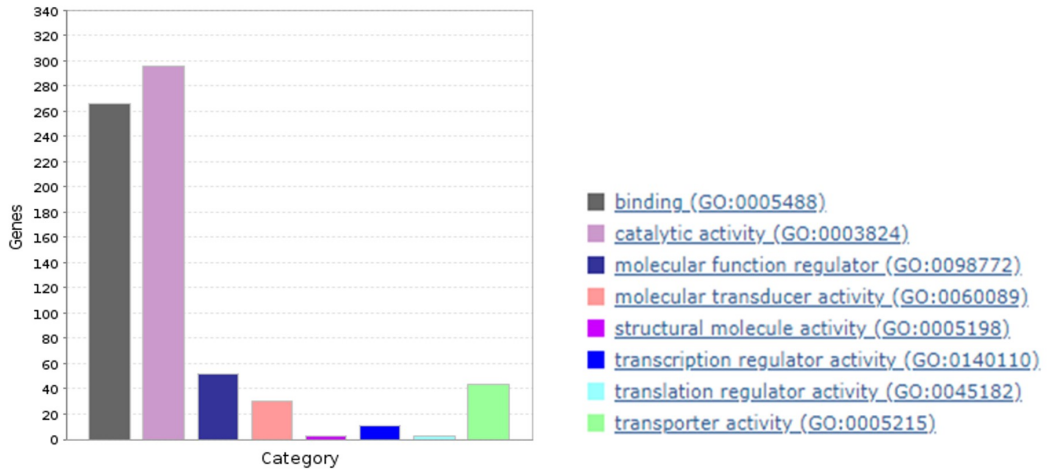
4.1.3 Gene Ontology Analysis

For the statistical comparison of GO molecular function similarity between validated and predicted viral targeted proteins, a permutation test has been devised that is based on GO similarity between proteins, and it was calculated by using the Geneontology tool. GO is a series of levels of labels, having general Molecular Function at the first level, and more specific molecular functions at higher levels. Validated and predicted viral targeted proteins were compared by limiting the comparison to the fifth level of the hierarchy of GO. This level was chosen because the molecular functions labels at this level are specific enough for a significant comparison, and yet quite enough for annotation of large numbers of proteins in our validated and predicted sets of proteins. The term Molecular Function describes activities which occur at the molecular level, for example “catalysis” or “transport”. Generally, Activities which are performed by individual gene products (RNA or proteins) are categorized into molecular functions, but there are

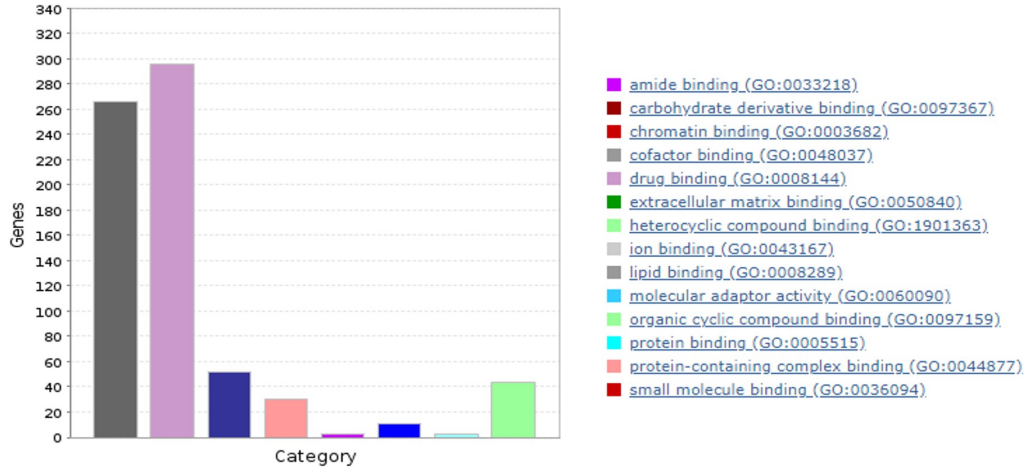
certain activities which are performed by specific molecular complexes formed by combinations of multiple gene products. For example transporter activity and catalytic activity are broader functional terms; while Toll-like receptor binding or adenylate cyclase activity are examples of narrower functional term. GO molecular functions are often added with the word "activity" in order to avoid misconception between names of gene products and their respective molecular functions . All the molecular functions are categorized at first level into (1) binding (2) Catalytic activity (3) Molecular function activity (4) Molecular transducer activity (5) Structural molecule activity (6) Transcription regulator activity (7) Translation regulator activity and (8) Transporter activity. Among the predicted proteins for E1, the highest percentage of genes was shown for the catalytic activity (43%) and binding (37%), at the first level, the lowest percentage of 0.3 was shown by structural molecule activity and transporter activity. For the second level of the binding the protein binding had the maximum numbers of proteins involved (60%). At the third level of protein binding the maximum number of proteins was shown to be involved in enzyme binding (64%, which in turn had the largest number of proteins involved in GTPase binding at fourth level. At the 5 level of protein binding the small GTPase binding had the maximum percentage of the proteins from previous level, involved. In Figure 4.1, graphs of GO analysis of molecular functions of proteins having interactions with E1 are shown. The GO statistically enriched (p-value$<$0.01) molecular functions uptil the 5th level are shown in Table 4.5 for predicted proteins having interactions with HPV proteins, individually. The p-value gives the significance of the results. Every value in the table corresponds to the respective function or activity with the protein predicted to be having interaction with respective HPV protein.

For the statistical comparison of GO molecular function similarity between validated and predicted viral targeted proteins, a permutation test has been devised that is based on GO similarity between proteins, and it was calculated by using the Geneontology tool. GO is a series of levels of labels, having general Molecular Function at the first level, and more specific molecular functions at higher levels. Every value in the table corresponds to the respective function or activity with the protein predicted.

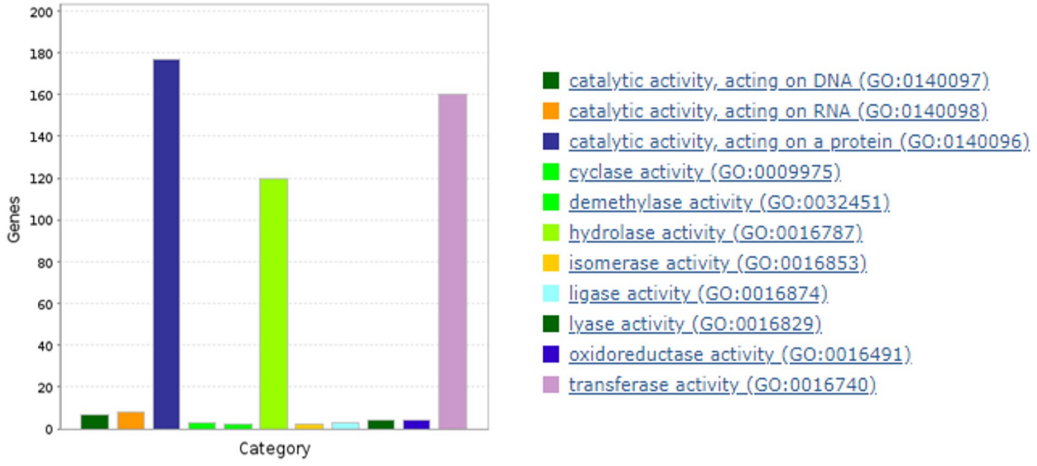
Level 1 for Molecular Functions



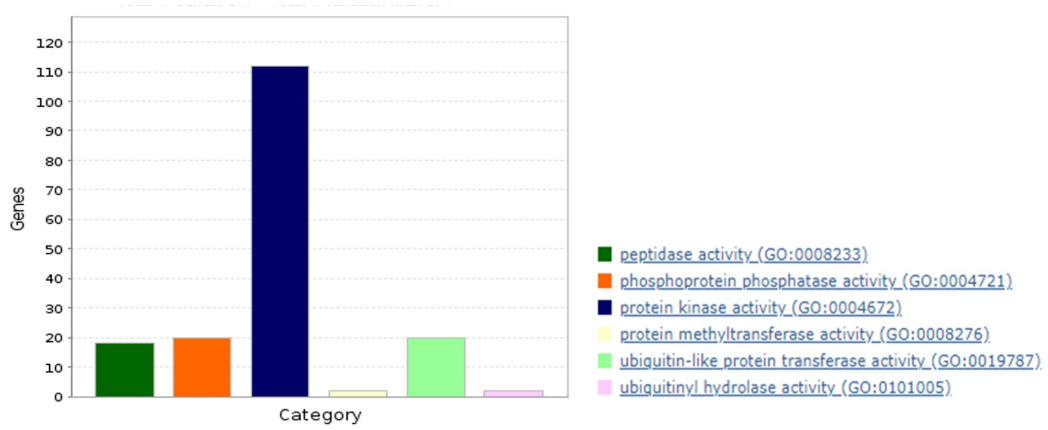
Level 2 for Binding



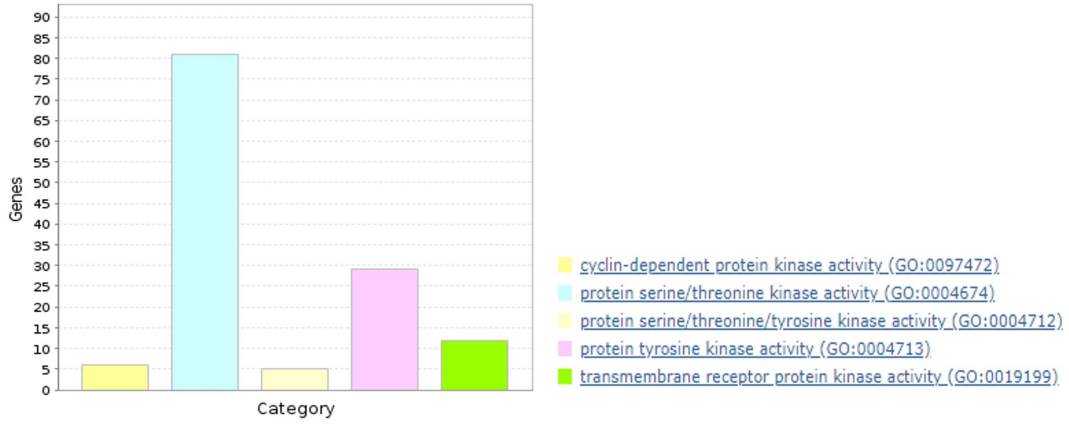
Level 2 for Catalytic activity



Level 3 for Catalytic activity, acting on DNA



Level 4 for protein kinase activity



Level 5 for serine/threonine kinase activity

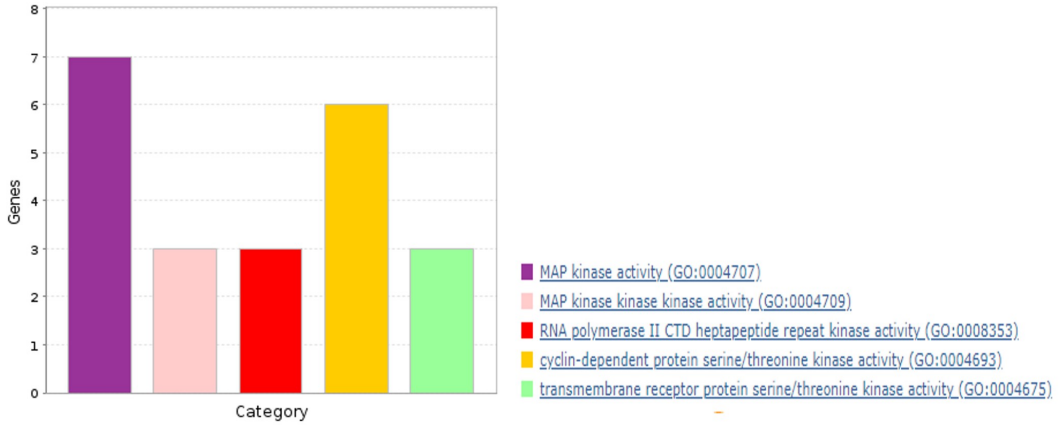


FIGURE 4.1: GO Molecular function analysis of Human proteins predicted to be having interactions with HPV E1; The graphs are showing the number of genes of the predicted proteins involved in the respective molecular function. The hierarchies of molecular functions from level 1 to level 5 are shown.

Table 4.5: GO Molecular Function analysis of predicted proteins for HPV proteins

S. No.	Molecular Function	E1	E2	E4	E5	E6	E7
1.	1-phosphatidylinositol-3-kinase activity	3.47E-07	3.47E-07	3.94E-09			6.39E-06
2.	1-phosphatidyl-inositol-4-phosphate 3-kinase activity	2.35E-04	2.35E-04	8.28E-07			1.80E-03
3.	Rac guanyl-nucleotide exchange factor activity	1.60E-06	1.60E-06				
4.	Rab guanyl- nucleotide exchange factor activity					9.54E-04	
5.	Phosphatidy-linositol-3- phosphatase activity	3.01E-03	3.01E-03			4.82E-02	2.20E-02
6.	Phosphotyrosine residue binding	2.39E-15	2.39E-15			2.78E-1	1.67E-12
7.	Insulin receptor substrate binding	3.57E-02	3.57E-02			5.69E-03	2.24E-04
8.	Insulin receptor binding					4.89E-02	1.86E-02
9.	Transmembrane receptor protein tyrosine phosphatase activity	1.80E-02	1.80E-02			5.36E-03	1.97E-03

10.	Transmembrane receptor protein tyrosine kinase activity					1.26E-09	4.57E-05
11.	Calcium dependent phospholipid binding	3.61E-15	3.61E-15	6.98E-24		6.26E-11	5.17E-13
12.	Proline rich region binding	1.80E-02	1.80E-02				
13.	SH3/SH2 adaptor activity	3.60E-12	3.60E-12			1.71E-08	1.59E-09
14.	RNA helicase activity	8.07E-18	8.07E-18		2.33E-74		
15.	Phosphatidylinositol phospholipase C activity	3.03E-04	3.03E-04	4.73E-05		2.19E-03	4.73E-03
16.	GTPase activator activity	6.05E-53	6.05E-53	1.01E-0		3.83E-52	8.42E-25
17.	Ephrin receptor binding	9.01E-04	9.01E-04			6.79E-03	2.10E-03
18.	Non-membrane spanning protein tyrosine kinase activity	1.04E-05	1.04E-05				
19.	Phosphatidylinositol-3 -kinase binding	7.31E-03	7.31E-03			1.74E-04	2.10E-03
20.	Phosphatidylinositol-3 -phospatase binding					1.66E-02	4.79E-03
21.	Structural constituent of muscle	5.07E-05	5.07E-05				

22.	Protein tyrosine/serine/threonine phosphatase activity	7.91E-05	7.91E-05		3.04E-07	3.88E-08
23.	Protein tyrosine/serine /threonine kinase activity				2.23E-04	4.57E-05
24.	Phosphatidylinositol- 3,4,5-trisphosphatase binding	5.54E-03	5.54E-03		4.58E-05	1.27E-06
25.	DNA helicase activity	1.18E-07	1.18E-07			
26.	Receptor tyrosine kinase binding	1.40E-04	1.40E-04		1.58E-03	6.29E-03
27.	Phosphatidylserine binding	2.00E-02	2.00E-02	6.90E-05	2.98E-03	6.50E-04
28.	Clathrin binding	2.68E-02	2.68E-02	1.02E-03		2.22E-02
29.	SH3 domain binding	3.15E-07	3.15E-07			1.54E-03
30.	Calcium ion binding	5.44E-48	5.44E-48	1.08E-07		
31.	Syntaxin binding	1.31E-02	1.31E-02	1.07E-05		
32.	Ion channel binding	8.95E-06	8.95E-06			
33.	Protein C-terminus binding	4.86E-05	4.86E-05		8.24E-04	
34.	Protein phosphatase binding	1.04E-02	1.04E-02		4.60E-08	6.50E-07
35.	Cadherin binding	1.80E-02	1.80E-02		3.92E-04	2.16E-03
36.	Actin binding	3.48E-02	3.48E-02		5.76E-11	3.51E-08

37.	ATP binding	1.82E-08	1.82E-08		4.38E-61	2.74E-57	8.30E-69
38.	DNA binding transcription factor activity	2.07E-03	2.07E-03				
39.	Protein containing complex binding	7.39E-03	7.39E-03	6.26E-04			
40.	phosphatidylinositol-3, 4-bisphosphate 5-kinase activity			1.22E-02			
41.	phosphatidylinositol-4, 5-bisphosphate 3-kinase activity			2.02E-02			
42.	Phosphatidylinositol bisphosphate binding					8.07E-05	7.95E-06
43.	Rab GTPase binding			7.72E-03			
44.	Rac GTPase binding					8.51E-06	
45.	Phosphatidylinositol binding			1.46E-02			
46.	Transmembrane transporter activity			1.69E-02			
47.	Transmembrane signaling receptor activity			7.14E-04		8.26E-03	
48.	RNA dependent ATPase activity				6.19E-03		
49.	Annealing helicase activity				1.21E-04		

50.	Four way junction helicase activity	9.27E-03		
51.	3' 5' DNA helicase activity	3.13E-03		
52.	Double-stranded RNA binding	9.66E-04		
53.	Single-stranded RNA binding	1.70E-03		
54.	Chromatin binding	1.29E-02		
55.	DNA binding	6.08E-06		
56.	Cyclin dependent protein serine/ threonine kinase activity		6.19E-10	6.85E-10
57.	Magnesium dependent protein serine/threonine phosphatase activity		3.94E-02	1.63E-02
58.	Microfilament motor activity		1.21E-06	2.50E-07
59.	Actin dependent protein kinase activity		4.82E-02	
60.	Actin dependent ATPase activity			2.20E-02
61.	Tau protein kinase activity		7.90E-05	2.07E-05
62.	Tau protein binding		7.68E-03	5.85E-05
63.	Non-membrane spanning protein tyrosine kinase activity		1.43E-10	1.12E-11
64.	Histone kinase activity		3.94E-02	

65.	Rho guanyl nucleotide exchange factor activity	2.38E-04	
66.	Rac guanyl nucleotide exchange factor activity		2.27E-02
67.	Receptor signaling complex adaptor activity	2.11E-0	1.94E-03
68.	Kinase regulator activity	1.65E-02	
69.	Protein domain specific binding	1.27E-07	
70.	Nuclear receptor activity		5.89E-14
71.	Steroid hormone receptor activity		1.14E-1
72.	Protein kinase C activity		1.63E-02
73.	Histone kinase activity		1.63E-02
74.	Proline rich region binding		4.19E-02
75.	Calmodulin binding		1.42E-19
76.	Identical protein binding		1.27E-11

4.1.4 KEGG Pathway Analysis

KEGG mapping or enrichment analysis is the method in which proteins or genes from manually generated datasets, are mapped into some interaction, relationship or pathway. It is an operation to create a new data set. On the basis of presence of ELMs in HPV proteins and their counter domains in human proteome, human proteins were predicted, which can interact with HPV proteins. As these interactions can lead to progression of HPV infection into cervical cancer, these predicted proteins (Table 4.4 and Appendix A-1 to A-6) were mapped on KEGG pathways. Three pathways were selected for our analysis, these were KEGG Cancer Pathways, Human Papillomavirus Infection Pathway and Viral Carcinogenesis Pathway. They were statistically analyzed for enrichment analysis for the predicted human proteins involved in interactions with E1, E2, E4, E5, E6, and E7. As aim of this study was to predict human proteins affected by HPV and analyze protein-protein interactions involved in progression of HPV infection into cervical cancer, so these three pathways were selected to validate the proteins predicted by our methodology. As these pathways cover mostly those proteins most which are involved in development of cancers from HPV infection.

4.1.4.1 KEGG Cancer Pathway

KEGG mapping or enrichment analysis is the method in which proteins or genes from manually generated datasets, are mapped into some interaction, relationship or pathway.

KEGG cancer pathway is the network of all the molecular objects like proteins involved in Cancer with experimental evidences. All the predicted proteins were mapped on KEGG Cancer Pathways (Appendix A-8, A-9 & A-10).The enrichment analysis showed proteins which were statistically significant ($p<0.01$). HPV E1, E2, E6 and E7 have been predicted for interaction with these proteins. Among the proteins of our data set, highest number of proteins found statistically significant on KEGG Cancer Pathway, were those which have interactions with E6 (35 out of 39) Table 4.1.

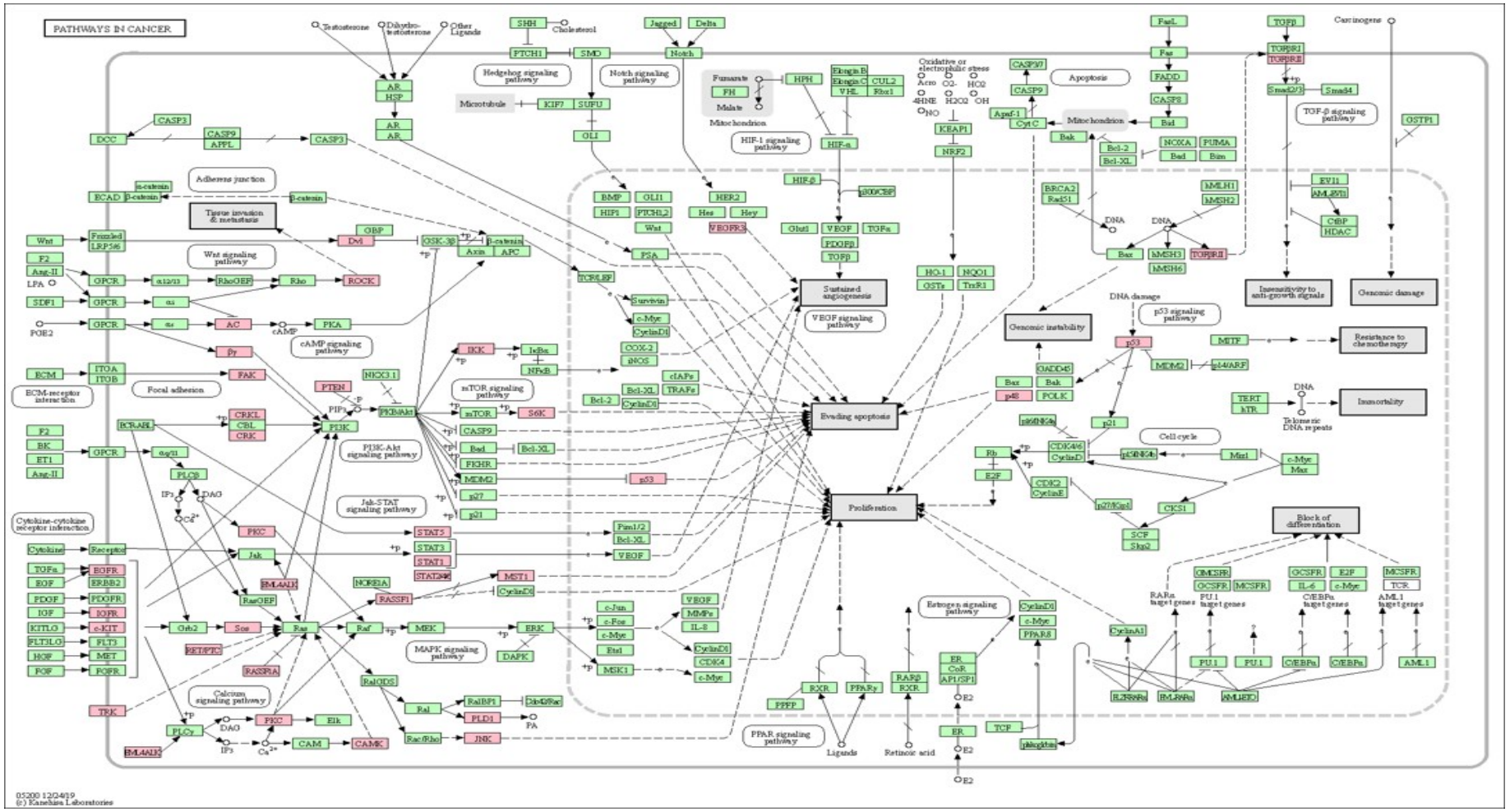


FIGURE 4.2: KEGG Cancer Pathway for enrichment with proteins predicted to be interacted by HPV16 E6.The proteins colored in pink are the statistically enriched ($p<0.01$) proteins among our predicted proteins affected by E6

Only few of predicted proteins from HPP, have been reported to have an interaction with HPV proteins, but due to involvement of all these proteins in cancer and showing the evidence of ability of interaction of HPV proteins with these proteins, we could predict these proteins (Table 4.6) to be affected by HPV proteins and play some role in progression of HPV infection into cervical cancer. These proteins include the proteins which are carcinogenic or enhance the process of cancer, once it occurs, and the proteins which suppress the tumorigenesis (Table 4.6). Among the effectors of the RhoA, which is a small GTPase, Rho-associated coiled-coil kinase (ROCK) is an important protein, which is member of AGC family of serine/threonine protein kinases. Protein kinases A, C and G are also included in this family. Various key cellular mechanisms like tumor growth, tumorigenicity, metastasis, tumor cell apoptosis/survival, angiogenesis, and chemoresistance, are regulated by ROCK, which plays many important roles in progression and development of tumor [239]. Genomic evidence from recent studies of oropharyngeal squamous cell carcinoma caused by HPV has shown activation of phosphatidylinositide 3-kinase (PI3K) pathway. PI3K pathway is known to be triggered, for activation, by mutations or the gene encoding amplification for p110α catalytically active subunit of phosphoinositide 3-kinase (PIK3CA), HRAS mutation and loss of phosphatase and tensin homolog (PTEN) gene activity [240]. In humans and other organisms, the tyrosine kinase signaling process is characterized by a major point of convergence, which possesses a network of essential signal transduction pathways having the Crk adaptor proteins (Crk and CrkL) as an essential part of the network. Human diseases like susceptibility to pathogen infections and cancers have been found to be associated with dysregulation of Crk proteins [241]. Signal transducers and activators of transcription (STATS) belong to a family consisting of transcription factors. All the Members of this family are involved in signaling through interferons (INF). Among these factors of transcription STAT1 and STAT3 show contrasting affect in causing cancer. Promotion of cell survival and proliferation, immune tolerance and cell motility, due to STAT3, is being considered as its oncogenicity, while on other side, enhancement of inflammation and adaptive and innate immunity, in tumor cells is triggered by STAT1 through the mechanisms of pro-optotic and anti-proliferative activities [242]. Son of Sevenless

(SOS), discovered intially in Drosophila melanogaster, is necessary for development of normal eye in Drosophila. There are two human homologues of SOS, SOS1 and SOS2. SOS proteins acts an important causative agent in the carcinogenesis of human bladder cancer [243]. Apoptotic signaling pathways are associated with specific death receptors, which critically, are characterized by presence of a Fas-associated death domain protein (FADD). Cancer cells have been reported to be characterized by having both the repression and overexpression of FADD. Mechanism of its regulation is still under investigations [244]. Epithelial tumorigenesis is contributed by constitutive activation or overactivation of Ras signaling pathways in many ways, cytoplasmic mislocalization of the cyclin-dependent kinase inhibitor is one of those. RalGDS and RalBP1 are GDP binding proteins being the components of these pathways [245]. Inflammation and aggressive tumors are associated with the reduced expression of TGFbRII in humans [246]. Vascular endothelial growth factor C (VEGF-C) that been found to cause lymphangiogenesis, alone. In gastric cancer, the stroma of primary gastric cancerous cells possess proliferated lymphatic vessels, induced by VEGF-C, through the increased activity of factor VEGFR-3, that is part of the endothelial cells in the lymphatic vessels. Due to these changes, the lymphatic vessel can be easily be invaded by the cancer cells. This may result due to increase in the contact points between cancer cells and the lymphatic vessels [247]. Some pathological conditions in humans like cancer, diabetes and obesity are contributed by mTOR/S6K signaling, which involves the enhancement of mTOR pathway by S6K as effector [248]. There are four isozymes (b1-b4) of the enzyme phospholipase C beta (PLCB), belonging to the class of phospholipases. These are encoded by different genes. Members of the PLCB gene family have been associated with several diseases, including cancers [249]. Homeostasis and epithelial tissue development in humans is regulated by an epidermal growth factor receptor (EGFR), that is its major physiological function. Tumorigenesis is driven by the EGFR in pathological conditions mostly like breast cancer, lung cancer and glioblastomas [250]. The cancer cells are survived and proliferated to progress towards malignancy by various factors. One of them is the Insulin-like Growth Factor 1 Receptor (IGF1R). It is a trans-membranous heterotetrametric protein. IGF1R signaling pathway has been seen in biology of lung

cancer, contributing majorly to the pathogenesis and behavior of this malignancy [251].

More than 50% of human proteins predicted in our work were found to be involved in interactions with other human proteins on the basis of motifs and domains operating through different classes of kinases. Tumorigenesis is strictly linked with the activation of receptor tyrosine kinases (RTKs) and their intracellular signal transduction pathways, that cause regulation of several cellular functions like motility, apoptosis, adhesions, angiogenesis and proliferation [252]. Protein kinase C (PKC), a prototypical class of serine/threonine kinases, exemplifies specific signaling molecules that link multiple cellular processes to cancer [253]. The tropomyosin receptor kinase (TRK) family has members TRKA, TRKB, and TRKC proteins, which are translated by *NTRK1*, *NTRK2* and *NTRK3* genes, respectively. Interaction between TRK proteins and neurotrophins causes induction of receptors phosphorylation, dimerization, and downstream signaling cascades activation through PI3K, PLC-gamma, and RAS/MAPK/ERK. Pathogenesis of some cancer types is related to aberrations in TRK pathway, including protein overexpression, gene fusions, and single nucleotide alterations [260, 261]. Ca2+/calmodulin-dependent protein kinase II (CaMKII) is a multiple functional serine/ threonine kinase, gains attention of scientists for its function in memory and learning. Recent researches have shown that a variety of malignant cancers have higher expressions of CaMKII [262]. The Proteases and Cytokines are principal mediators of inflammation and also play some role as effective cancer promoters. Their synthesis and functioning is regulated by the p38 and JNK MAPK pathways [263]. Embryonic development and the pathogenesis of human diseases like cardiovascular diseases and cancer are essentially characterized by a major activity of a specific cytoplasmic protein tyrosine kinase called as Focal adhesion kinase (FAK) [264].

80% of thyroid carcinomas are of papillary histotype (PTC). Despite of having so much experimental knowledge about its pathogenesis epidemiology, biological, and clinical behavior, but the only validated risk factor for PTC is the exposure of ionizing radiations. The pathogenic role is played by rearrangements of the

Rearranged during Transfection (*RET*) proto-oncogene in PTC. The first *RET* rearrangement, called *RET/PTC*, activates constitutively the synthesis of the RET tyrosine-kinase domain in the follicular cell, thus generating the signaling cascade of the MAPK pathway and proliferation in an uncontrolled manure. Uptil now, thyroid cancer has been reported with having 13 various types of rearrangements of *RET/PTC* [254].

The activity and function of PLD are regulated by and widely dependent on hormones, neurotransmitters, lipids and small monomeric GTPases. Inner membranes of endosomes, secretory granules, Golgi apparatus and lysosomes of mammalian cells are characterized by having an enzyme PLD1. PLD1 is transported readily towards the plasma membrane on an external stimulus by the protein binding. Low intrinsic activity is exhibited by PLD1, and this results in its unability for transduction of an extracellular signaling messenger at its basic level. For extracellular transduction to occur, activation by some proteins like Rho, Rac, Arf and protein kinase C is required. G protein-coupled receptors (GPCRs) and receptor tyrosine kinases (RTKs) are utilized for the stimulation of this process, through a wide range of agonists [255].

Fibroblast growth factor (FGF) has the ability to produce a variety of responses in different types of cells. The receptors which interact with FGFs include FGF receptors (FGFRs) which are tyrosine kinase receptors, having high affinity for it. The regulation of FGF activity is dependent on pattern of its expression or synthesis, and specificity for binding between FGF FGFRs [256].

Contrary to these carcinogenic proteins, there is also a group of proteins which act as tumor suppressor proteins. HPV proteins also have been found to have interactions with these tumor suppressor proteins. Among the tumor repressor proteins p53 is one of the major proteins, whose expression leads to suppress tumorigenesis, while any mutation resulting in loss of its activity leads to cancer development [257]. Wnt/ β-catenin signaling pathway is mediated by cytoplasmic family of Dishevelled (Dvl) proteins, which have been recently reported to have a link with cancers [258]. The expression of p53 is dependent on NF-κB, as its activation helps p53 to suppress tumor formation. Some experiments has shown

destabilization of p53 by activation of NF-κB/IKK. NF-κB is also considered as tumor suppressor protein under some specific conditions [259].

The cell signaling is involved by a scaffold tumor suppressor protein known as the RASSF1A. Researchers have shown that RASSF1A protein lies at the junction of a complex network of signaling molecules, that includes different cellular regulators that control cellular homeostasis, like MST2/Hippo Ras, death receptor and p53 pathways. Solid tumors are commonly characterized by having the loss of expression of RASSF1A that is oftenly caused due to gene silencing via methylation of DNA. For treatment of different tumors it is a promising thoroughfare, to re-express RASSF1A or target effector modules of its complicated signaling network, therapeutically [260].

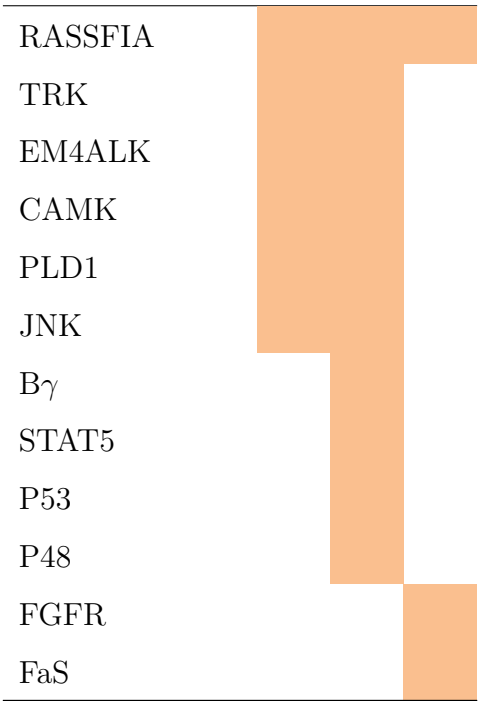
Apoptosis can be induced by a signal transduction pathway, that involves the binding between of Fas with its ligand, known as FasL. Fas is member of superfamily of tumor necrosis factor (TNF). It is a type I transmembrane protein, and its combination with FasL or a related membrane associated antibody results in activation of caspases, which leads towards the apoptosis. Inhibition of apoptosis, proliferation and unusual survival of tumor cells, and promotion of malignancies, are consequences of impaired functioning of FasL or Fas [261]. Additionally, some other proteins also have been found which can interact with HPV proteins and have both the tumorigenic and tumor suppressor abilities, depending on the cellular conditions. Specific gene expressions and mitochondrial functions are regulated by many cytokine receptors through stimulation of the Signal Transducer and Activator of Transcription (STAT) 3 and 5 proteins. These stimulators have the ability to both enhance tumorigenesis and suppress tumor formation, depending on the cellular environment and other contexts [262]. The ErbB3-binding protein 1 (Ebp1) has been reported as either an oncogenic regulator or a tumor suppressor in a variety of cancers [263].

ROCK, PTEN, CRKL, CRK, STAT1, STAT2, Sos and RalBP1 have been predicted for having interactions with HPV E1, E2 and E6. FADD and RalGDC has been found to be affected by only E1. Both E2 and E6 have been shown interactions with FGFBRII, VEGFR3, Dvl, IKK, S6K, AC, FAK, PLCB, PKC, EGFR,

IGFR, c-KIT, MSTI, RET/PTC, RASSFIA, TRK, EM4ALK, CAMK, PLD1 and JNK. Bγ, STAT5, P53 and P48 have shown interactions with E6 only. While E7 alone can interact with RASSF1A, FGFR and FaS. No proteins have significantly shown enrichment, possessing ability of interactions with E4 and E5.

TABLE 4.6: List of predicted proteins giving significant enrichment analysis in KEGG Pathways in Cancers. The colored blocks show the interaction of respective protein with viral protein.

Proteins	**E1**	**E2**	**E6**	**E7**
ROCK				
PTEN				
CRKL				
CRK				
STATI				
STAT246				
Sos				
FADD				
RalGDC				
RalBP1				
FGFBRII				
VEGFR3				
Dvl				
IKK				
S6K				
AC				
FAK				
PLCB				
PKC				
EGFR				
IGFR				
c-KIT				
MSTI				
RET/PTC				



4.1.4.2 KEGG Human Papillomavirus Infection Pathway

The KEGG Human Papillomavirus Infection Pathway clusters the proteins involved in infection and pathogenesis by HPV in a network or pathway. The causing agents of cancers like that of cervix and other genital organs include majorly the high-risk human papillomaviruses (HPVs). Additionally many oropharyngeal cancers are also developed by HPV infections. Various pathways of host cells are either activated or repressed by HPVs, for promoting their own life cycles. This also majorly includes affecting DNA damage response. activate and repress a number of host cellular pathways to promote their viral life cycles, including those of the DNA damage response. Proteins which were significantly enriched in the Human papillomavirus pathway (Appendix A-11 to A-15) are shown in Table 4.7. Among these TRADD, p53, ECM and PP2A are tumor suppressor proteins. HPV proteins interact with these proteins to suppress their activity which leads to tumorigenesis. While other proteins including PTEN, Dlg, Pals1, FADD, SOS, STAT1, STAT2, ATM, Mi2β, Dvl and Hugl-1 are tumorigenic proteins. The ataxia telangiectasia-mutated (ATM) and ATM and Rad3-related (ATR) DNA damage repair pathways, are activated by HPVs, which are essential requirements

for replication of viral genome. The stability and integrity of maintenance of host genome is dependent on these pathways. HPVs oftenly cause their dysregulation or mutations in tumorigenesis [264]. Epigenetic reprogramming of the cell occurs in an extensive manure, when the high-risk HPV E7 is expressed inside the cell. In turn, entry of cell into cell cycle is stimulated and progressed after expression of E7. HPV 16 E7 also interacts with Mi2β, which is a component of the nucleosome remodeling and deacetylase complex (NuRD complex). This interaction is considered important in blocking of the activity of histone deacetylases 1 and 2 [265]. A variety of proteins which play some role in cell polarity regulation are characterized by having PDZ domains. These proteins are targeted by the oncoprotein E6 being having the ability to interact with such proteins on the basis of PDZ domain. Different types of E6 belonging to different types of HPVs, target different protein, like Scribble, Dlg1, MUPP1, MAG-I, II and III. MUPP1 is part of subapical tight junctions proteins, while the MAG proteins are situated at regions of cell to cell contact. Thus E6 targets the same process but through different intermediates. Scribble, and Dlg1 are examples which are targets of E6 of HPV-16 and HPV-18, respectively. hScrib, hDlg and Hugl-1 are among the proteins which are undetectable at the last stages of some malignancies. They are undetectable at the advanced stages of many types of epithelial cancers, which make the identification of cause of cancers difficult, that either they are due to E6 or other factors [266].

Cell transformation is mediated due to the expression of two pleiotropic oncoproteins by human papillomaviruses which cause alterations in major regulatory cellular pathways. However, these viruses or viral proteins are not completely carcinogenic, and for tumor progression and establishment, more changes inside and in the microenvironment of the infected cells are necessary requirements. For example, alterations in basic components of the extracellular matrix, matrix metalloproteinases and few of their regulators like tissue inhibitors of metalloproteinases, have been found to play a main role in human papillomaviruses-associated diseases [267]. As ECM act as barrier in progression of cancer [268]. Various human malignancies have been reported to show unusual mutations, expressions, and somatic alterations of the scaffold of Protein Phosphatase 2A (PP2A) and its regulatory

subunits. Many human malignant tumors have been recently identified as having cancerous inhibitors of protein phosphatase 2A (CIP2A) . In cervical cancer also this CIP2A is overexpressed, and this overexpression is caused by E7 [269]. In KEGG Human Papillomavirus Infection Pathway PP2A has shown significant interaction with E4 and E7. Another tumor suppressor protein Pals1 regulates cell polarity, cancer progression and Hippo signaling. It has been shown to be targeted by E6 through PDZ domain [270]. Table 4.7 shows the significant proteins in enrichment analysis of Human Papillomavirus Infection Pathway.

TABLE 4.7: List of predicted proteins giving significant enrichment analysis in KEGG Human Papillomavirus Infection Pathway. The colored blocks show the type of individual HPV protein having interaction with respective protein

Proteins	**E1**	**E2**	**E4**	**E5**	**E7**
PTEN	■	■			
Dlg	■	■			
Pals1	■	■			
FADD	■	■			
SOS	■	■			
STAT1	■	■			
STAT2	■	■			
ATM	■	■	■		
Mi2β	■	■		■	
Hugl-1			■		
Dvl			■		
ECM			■		
PP2A			■		■
p53			■		
Fas					■

4.1.4.3 KEGG Viral Carcinogenesis Pathway

KEGG viral Carcinogenesis Pathway shows the relationships and mechanisms of integration among different molecular objects involved in development of cancers

from different viruses like HPV, Hepatitis, HIV, etc. The proteins, among the predicted proteins, which have shown significant enrichment analysis with KEGG Viral Carcinogenesis Pathway (Appendix A-16 to A–20), include14_3_3, 14-3_3ϵ, hDLG, Actin. Dlg1, Mi2β, SyK, P53, CDC20. TRADD, MK2, GapIm, JAK3, and hSCRIB. All the eukaryotic cells possess a family of acidic proteins which are highly conserved and 14-3-3 proteins belong to this family. There are seven members of this protein family including β, γ, ϵ, η, τ , ζ and α. These members function both as heterodimers and homodimers. A variety of functional regulators belonging to different biological processes are associated by these proteins. This binding occurs through interactions with particular motifs of phosphoserine and phosphothreonine. These interactions enable them for regulating the cell cycle, apoptosis, intracellular protein trafficking, DNA-damage response, DNA replication, and DNA transcription. A major role in virus infections is played by these 14-3-3 proteins [271]. E1 and E2 has been predicted for having interactions with proteins containing 14-3-3 domain, which include 14-3-3 protein also known as PP2A, epididymis lumen protein, epididymis secretory protein and protein activating Tyrosine 3-monooxygenase/ tryptophan 5-monooxygenase. Epididymis secretory protein has been reported recently as tumor marker in different tumors [272]. In present study ELM LIG_14-3-3_CanoR_1 has been seen in E1 and E2 which interacts with 14-3-3 domain.

Spleen tyrosine kinase (Syk), a non-receptor protein tyrosine kinase, has recently been recognized as a new candidate tumor suppressor. A variety of cancers, caused by viruses, are characterized by decrease or loss of Syk expression which is associated with poor prognosis and a malignant phenotype [273]. Increased expression of CDC20 has been observed in hepatic cancer due to viral hepatitis [274].

Tumor necrosis factor receptor (TNFR)-associated death domain (TRADD) protein is a main adaptor of the signaling complex TNFR1 which controls inflammatory signals as well as cell death. A research on a chemical-induced carcinogenesis model on mice has shown the acceleration of tumor formation due to deficiency of TRADD [275]. Being predicted to be interacted by HPV proteins and significant part of KEGG Viral Carcinogenesis Pathway (Table 4.8), TRADD activity might

be suppressed by HPV proteins. This may direct viral pathogenesis to carcinogenesis. MK2 has been shown in studies to depress different tumor suppressor proteins in HIV, Hepatitis and influenza viruses [276, 277]. In our study HPV E6 has been significantly appeared as interacting partner of MK2. This interaction could also contribute to the carcinogenic ability of HPV. The nasopharyngeal carcinoma has been considered to have highly invasive nature, that is contributed by the principal oncoprotein of Epstein–Barr virus (EBV), known as latent membrane protein 1 (LMP1). Janus kinase 3 (JAK3) and Extracellular signal-regulated kinase 1/2 (ERK1/2) pathways are signaled by LMP1, that is triggered by the activation of STAT3 as well as STAT transactivation activity.

TABLE 4.8: List of predicted proteins giving significant enrichment analysis in KEGG Viral carcinogenesis Pathways. The colored blocks show the individual HPV protein having interaction with respective significant protein

Proteins	**E1**	**E2**	**E4**	**E5**	**E6**
14_3_3	■	■			
14-3_3ϵ	■	■			
hDLG	■				■
Actin	■				
Dlg1	■				■
Mi2β	■			■	
SyK	■				■
P53			■		■
CDC20			■		■
TRADD					■
MK2					■
GapIm					■
JAK3					■
hSCRIB					■

Among these proteins some proteins were found common in all three pathways analyzed like PTEN, p53, Dlg1, Dvl, Fas, FADD [240, 244, 261, 264, 266]. These proteins have been found having an association with the cancers development in

various studies. PTEN, Dlg1, Dvl and FADD are tumorigenic proteins, while p53 and Fas are tumor suppressors.

P53 is cancer suppressor protein. E4 and E6 have been found to interact with p53. In viral carcinogenesis and cancer pathways p53 has shown interactions with E6 commonly. It reflects the inability of the p53 to suppress cancer development after it has been blocked to interact by E6 and E4 with its counter proteins in order to perform its function. PTEN, Fas, FADD, SOS, STAT1, are among the proteins common in HPV pathogenesis pathway and cancer pathway. E1 and E2 have been found to interact with PTEN, STAT1, SOS, and FADD in HPV pathogenesis pathway as well as in cancer pathway, because these are the primary proteins required for the pathogenesis and HPV infection that is the prerequisite for HPV to cause malignancy. Fas, being the component of tumor necrosis factor has been found to be impaired in cancers, is interacted by E7 that is the major tumorigenic HPV protein. Fas has not been found to have any interaction with E1 and E2 or E6 in any of the studied pathways. hDlg has been found to be interacted by E1 and E2 in HPV infection pathway while both E1 and E6 can interact with hDlg in viral carcinogenesis pathway as hDlg has been proved by studies to be affected in development of cancers [243, 245, 246, 247, 260, 264, 265].

4.2 Elastic Network Modeling of HPV Proteins E1, E2, E4, E5, E6 and E7

The first objective of this study was to find ELMs of HPV proteins and their interacting domains of host proteins, and prediction of human proteins which could have possibility of interactions with HPV proteins, on the basis of interactive relationships between predicted ELM and CDs. Second objective was to analyze dynamic fluctuations of amino acid residues which stabilize host pathogen proteins interactions. The major and third object was the elastic network modeling of HPV proteins. ENMs of viral proteins have been used to have an insight into the flexible regions of viral proteins and nature of their role in establishing these interactions.

ENMs were modelled by using 3D structures of HPV proteins as input (Appendix A-39). 3D structures of HPV proteins E1, E2, E4, E5, E6 and E7 were modeled by using FASTA amino acid sequences of these proteins, with use of RaptorX. Pdb files of 3D structures of HPV16 proteins were used as input for generating Elastic Network Models of the proteins.

4.2.1 Analysis of ENMs of E1

The most conserved and largest open-reading frame of the genome of HPV encodes for the E1. The E1 protein of HPV is splitted into four major domains, a DNA binding domain, a N-terminal domain, a helicase domain at C-terminal region and oligomerization domain.

4.2.1.1 Motifs on E1

Mapping of E1 amino acid sequence on ELM, E1 was found to have 7 motifs which interact with the cyclin domain of proteins, (Table 4.9) and have the possibility to be involved in phosphorylation of proteins through cyclin/cdk complexes. A CDK phosphorylation site should be present in such proteins. Two of these motifs of E1 have been found in N-terminal region which interact with the domains having CDK phosphorylation sites. Being possessing these motifs with ability to interact with cyclin/CDK domains, and being acting as a potent NES, E1 has the potential to interlocalize between nucleus and cytoplasm. This property of E1 of interacting through cyclin/CDK domains, could play a major role in transformation of HPV infection into cervical carcinoma.

There are 200 amino acid residues which form the the N-terminal part of E1. Different short but conserved amino acid sequences are in this region, which form specific motifs. These motifs are conserved variably among different PVs. They include a bi-partite nuclear localization signal (NLS), a cyclin-binding motif (CBM) that interacts with cyclin A/E in complex with cyclin-dependent kinase 2 (Cdk2) and phosphorylation sites for Cdk2, a Crm1-dependent nuclear export

signal (NES), and other kinases [278]. Transporting factors are involved in recognition of the short sequences of peptides, possessed by the proteins that travel between the nucleus and cytoplasm oftenly. Proteins having a nuclear localization sequence (NLS) are transported into the nucleus from cytoplasm and take help of a family of importins α/β heterodimers. On other hand, those proteins having a nuclear export sequence (NES) form a complex with exportins as well as RanGTP for the transportation from the nucleus to the cytoplasm. These activities of the NES and NLS are regulated through posttranslational modifications, which include phosphorylation, that can control the accessibility by and the affinity for the transporting proteins [279].

TABLE 4.9: The motifs with sequence and position of amino acid residues of ELMs in N-terminal region of E1

S. No.	Motif	Amino acid Sequence	Amino acid Residue Position in E1
1.	DOC_CYCLIN_1	KYLV	89-92
2.	DOC_CYCLIN_1	RRLF	124-127
3.	DOC_CYCLIN_1	KELY	212-215
4.	DOC_CYCLIN_1	KTLL	252-255
5.	DOC_CYCLIN_1	KLLCV	296-300
6.	DOC_CYCLIN_1	KSLF	483-486
7.	DOC_CYCLIN_1	RPLV	552-555
8.	MOD_CDK_SPK_2	NNISPR	104-109
9.	MOD_CDK_SPxxK_3	NNISPRLK	104-111

E1 of HPV31 has been found to have an unstructured sequence of 82 amino acid residues at N-terminal region, and is said to be having characteristics of a genuine intrinsically disordered domain (IDD) [280]. IDDs are commonly found among viral proteins and they have the ability to accommodate protein interactions with different protein partners by using a mechanism of folding-upon-binding. These IDDs consist of larger proportions of charged amino acids, which participate in different mechanisms of post-translational modifications. N terminal of E1 is involved in phosphorylation. IDD of HPV31E1 possesses a binding site which binds with

a p80/Uaf1 protein. This binding is based on a conserved and short hydrophobic motif Φ-x-x-Φ-Φ, where Φ represents some hydrophobic amino acid residue. This sequence makes an alpha helix of amphipathic nature. In an experiment substitution of three amino acid changed the nature of sequence and resulted in twofold reduction of viral DNA replication in the host. Thus it shows that this small N terminal region composed of short linear motif is important in regulating the activity of E1. On mapping E1 on ELMs a motif MOD_SUMO_rev_2 was found at position 22-29. This motif interacts with UQ_con domain which is a part ubiquitin based protein and has been reported previously to be the part of WD repeat protein p80 [280]. So our work validates the interaction of E1 with p80 and shows that this interaction is stabilized by a specific motif of MOD_SUMO_rev_2.

4.2.1.2 Interaction of E1 with Histone H1 and Ini1/hSNF5

In present study the information about the motifs of E1, their CDs, and the ENMS has been used to analyze the interactions of E1 with different proteins like histones, and chaperones.

The genome of HPV has been found to be associated with nucleosomes. E1 has been reported to establish interactions with host proteins, which modify structure of chromatin. In a study, histone H1 has been found to be interacted by the 185 amino acids long N terminal region of HPV11 E1. It binds with nucleosome and creates a more rigid and ordered DNA structure [281]. In vitro E1 has shown the capability to displace Histone H1. This means that H1 could cause some resistance to E1 in producing the effects required for its stability in host cell, and E1 interacts with H1 to relieve that stress. The chromatin remodeling complex AWI/SNF is composed of different subunits, and Ini1/ hSNF5 subunit among these has been observed in a study to be interacted by HPV18 E1. This interaction has been observed by both HPV11E1 and BPV1E1. The region of E1 which binds with Ini1/ hSNF5 was spanning over the amino acids 147 to 444, which is comprised of DNA binding Domain (DBD) and oligomerization domain. Substitution of the most conserved amino acid residues in this region lead to decreased binding of E1 with Ini1/ hSNF5, causing reduced replication of viral DNA. This concludes that

binding of E1 with Ini1/ hSNF5 is major requirement for the HPV genome to replicate in host cell [282]. However, given that by these amino acid substitutions most conserved sequences of E1 are affected, and this also impairs some other activities of the E1 protein, like ori-binding, ATP-hydrolysis, and oligomerization, which are major requirements for replication of viral DNA.

TABLE 4.10: The motifs having CDs of P Kinase in N terminal region of E1

S. No.	**Motifs**
1.	DOC_MAPK_gen_1
2.	DOC_MAPK_MEF2A_6
3.	MOD_CDK_SPK_2
4.	MOD_CDK_SPxxK_3
5.	MOD_CK1_1
6.	MOD_CK2_1
7.	MOD_GSK3_1
8.	MOD_NEK2_1
9.	MOD_PKA_2
10.	MOD_PLK
11.	MOD_ProDKin_1
12.	TRG_NLS_MonoExtN_4

Various cellular proteins and viral proteins like EBNA2 , ALL1, c-MYC, p53 and HPV18-E1 have been found to have some interactions with a protein INI1/hSNF5. This protein is composed of a C-terminal coil, an N-terminal helical winged structure, two repeated sequences of 60 amino acids, termed as Repeat 1 or RPT1 and Repeat 2 or RPT2 [283]. RPT regions play a major role in stabilizing interaction of proteins with INI1/hSNF5, as it is believed to have domains important in protein-protein interactions [284]. The interaction between viral proteins and RPT1 region of INI1/hSNF5 has been shown to be revealed by some kinases. Table shows the motifs having CDs of P Kinase, found in E1 by mapping on ELMs. These motifs are found in N terminal region and possibly may involve in interaction of E1 with INI1/hSNF5, as all these motifs interact with counter domains of Protein Kinases (Table 4.10).

Hsp40 and Hsp70 are the two chaperones protein molecules, by which promotion of the binding of HPV11E1 to origin of replication has been observed, in in vitro studies. These chaperones heighten the attachment of DNA by E1 as a hexamer, differently and by an independent mechanisms, which gives the binding an additive effect.it is an ATP requiring process, in which Hsp 70 interacts with E1 to enhance the binding with DNA. While the Hsp40 promotes the binding of E1 with the origin of replication, in the form of double hexamer. These two chaperones invigorate the synthesis of viral DNA by two mechanisms. One is the dissociation of E1 from E2 at the origin of replication, and secondly by promoting the assembling of E1 as double hexamer to accomplish replication [285]. Hsp 40 and Hsp 70 interact with proteins through a zinc Finger region. The motif of LIG_MYND_2 is the only motif found in E1 by mapping on ELMs. This motif interacts with the zinc finger binding domain. This motif is localized at position 560-564, of C-terminal region. This confirms the involvement of N-terminal region in its recruitment to DNA with the aid of Hsp 40 and Hsp 70.

Replication Protein A (RPA) [286] binds with the E1 through a zinc finger containing region. Yeast p70, p180C heterodimer also has OB-fold and Zn-finger domains and possesses a micro-molar affinity to dsDNA [287]. As reported that E1 interacts with this machinery through C terminal region, which in our study has been found to have a LIG_MYND_2 motif at position 560-564, which binds with zinc finger domain of proteins. This validates the study of group which reported about the interactivity of E1 with both p70 and p180.

All the proteins in cellular systems and living environment possess some strain energy. It is defined as the energy which is stored in a body due to its deformation. The strain energy per unit volume is called strain energy density while the area under the stress-strain curve measures the degree of deformation. Proteins being flexible molecule when experience environmental forces and stress, undergo change in their conformation. But still keep overall energy constant, when this applied force is released, the whole protein returns to its original shape. Deformation energies was also the one of the parameters calculated by ENMs (Appendix A22). Proteins interaction energies are based on individual free energies of amino acid

residues and are related to the nature of R group of that amino acid residue and types of amino acids present in its neighborhood in that specific protein. These free energies are responsible for creating flexibility in specific regions of proteins. This phenomenon also forms the basis of correlation analysis of individual amino acid residues in relation with other amino acid residues present within the protein.

4.2.1.3 ENMs of E1

ENM study of E1 reveals the higher degrees of fluctuation in both the C-terminal and N-terminal domains of E1. Modes 1, 2, 4, 6, 10, 13 and 18 has exhibited higher flexibility in N terminal regions as shown by red color of ENMs of E1. While the modes 1, 4, 5, 8, 11, 14, 16 and 19 show higher fluctuations in C terminal region. As graph of B factors for E1 (Figure 4.3) shows the higher values of B factors in N terminal and C terminal region. This shows the involvement of these terminal regions in interaction of E1 with other proteins as discussed earlier. Remaining modes does not show any evidence of involvement of the their respective flexibles regions in type of protein-protein interactions.

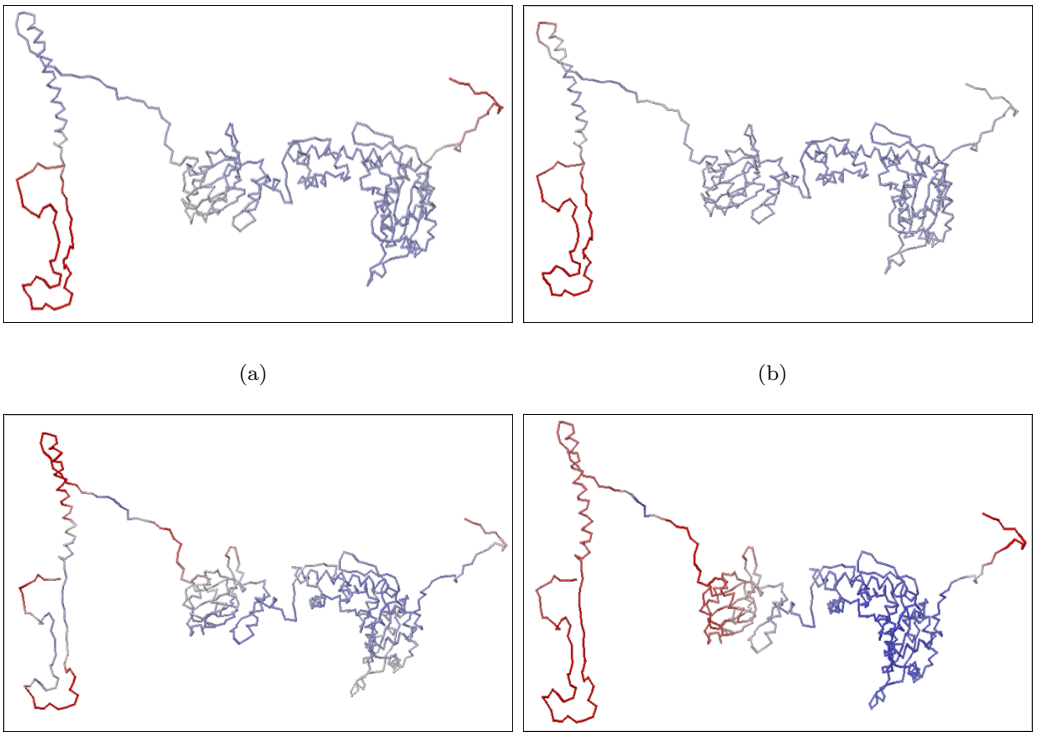
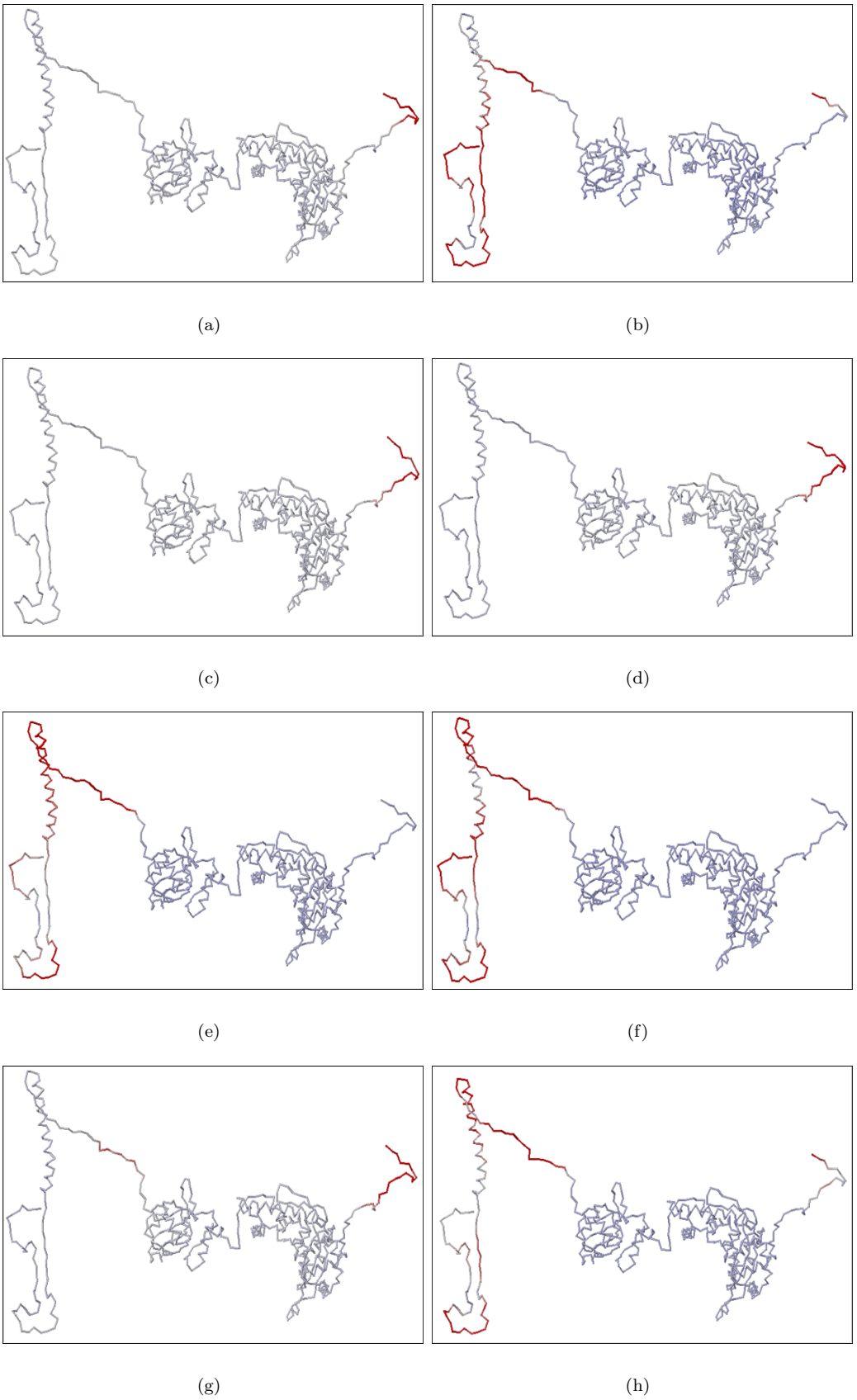
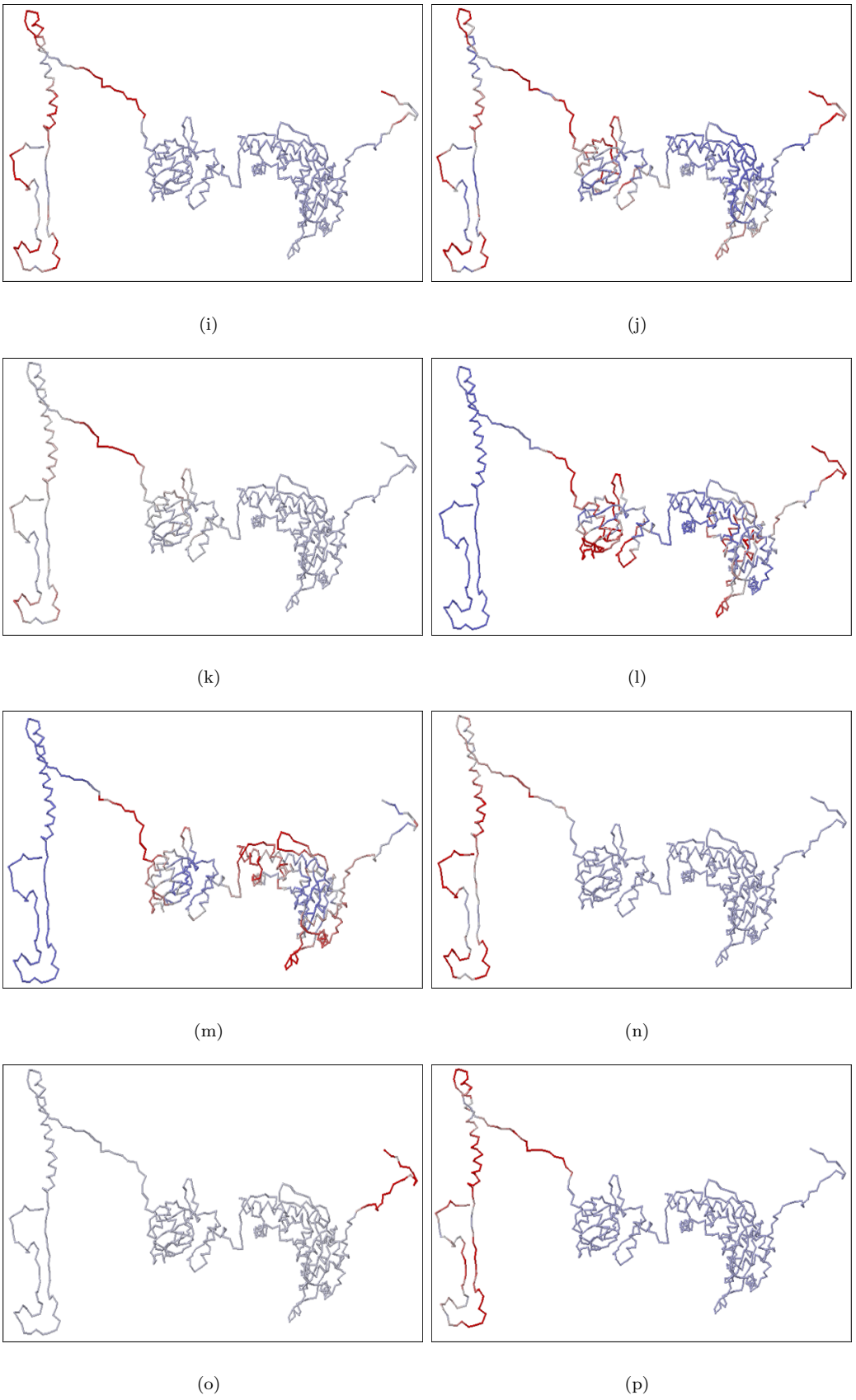


FIGURE 4.3: (A) ENMs of E1. Red color is showing the amino acid residues having highest degrees of fluctuation or elasticity in (a) Mode 1 (b) Mode 2 (c) Mode 3 (d) Mode 4



(a) (b) (c) (d) (e) (f) (g) (h)

FIGURE 4.3: (B) ENMs of E1. Red color is showing the amino acid residues having highest degrees of fluctuation or elasticity in (a) Mode 5 (b) Mode 6 (c) Mode 7 (d) Mode 8 (e) Mode 9 (f) Mode 10 (g) Mode 11 (h) Mode 12



(i) (j)

(k) (l)

(m) (n)

(o) (p)

FIGURE 4.3: (C) ENMs of E1. Red color is showing the amino acid residues having highest degrees of fluctuation or elasticity in (i) Mode 13 (j) Mode 14 (k) Mode 15 (l) Mode 16 (m) Mode 17 (n) Mode 18 (o) Mode 19 (p) Mode 20

4.2.1.4 Correlation Analysis of E1

We also had the correlation analysis of HPV proteins showing the correlation of all the amino acids of HPV proteins with each amino acid within the chain. Appendix A-21 shows the correlation analysis of E1 protein, red coloration shows the correlation among the residues, and blue exhibits anti correlation among the respective residues. Small fluctuations Re shown as red and large inter-residue fluctuations as blue. The distance fluctuations of E1 (Appendix A-23) depicting that the N terminal region shows larger inter residual fluctuations in most of the modes. Graph of B factors of amino acid residues of E1 in N terminal show higher degrees of fluctuations, as appear in figure 4.4. Higher degrees of fluctuations show higher ability of that region in maintaining interactions with other proteins. Deformation energies have also been found to be higher in amino acids of N terminal region (Appendix A-22). Higher the deformation energies, more is the capacity of that region to show flexibility. All these evidences validate that N terminal region of E1 is involved in stabilizing its interactions with different proteins.

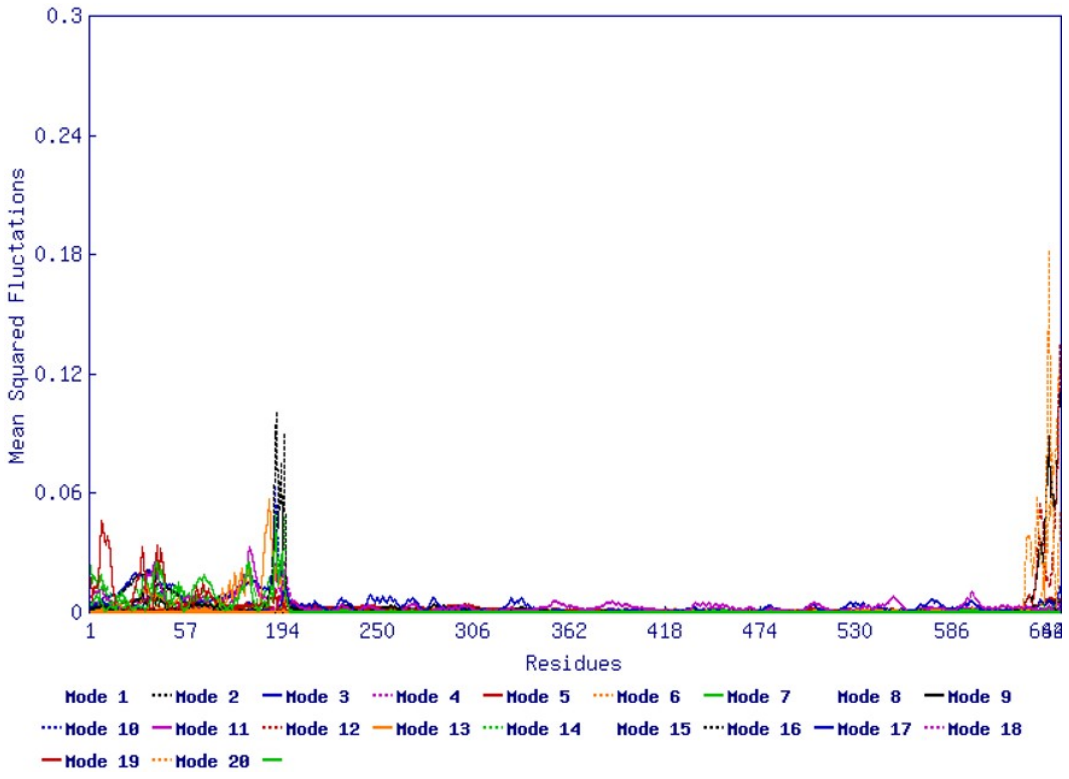


FIGURE 4.4: Graph of B factors for ENMs of E1 in 20 different modes. Mode 6 shows the highest value of B factors, reflecting the maximum degree of fluctuation

4.2.2 Analysis of ENMs of E2

E2 becomes functional after some post translational modifications. These include phosphorylation, acetylation and Sumoylation. Phosphorylation has been observed majorly at amino acid 253, 298, and 301, while minor sites of phosphorylation are 36, 76, 235 and 286 [288].

4.2.2.1 Motifs on E2

9 ELMs have been found involving these amino acids (Table 4.11), which experience phosphorylation in order to maintain an interaction with proteins as well as to perform their functions. This validates the role of these specific amino acids in phosphorylation of E2.

TABLE 4.11: ELMs found in the phosphorylated regions of E2.

ELMs	Amino acids Sequence/ Position
LIG_FHA_1	252-258 (HTTKLLH)
MOD_PKA_2	258-264 (HRDSVDS)
DOC_MAPK_MEF2A_6	68-77 (KALQAIELQL)
MOD_PIKK_1	231-237 (TEETQTT)
DOC_WW_Pin1_4	283-288 (SNTTPI)
LIG_FHA_1	283-289 (SNTTPIV), 284-290 (NTTPIVH)
MOD_CK1_1	283-289 (SNTTPIV)
MOD_N-GLC_1	283-288 (SNTTPI)
MOD_ProDKin_1	283-289 (SNTTPIV)

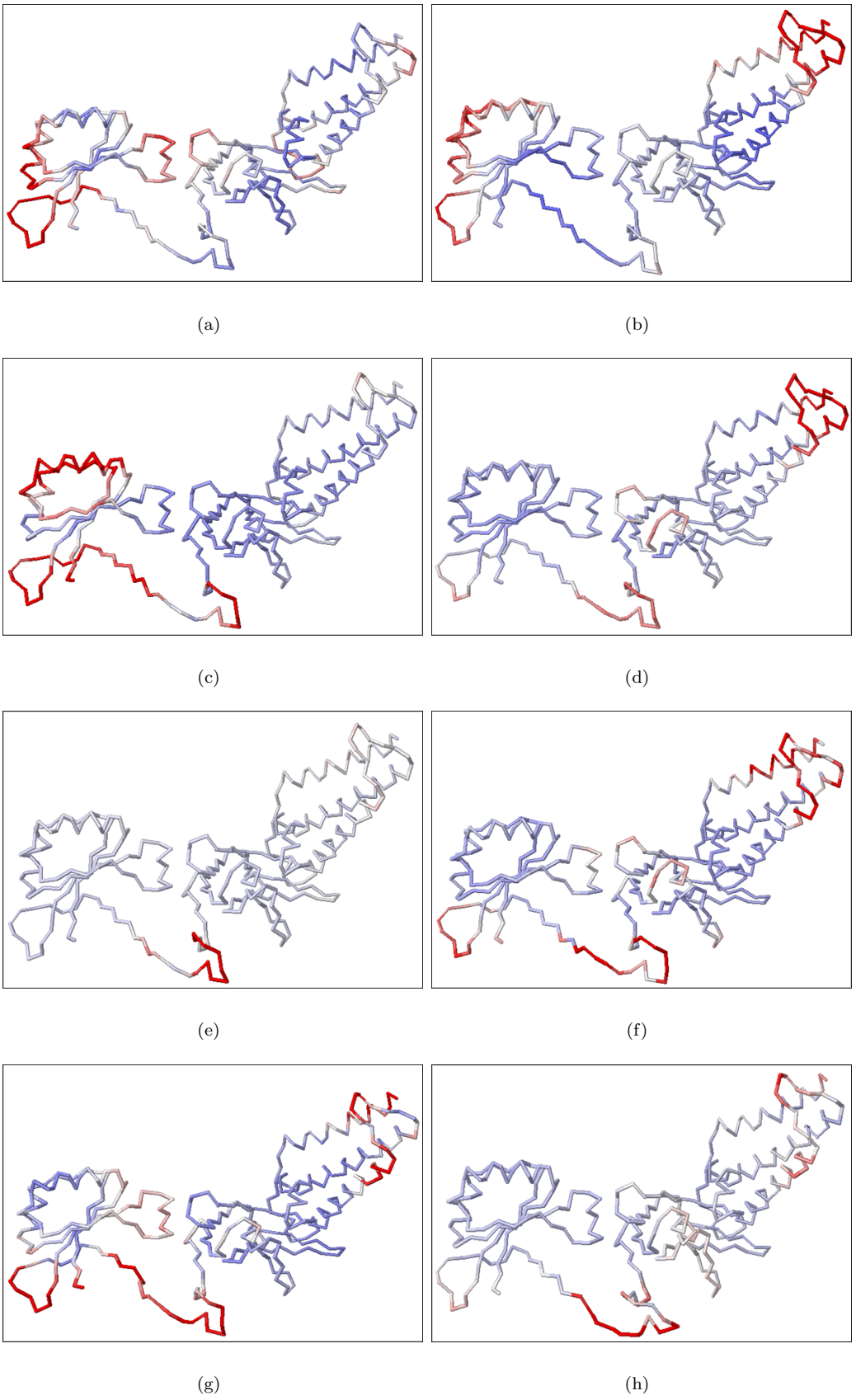
ELMs, LIG_FHA_1 and DOC_MAPK_gen_1 have been found to be associated with the Lysine residues present at 111 and 112 position, among which the Lysine at 111 position is most highly conserved among the HPV types. This Lysine plays a major role in Acetylation of E2 by p300, CBP, and pCAF. This motif interacts with FHA containing domains. Another protein transcription factor 19 also contains FHA domain and could play a role in enhancing capability of E2 during propagation of viral particle in the host cells. That has different lengths and amino acid sequences.

ELMs, LIG_FHA_1 and DOC_MAPK_gen_1 have been found to be associated with the Lysine residues present at 111 and 112 position, among which the Lysine at 111 position is most highly conserved among the HPV types. This Lysine plays a major role in Acetylation of E2 by p300, CBP, and pCAF. This motif interacts with FHA containing domains. Another protein transcription factor 19 also contains FHA domain and could play a role in enhancing capability of E2 during propagation of viral particle in the host cells. That has different lengths and amino acid sequences.

4.2.2.2 ENMS of E2

The HPVE2 is composed of two major domains, one full-length C-terminal domain comprised of 100 amino acids, which has capacity of binding with DNA also called "dimerization" domain, and an N-terminal domain of about 200 amino acids called "transactivation" domain. These domains are linked through a flexible linker region, called as the "hinge", that has different lengths and amino acid sequences, in different types of PVs. E2 helps the E1 helicase to initiate the replication by loading it on the origin of replication. The origin of replication has two binding sites, one for E1 and other for E2, and an AT rich region. The primary protein for replication is E! while E2 supports and enhances the activities of E1. A double hexameric helicase is converted from E1 as it is loaded, followed by displacement of E2. Nucleosomes are also displaced by E2 from the origin to facilitate repression [288].

In ENMs of E2, mode 1, 2, 4, 6, 7, 8, 9, and 10 were found to be having higher degrees of flexibility in N-terminal region. Table 4.12 shows the motifs present in flexible regions of E2. While C-terminal region was found to be highly fluctuating in modes 1, 2, 3, and 13. Hinge region was found to be most flexible in majority of modes (Figure 4.5), having higher values of B factors in graphs (Figure 4.6). This validates the previous findings of the flexibility of hinge regions [288]. The other modes does not show fluctuations in these particular regions. In ENMs of E2, mode 1, 2, 4, 6, 7, 8, 9, and 10 were found to be having higher degrees of flexibility in N-terminal region. The HPVE2 is composed of two major domains, one full-length C-terminal domain comprised of 100 amino acids.



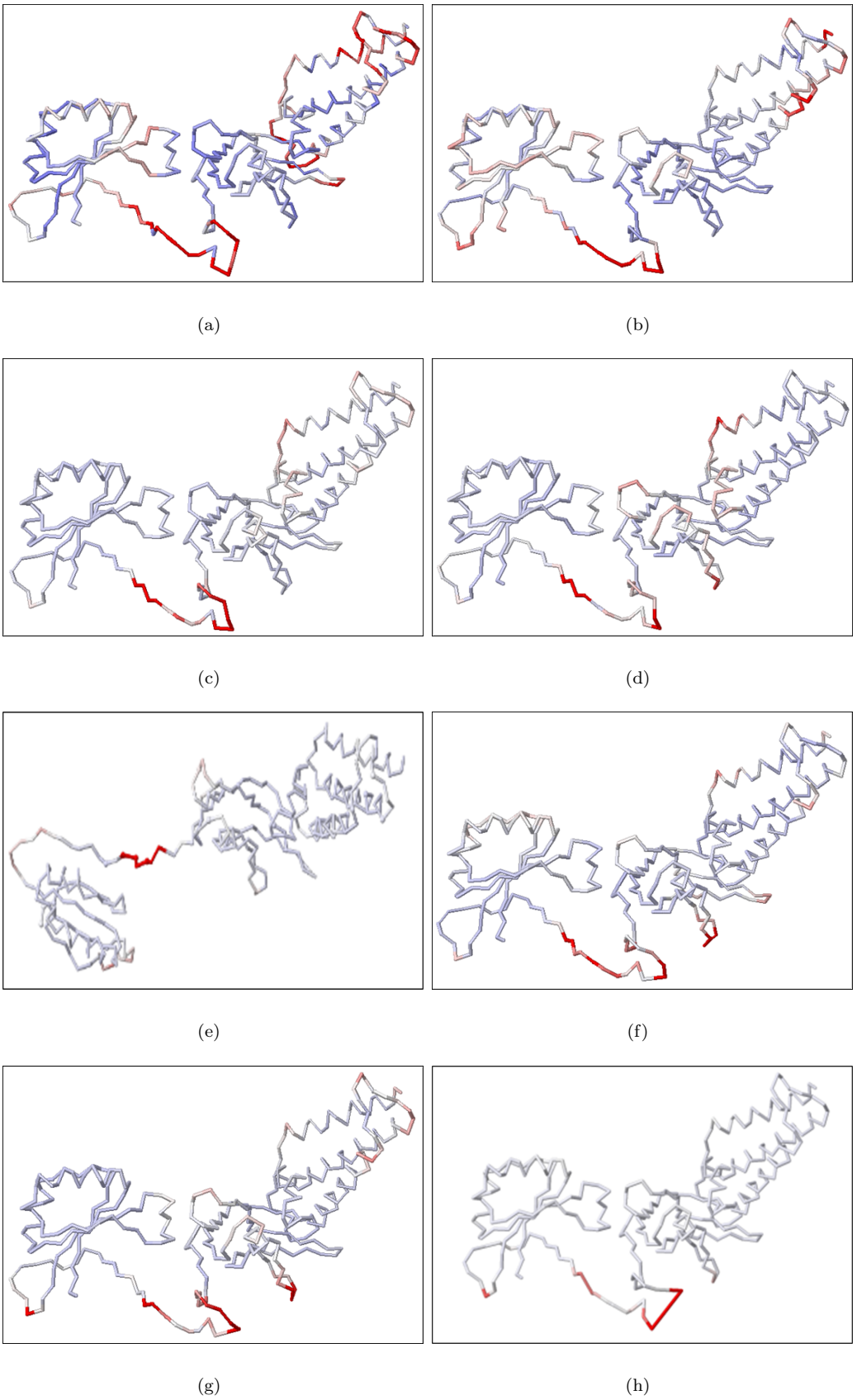
(a) (b)

(c) (d)

(e) (f)

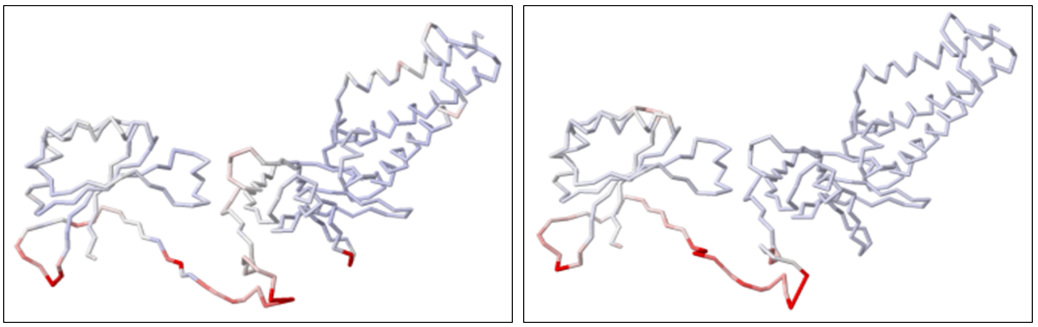
(g) (h)

FIGURE 4.5: (A) ENMs of E2. Red color is showing the amino acid residues having highest degrees of fluctuation or elasticity (a) Mode 1 (b) Mode 2 (c) Mode 3 (d) Mode 4 (e) Mode 5 (f) Mode 6 (g) Mode 7 (h) Mode 8

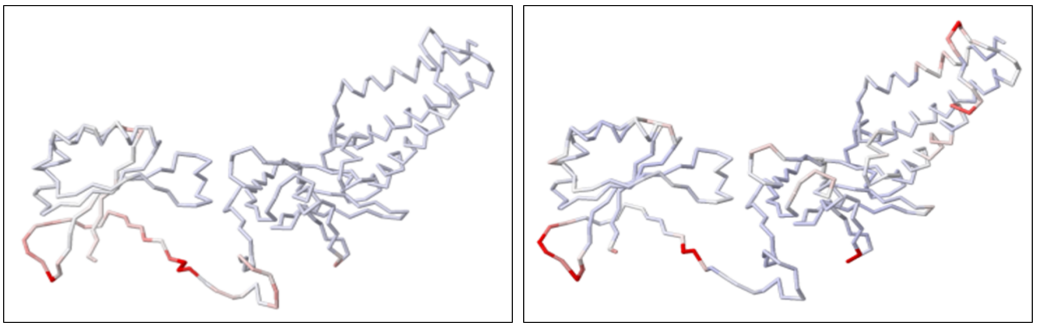


(a) (b) (c) (d) (e) (f) (g) (h)

FIGURE 4.5: (B) ENMs of E2. Red color is showing the amino acid residues having highest degrees of fluctuation or elasticity (a) Mode 9 (b) Mode 10 (c) Mode 11 (d) Mode 12 (e) Mode 13 (f) Mode 14 (g) Mode 15 (h) Mode 16



(i) (j)



(k) (l)

FIGURE 4.5: (B) ENMs of E2. Red color is showing the amino acid residues having highest degrees of fluctuation or elasticity (i) Mode 17 (j) Mode 18 (k) Mode 19 (l) Mode 20

4.2.2.3 Correlation Analysis of E2

Correlation analysis, deformation energies and distance fluctuations of E2 (Appendix A-24 to A-26) also provide evidences to have greater abilities of flexibility by these specific regions.

4.2.2.4 Hot Spot residues of E2

As the hinge regions play an important role in positioning of E2 for its interaction with DNA as well as recruiting E1 to DNA for replication. The amino acids found to be conserved throughout the different types of HPV include proline, serine, threonine, glycine and arginine residues. These amino acids could be predicted as hot spot residues of E2, thus becoming the potential targets for drugs in order to paralyze E2 to perform its functions. As the drugs will target the interaction of E2 with the host proteins and DNA by focusing on its hot spot residues, it will decrease the ability of E2 to interact with host machinery for progression of infection and carcinogenesis.

TABLE 4.12: 20 modes of E2 with the motifs they have

S. No.	Modes	Amino acid Residue Position in flexible regions	Human Motifs
1.	Mode 1:	206-217	DOC_WW_Pin1_4
		248-251	LIG_SH3_3
		274-283	MOD_N-GLC_1
			MOD_ProDKin_1
			Nend_UBRbox_2
2.	Mode 2:	1-6 DEG_	MOD_GSK3_1
		120-127	MOD_CK1_1
		179-185	MOD_CK2_1
		191-202	MOD_GSK3_1
		213-220	CLV_C14_Caspase3-7
		250-265	LIG_FHA_1
		364-365	MOD_PKA_2
			LIG_LIR_Gen_1
3.	Mode 3:	1-6	DEG_Nend_UBRbox_2
		48-52	LIG_LIR_Gen_1
		120-124	DOC_MAPK_MEF2A_6
		179-202	MOD_CK1_1
		211-220	MOD_PIKK_1
		237-248	LIG_LIR_Gen_1
		363-365	MOD_GSK3_1
			MOD_CK1_1
4.	Mode 4:	199-208	MOD_GSK3_1
		214-216	CLV_C14_Caspase3-7
		227-230	LIG_FHA_1
		253-265	MOD_GlcNHglycan
			MOD_PKA_2
			MOD_CK1_1
5.	Mode 5:	198-218	MOD_GSK3_1
		228-244	DOC_WW_Pin1_4
		282-291	LIG_FHA_1
		364-365	MOD_N-GLC_1
			MOD_ProDKin_1
			LIG_LIR_Gen_1
6.	Mode 6:		DEG_Nend_UBRbox_2
		1-6	LIG_LIR_Gen_1
		43-65	LIG_SUMO_SIM_par_1
			MOD_GSK3_1
7.	Mode 7:	1-3	DEG_Nend_UBRbox_2
		48-61	LIG_LIR_Gen_1
		118-130	MOD_GSK3_1
		176-181	MOD_CK1_1
		190-192	MOD_CK2_1
		196-199	MOD_GSK3_1

S. No.	Modes	Amino acid Residue Position in flexible regions	Human Motifs
8.	Mode 8:	2-13	LIG_LIR_Gen_1
		52-66	LIG_SUMO_SIM_par_1
		176-181	MOD_CK1_1
		197-199	MOD_CK2_1
			MOD_GSK3_1
9.	Mode 9:	198-200	MOD_CK1_1
			MOD_CK2_1
			MOD_GSK3_1
10.	Mode 10:	176-180	MOD_CK1_1
		196-197	MOD_CK2_1
		199-200	MOD_GSK3_1
11.	Mode 11:	124-126	MOD_CK1_1
		196-203	MOD_GSK3_1
12.	Mode 12:	215-217	LIG_FHA_1
		256-259	MOD_PKA_2
13.	Mode 13:	198-204	MOD_CK1_1
			MOD_CK2_1
			MOD_GSK3_1
14.	Mode 14:	215-217	LIG_FHA_1
		225-227	
			LIG_SH2_GRB2
15.	Mode 15:	16-22	MOD_N-GLC_1
		122-126	MOD_CK1_1
		197-200	MOD_CK2_1
			MOD_GSK3_1
16.	Mode 16:	198-200	MOD_CK1_1
			MOD_CK2_1
			MOD_GSK3_1
17.	Mode 17:	198-204	MOD_CK1_1
			MOD_CK2_1
			MOD_GSK3_1
18.	Mode 18:	61-63	LIG_SUMO_SIM_par_1
		162-164	LIG_SH3_3
		197-198	MOD_CK1_1
		202-203	MOD_GSK3_1
19.	Mode 19:	202-203	LIG_SH3_3
		212-216	MOD_GSK3_1
		255-260	LIG_FHA_1
			MOD_PKA_2
20.	Mode 20:	212-218	LIG_FHA_1
		255-260	MOD_PKA_2
		364-365	LIG_LIR_Gen_1
21.	Mode 20:	212-218	LIG_FHA_1
		255-260	MOD_PKA_2
		364-365	LIG_LIR_Gen_1

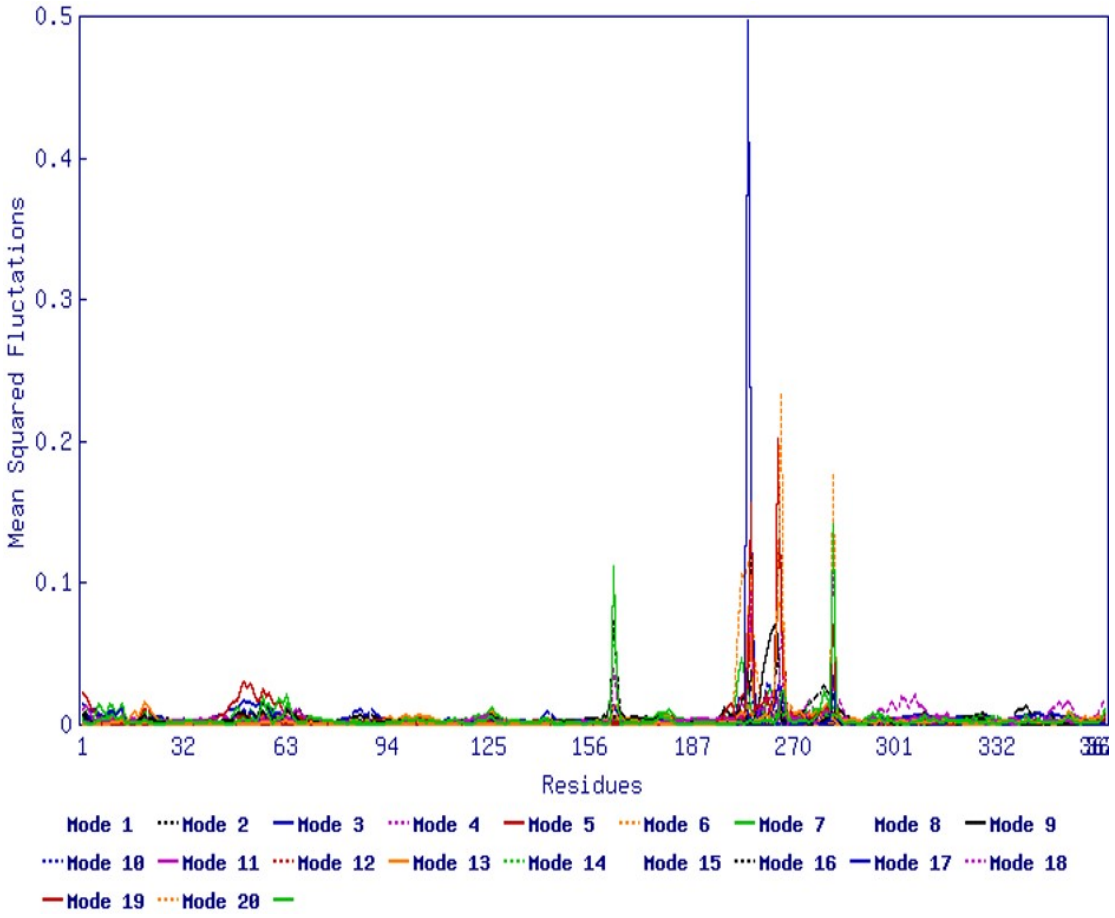


FIGURE 4.6: Graphs of B factors for 20 modes of E2. Mode 3 has highest B factors showing the maximum degree of fluctuations on the region between amino acid residue 200 and 250

4.2.3 Analysis of ENMs of E4

E4 has been reported to interact with cytokeratin [289] and different kinases including Cyclin B/Cdk1 [290], Cyclin a/Cdk2, Cyclin E/Cdk2 [291], p42 MAPK, protein kinase C [292], Protein kinase A [293], and SR protein kinase 1. Many E4 proteins are characterized by an N-terminal leucine-rich LLXLL motif. Initially it was found that if these residues are deleted from E4 of HPV1 and HPV16, it results in the inability of these mutated proteins to localize to the keratin based cytoskeleton. Latterly, in a research it was found that some other regions of E4 protein also play vital role for its interaction with cytokeratin. These include residues 61 to 92 of C-terminal region, in which the most conserved and important are residues 84 to 92 (LTVIVTLHP). Further analysis showed that in particular valines 86 and 88 [294].

4.2.3.1 Motifs on E4

In this study these findings were validated on the basis of presence of specific motifs in E4 in these particular regions (Table 4.13), both involving in interactions of E4 with cytokeratin. 9 different motifs were found in E4 which interact with their respective domains operating through the specific kinases. These specific kinases include MAP kinases, kinases of NEK and PIKK family.

TABLE 4.13: ELMs found in N-terminal and C terminal regions of E4 protein involved in interaction with Cytokeratins

ELMs	Amino acids Sequence/Position
DOC_MAPK_MEF2A_6	81-90 (KDGLTVIVTL)
LIG_FHA_1	83-89 (GLTVIVT)
LIG_LIR_Gen_1	64-70 (ETQWTVL), 63-70 (TQWTVL)
LIG_SUMO_SIM_anti_2	82-89 (DGLTVIVT)
LIG_SUMO_SIM_par_1	84-89 (LTVIVT)
MOD_CK1_1	60-66 (SCCTETQ)
MOD_GSK3_1	65-72 (TQWTVLQS), 70-77 (LQSSLHLT), 82-89 (DGLTVIVT)
MOD_NEK2_1	70-75 (LQSSLH), 74-79 (LHLTAH)
MOD_PIKK_1	62-68 (CTETQWT)
LIG_eIF4E_1	10-16 (YPLLKLL)

10 motifs have been found in E4 which interact with Protein kinases domain, thus facilitating the interaction of E4 with different kinases to phosphorylate or interact with keratins (Table 4.14).

TABLE 4.14: ELMs found in E4 with their CDs

S. No.	ELMs	CDs	Amino acids Sequence/Position
1.	CLV_NRD_NRD_1	Peptidase _M16	40-42 (RRL)
2.	DOC_CKS1_1	CKS	21-26 (PTTPPR), 52-57 (PETPAT), 55-60 (PATPLS)

3.	DOC_MAPK_MEF2A_6	Pkinase	81-90 (KDGLTVIVTL)
4.	MOD_CDK_SPxK_1	Pkinase	20-26 (WPTTPPR)
5.	MOD_CK1_1	Pkinase	60-66 (SCCTETQ)
6.	MOD_GSK3_1	Pkinase	65-72 (TQWTVLQS), 70-77 (LQSSLHLT), 82-89 (DGLTVIVT)
7.	MOD_NEK2_1	Pkinase	70-75 (LQSSLH), 74-79 (LHLTAH)
8.	MOD_PKA_1	Pkinase	40-46 (RRLSSDQ)
9.	MOD_PKA_2	Pkinase	40-46 (RRLSSDQ)
10.	MOD_ProDKin_1	Pkinase	20-26 (WPTTPPR), 29-35 (PKPSPWA), 48-54 (QSQTPET), 5 1-57 (TPETPAT), 54-60 (TPATPLS)
11.	DOC_PP2A_B56_1	Armadillo type fold	59-65 (LSCCTET)
12.	DOC_WW_Pin1_4	WW	20-25 (WPTTPP), 29-34 (PKPSPW), 48-53 (QSQTPE), 5 1-56 (TPETPA), 54-59 (TPATPL)
13.	LIG_WW_3	WW	23-27 (TPPRP)
14.	LIG_BIR_II_1	BIR	1-5 (MADPA)
15.	LIG_BIR_III_1	BIR	1-5 (MADPA)
16.	LIG_eIF4E_1	IF4E	10-16 (YPLLKLL)
17.	LIG_FHA_1	FHA	83-89 (GLTVIVT)
18.	LIG_LIR_Gen_1	Atg8	64-70 (ETQWTVL), 63-70 (TQWTVL)
19.	LIG_SH3_1	SH3_1, SH3_9	30-36 (KPSPWAP)

20.	LIG_SH3_2	SH3_1, SH3_9	21-26 (PTTPPR), 33-36 (KPSPWAP)
21.	LIG_SH3_3	SH3_1, SH3_9	18-24 (STWPTTP), 21-27 (PTTPPRP), 30-36 (KPSPWAP), 49-55 (SQTPETP), 52-58 (PETPATP)
22.	LIG_SUMO_SIM_anti_1	Rad60-SLD	82-89 (DGLTVIVT)
23.	LIG_SUMO_SIM_par_1	Rad60-SLD	84-89 (LTVIVT)
24.	MOD_LATS_1	Pkinase_C	39-45 (HRRLSSD)
25.	MOD_PIKK_1	PI3_PI4_ kinase	62-68 (CTETQWT)
26.	TRG_ENDOCYTIC_2	Adap_comp _sub	10-13 (YPLL)
27.	TRG_ER_diArg_1	WD40	40-42 (RRL)
28.	TRG_NLS_MonoExtC_3	Arm	36-41 (PKKHRR)
29.	TRG_NLS_MonoExtC_4	Arm	36-41 (PKKHRR)

4.2.3.2 ENMs of E4

The graph of B factors of E4 show highest degree of fluctuations at position 1-10, 25-35, 60-70, and at C terminal end (Figure 4.7). Most of these regions are those containing the motifs involved in interaction with kinases (Table 4.14). This validates the interaction of E4 with kinases and gives the detailed dynamics of these regions involved in interactions, at residual level, thus making these sites potential targets of drugs, in order to prevent propagation of viral genome inside the human host. Table 4.15 is showing the motifs present in flexible regions of E4.

4.2.3.3 Correlation Analysis of E4

ENMs, correlation analysis, and deformation energies of E4 (Appendix A-27 to A-29) also show the higher flexibility and co-ordination of amino acids in these regions (Figure 4.8) In Correlation analysis, deformation energies and distance fluctuations the higher degrees of fluctuations have been observed in these specific regions which show ability to interact with kinases.

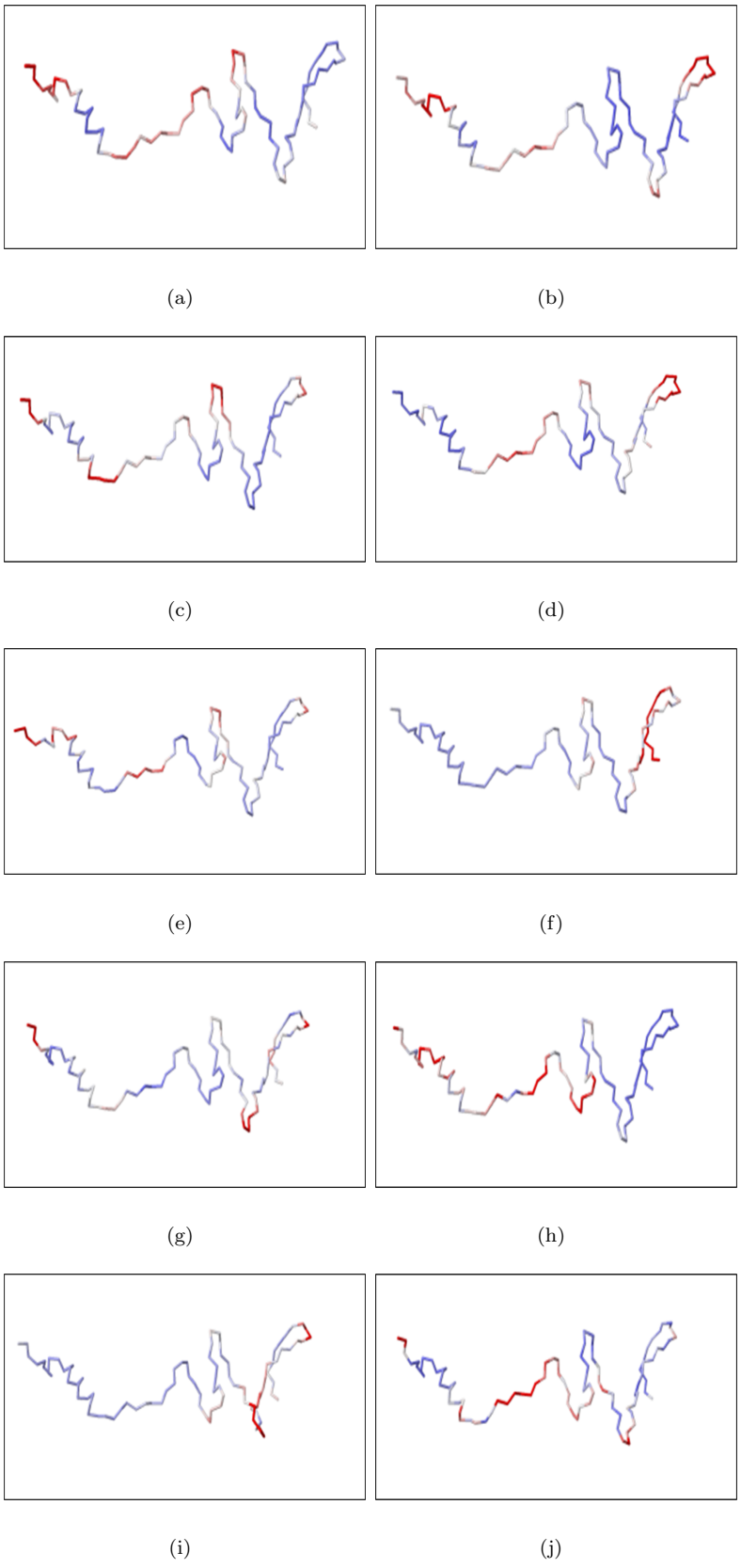


FIGURE 4.7: (A) ENMs of E4. Red color is showing the amino acid residues having highest degrees of fluctuation or elasticity in (a) Mode 1 (b) Mode 2 (c) Mode 3 (d) Mode 4 (e) Mode 5 (f) Mode 6 (g) Mode 7 (h) Mode 8 (i) Mode 9 (j) Mode 10

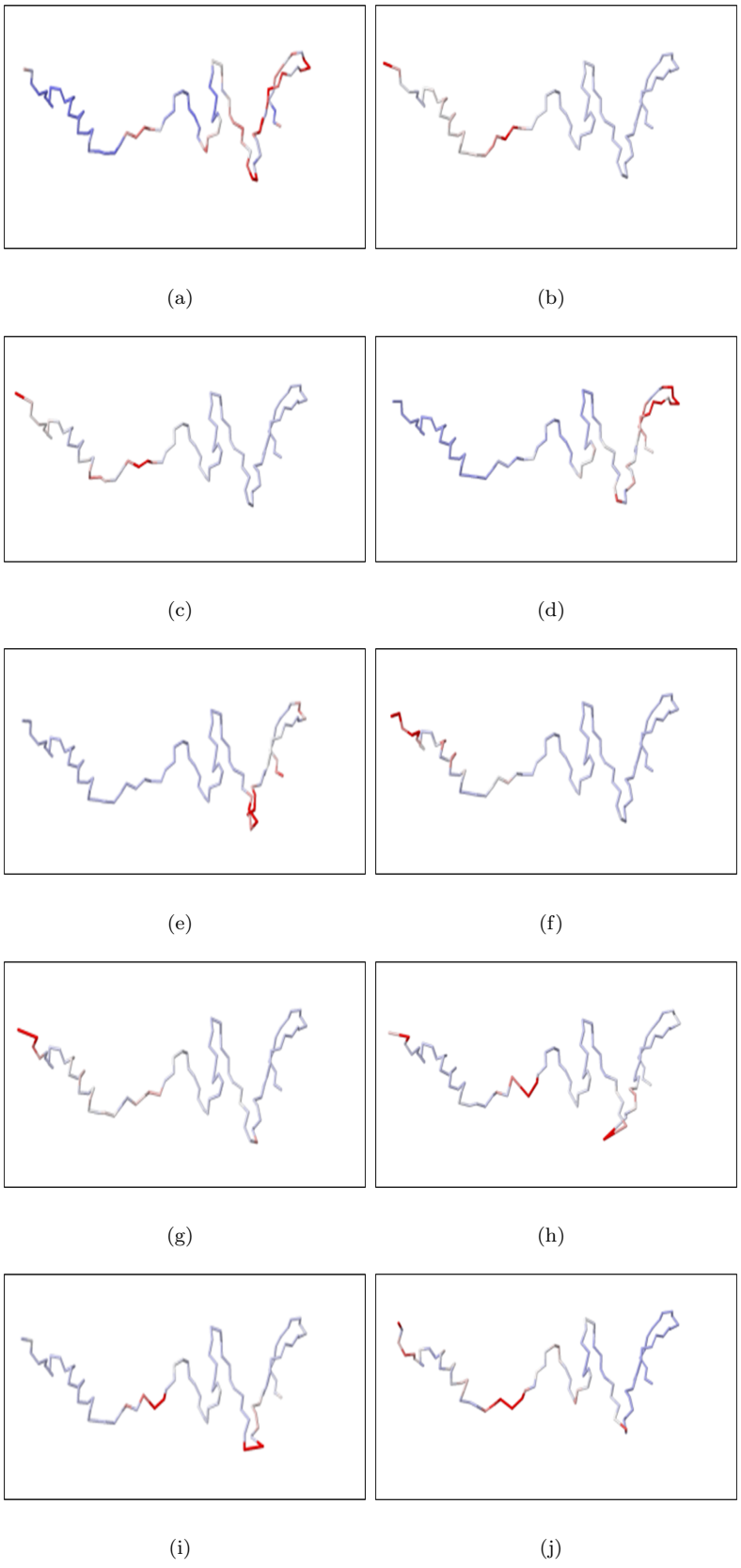


FIGURE 4.7: (B) ENMs of E4. Red color is showing the amino acid residues having highest degrees of fluctuation or elasticity in (a) Mode 1 (b) Mode 2 (c) Mode 3 (d) Mode 4 (e) Mode 5 (f) Mode 6 (g) Mode 7 (h) Mode 8 (i) Mode 9 (j) Mode 10

TABLE 4.15: Motifs present in flexible regions of E4 in 20 Modes

S. No.	Modes	Amino acid Residue Position	Human Motifs
1.	Mode 1:	1-6 8-11 24-36 51-54	DOC_CKS1_1 DOC_WW_Pin1_4 DOC_WW_Pin1_4 LIG_BIR_II_1 LIG_BIR_III_1 LIG_eIF4E_1 LIG_LIR_Gen_1 LIG_LIR_Gen_1 LIG_SH3_1 LIG_SH3_2 LIG_SH3_3 LIG_WW_3 MOD_CDK_SPxK_1 MOD_PIKK_1 MOD_ProDKin_1 TRG_ENDOCYTIC_2
2.	Mode 2:	6-13 27-29 63-65 79-86	LIG_LIR_Gen_1 LIG_eIF4E_1 TRG_ENDOCYTIC_2 LIG_LIR_Gen_1 MOD_CK1_1 MOD_PIKK_1 LIG_SUMO_SIM_anti_2 LIG_SUMO_SIM_par_1 MOD_GSK3_1 MOD_GSK3_1
3.	Mode 3:	1-4 21-25 51-55 81-82	DOC_CKS1_1 DOC_MAPK_MEF2A_6 DOC_WW_Pin1_4 LIG_BIR_II_1 LIG_BIR_III_1 LIG_SH3_2 LIG_SH3_3 LIG_WW_3 MOD_CDK_SPxK_1 MOD_GSK3_1 MOD_PIKK_1 MOD_ProDKin_1
6.	Mode 6:	71-73 83-92	DOC_MAPK_MEF2A_6 LIG_FHA_1 LIG_SUMO_SIM_anti_2 MOD_GSK3_1 MOD_NEK2_1
7.	Mode 9:	60-65 80-83	DOC_MAPK_MEF2A_6 LIG_SUMO_SIM_anti_2 MOD_GSK3_1

			LIG_SH3_1
			LIG_SH3_2
			LIG_SH3_3
			LIG_SUMO_SIM_anti_2
8.	Mode 4:	28-32 77-86	LIG_SUMO_SIM_par_1
			LIG_WW_3
			MOD_GSK3_1
			MOD_ProDKin_1
			MOD_NEK2_1
			DOC_CKS1_1
			DOC_MAPK_MEF2A_6
			DOC_WW_Pin1_4
			LIG_BIR_II_1
			LIG_BIR_III_1
		1-5	LIG_eIF4E_1
		8-13	LIG_LIR_Gen_1
9.	Mode 5:	27-31	LIG_LIR_Gen_1
		52-54	LIG_SH3_1
		80-82	LIG_SH3_3
			LIG_WW_3
			MOD_CDK_SPxK_1
			MOD_ProDKin_1
			TRG_ENDOCYTIC_2
			DOC_MAPK_MEF2A_6
			DOC_PP2A_B56_1
			LIG_BIR_II_1
		1-3	LIG_BIR_III_1
10.	Mode 7:	62-67	LIG_LIR_Gen_1
		80-82	MOD_CK1_1
			MOD_GSK3_1
			MOD_PIKK_1
			LIG_LIR_Gen_1
			LIG_eIF4E_1
			LIG_LIR_Gen_1
		6-12	CLV_NRD_NRD_1
11.	Mode 8:	14-15	MOD_LATS_1
		40-45	MOD_PKA_1
			TRG_ER_diArg_1
			TRG_NLS_MonoExtC_3
			TRG_NLS_MonoExtN_4
			DOC_PP2A_B56_1
			MOD_CK1_1
		62-64	MOD_PIKK_1
		71-72	MOD_NEK2_1
12.	Mode 11:	75-76	MOD_GSK3_1
		80-83 DOC_	MAPK_MEF2A_6
			LIG_SUMO_SIM_anti_2
			MOD_GSK3_1

			DOC_CKS1_1
			DOC_WW_Pin1_4
			LIG_BIR_II_1
			LIG_BIR_III_1
			LIG_LIR_Gen_1
			LIG_SH3_1
		1-3	LIG_SH3_2
13.	Mode 10:	21-22	LIG_SH3_2
		26-34	LIG_SH3_3
		63-65	LIG_WW_3
			MOD_CDK_SPxK_1
			MOD_CK1_1
			MOD_GSK3_1
			MOD_PIKK_1
			MOD_ProDKin_1
			LIG_BIR_II_1
		1-3	LIG_BIR_III_1
14.	Mode 12:		DOC_WW_Pin1_4
		28-30	MOD_ProDKin_1
			LIG_BIR_II_1
		1-3	LIG_BIR_III_1
15.	Mode 13:		DOC_WW_Pin1_4
		28-30	MOD_ProDKin_1
			DOC_PP2A_B56_1
			MOD_CK1_1
			MOD_PIKK_1
		62-64	DOC_MAPK_MEF2A_6
16.	Mode 14:	75-78	LIG_FHA_1
		80-84	MOD_GSK3_1
			LIG_SUMO_SIM_anti_2
			MOD_NEK2_1
			DOC_PP2A_B56_1
		62-63	LIG_LIR_Gen_1
17.	Mode 15:	65-69	MOD_CK1_1
		91-92	MOD_GSK3_1
			MOD_PIKK_1
18.	Mode 16:	1-6	LIG_BIR_II_1
			LIG_BIR_III_
19.	Mode 17:	1-4	LIG_BIR_II_1
			LIG_BIR_III_1
			DOC_WW_Pin1_4
			LIG_SH3_3
		1-2	LIG_WW_3
20.	Mode 20:	26-30	MOD_ProDKin_1
		61-63	DOC_PP2A_B56_1
			MOD_CK1_1
			MOD_PIKK_1

			LIG_BIR_II_1
			LIG_BIR_III_1
			DOC_WW_Pin1_4
			LIG_SH3_1
		2-4	LIG_SH3_3
21.	Mode 18:	29-32	MOD_ProDKin_1
		63-65	DOC_PP2A_B56_1
			LIG_LIR_Gen_1
			MOD_CK1_1
			MOD_PIKK_1
			DOC_WW_Pin1_4
			LIG_SH3_1
			LIG_SH3_3
			MOD_ProDKin_1
22.	Mode 19:	29-32	DOC_PP2A_B56_1
		63-66	LIG_LIR_Gen_1
			MOD_CK1_1
			MOD_GSK3_1
			MOD_PIKK_1

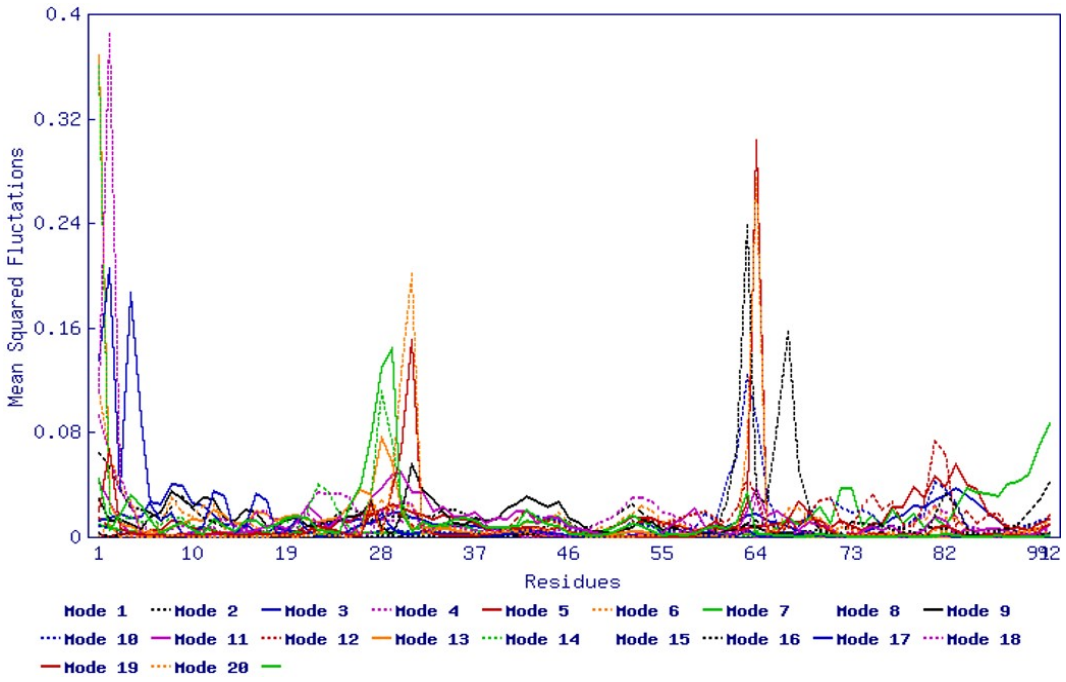


FIGURE 4.8: Graph of B factors of ENMs for E4

4.2.4 Analysis of ENMs of E5

Several animal and human papillomaviruses encode short trans-membranous protein E5. Such proteins exhibit transformation activity in animals and cultured cells, and apparently they have also some role in the virus life cycle. Many cellular proteins activities are thought to be modulated by E5 protein [295].

4.2.4.1 Motifs on E5

In our work as mentioned earlier that the ELM source only maps and finds those motifs which are present on the surface of a protein. Motif matches which are buried in inner folded cores of domains of globular proteins are not reliable candidates. When we mapped E5 amino acid sequence on ELMs. We found only one significant motif on the surface region. But by applying the structural filter we got able to view the motifs lying in hidden regions. As in case of E5 due to involvement of its helices in interactions these motifs could play a major role. We found 22 motifs on E5 (Table 4.16)

TABLE 4.16: ELMs found in alpha helices of E5

ELM	**Amino Acid Position/ Sequence**	**CDs**
CLV_PCSK_SKI1_1	30-34	Peptidase S8
DEG_APCC_DBOX_1	29-37	WD40
DOC_CYCLIN_RxL_1	27-38	Cyclin_N
DOC_MAPK_JIP1_4	30-36	Pkinase
DOC_MAPK_MEF2A_6	58-66	Pkinase
DOC_SPAK_OSR1_1	79-83	OSR1_C
LIG_14-3-3_CterR_2	79-83	14-3-3
LIG_eIF4E_1	68-74	eIF4e
LIG_FHA_1	7-13	FHA
	38-44	
LIG_LIR_Gen_1	37-42	Atg8
	38-42	
LIG_NRBOX	44-50	Hormone recep
LIG_Pex14_2	15-19	SH3_1
		Px14 N
LIG_SH2_PTP2	68-71	SH2
LIG_SH2_STAT5	39-42	SH2
	63-66	
	68-71	

LIG_SH3_3	64-70	SH3
LIG_SUMO_SIM_anti_2	40-47	Rad60_SLD
	41-47	
LIG_SUMO_SIM_par_1	30-35	Rad60_SLD
	32-38	
MOD_CK1_1	35-41	Pkinase
	37-43	
MOD_GSK3_1	3-10	Pkinase
	34-41	
MOD_NEK2_1	32-37	Pkinase
	34-39	
	49-54	
MOD_Plk_4	73-78	PI3_PI4 kinase
	7-13	
	38-44	
TRG_ENDOCYTIC_2	39-42	Adap_comp_sub
	63-66	
	68-71	

TABLE 4.17: Proteins having validated interactions with HPV16 E5 (Reported in literature and taken from VirusHost database).

EGF Receptor	**P38 MAP kinase**
Erk1/2	Phospholipase C-gamma 1
AKT	ErB4
Cyclooxygenase-2	c-jun, c-fos
VEGF	EVER1 AND EVER2
c-Cbl	ZNt-1
16K vacuolar-ATPase subunit	Bap 31
MHC class I antigens	Calnexin
MHC class II antigens	Calpactin 1
CD1d	Karyopherin beta3

Connexin 43	Endothelin-1 receptor
Fas	Bcl-2
Cyclin dependent kinase inhibitor p21	Bax
Cyclin dependent kinase inhibitor p27	

HPV E5 protein has the ability to modulate the activity of a large number of proteins (Table 4.17), despite of having smaller size, and hydrophobic nature. It lack the highly soluble and globular domains which paly main role in mediation of protein-protein interactions. Preferably, an alternative mechanism is adopted by the E5 in order to engage their protein targets. An interface for stabilizing protein-protein interactions is provided by the hydrophobic region of the E5. In the hydrophobic environment of membrane, the amide of main chain and carboxyl groups of the transmembrane domains get attached with each another through hydrogen bonds thus minimizing the membrane insertion energetic cost, resulting the creation of an α-helix. Thus, despite of their smaller size and infrequent composition, the E5 proteins accept an energetically promising, well-formed assembly in cell membranes. In these arrangements the amino acid R groups are outwardly displayed on the surface of the helical axis. By this the well exposed side chains can easily interact with other molecules in a sequence specific way, within the cell membrane. With the help of these interactions, E5 can form homo-oligomers, or with other cellular proteins spanning transmembrane segments, ultimately creating protein complexes [295].

The cellular functions of an organism rely on signaling networks, which are stabilized majorly by protein kinases. In eukaryotes, the mitogen-activated protein kinases (MAPKs) are among the most important and most conserved members of this family of protein kinases. Studies have shown that one of the mechanisms by which 16HPVE5 stimulates EGFR signaling pathways, is based on its mechanism of affecting MAPK. In our study we found two motifs DOC_MAPK_JIP1_4 and DOC_MAPK_MEF2A_6, which interact with proteins through the protein kinases domains based on MAP kinases. This validates the possibility of E5 to interact through MAPK. E5 has also been reported to interact with p38, ERK1/2, c-jun

amino-terminal kinase, which are members of MAPK family. Our study also validates these interactions to occur through these motifs. Additionally, 4 motifs were also found which interact with protein kinases domains. These motifs could also play role in regulating other activities of E5 involving protein kinases. Oncogenesis has been shown to be promoted by abnormal synthesis and functioning of proteins of the PI3K-Akt pathway. The proteins affected by E5 also include Akt [295].This study has predicted the motif of MOD_Plk_4 interacting with the domains undergoing phosphorylation by PI3 and PI4. Same is the mechanism adopted by Akt proteins to operate the signaling pathways. In inflammatory process, Arachidonic acid is converted to thromboxane and prostaglandin by a key enzyme Cyclooxygenase (COX). COX-1 and COX-2 are two isoforms of COX. Chronic inflammation interposed by COX-2 plays some role in cancer progression and carcinogenesis. It is caused by various factors including bacterial infections and chemical irritants. MAPK activity is increased by Prostaglandin E2, that is the product of COX 2. As this mechanism of oncogenicity is associated with viral infection that is HPV. The E5 of HPV through its MAPK motifs could possibly increase the MAPK activity thus promoting the infection towards cervical cancer. E5 interacts with many cyclin dependent kinase inhibitors including P38 MAP kinase, p27, Phospholipase C-gamma 1, ZNt-1, ErB4, c-fos, c-jun, EVER1, EVER2, Bap 31, Calnexin, Calpactin 1, Karyopherin beta3 and Endothelin-1 receptor. E5 has a motif DOC_CYCLIN_RxL_1 at position 27-38. Here we predict the involvement of this motif in stabilizing interaction of E5 with these Cyclin dependent kinase inhibitors as all of these have CYCLIN_N domain to interact, with which E5 could also interact through motif DOC_CYCLIN_RxL_1 [295].

4.2.4.2 ENMs of E5

In elastic network modeling we found the highest values of B factors in regions of position 30-45 (Figure 4.9). This spans the part of E5 comprising of the motifs involved in interaction with MAPK, protein kinases and Cyclin based domains. This again validates the role of these motifs in establishing the already reported interactions with different proteins, as discussed above. The ENMs also show higher degrees of fluctuations in these regions (Figure 4.10).

4.2.4.3 Correlation Analysis of E5

These regions have shown higher values of B factors (Figure 4.9), higher correlations (Appendix A-30), deformation energies (Appendix A-31) and distance fluctuations (Appendix A-32) among their residual amino acids at these specified positions involved in stabilizing these interactions.

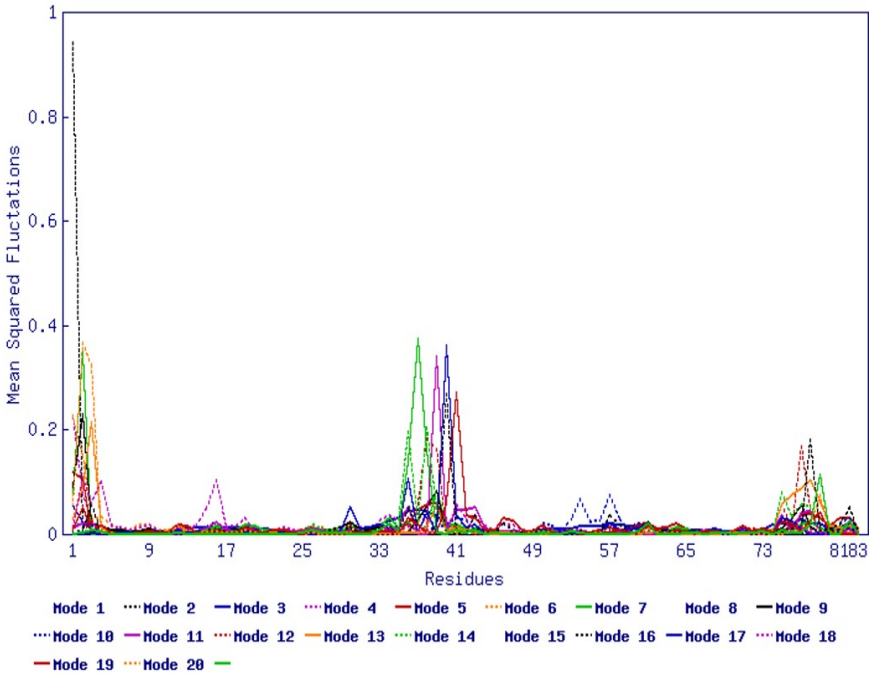


FIGURE 4.9: Graph of B factors E5

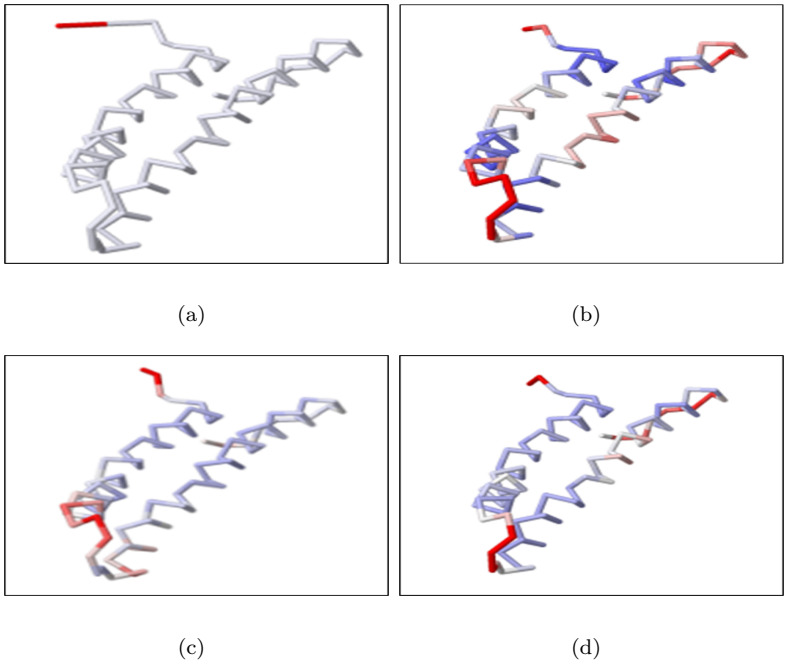
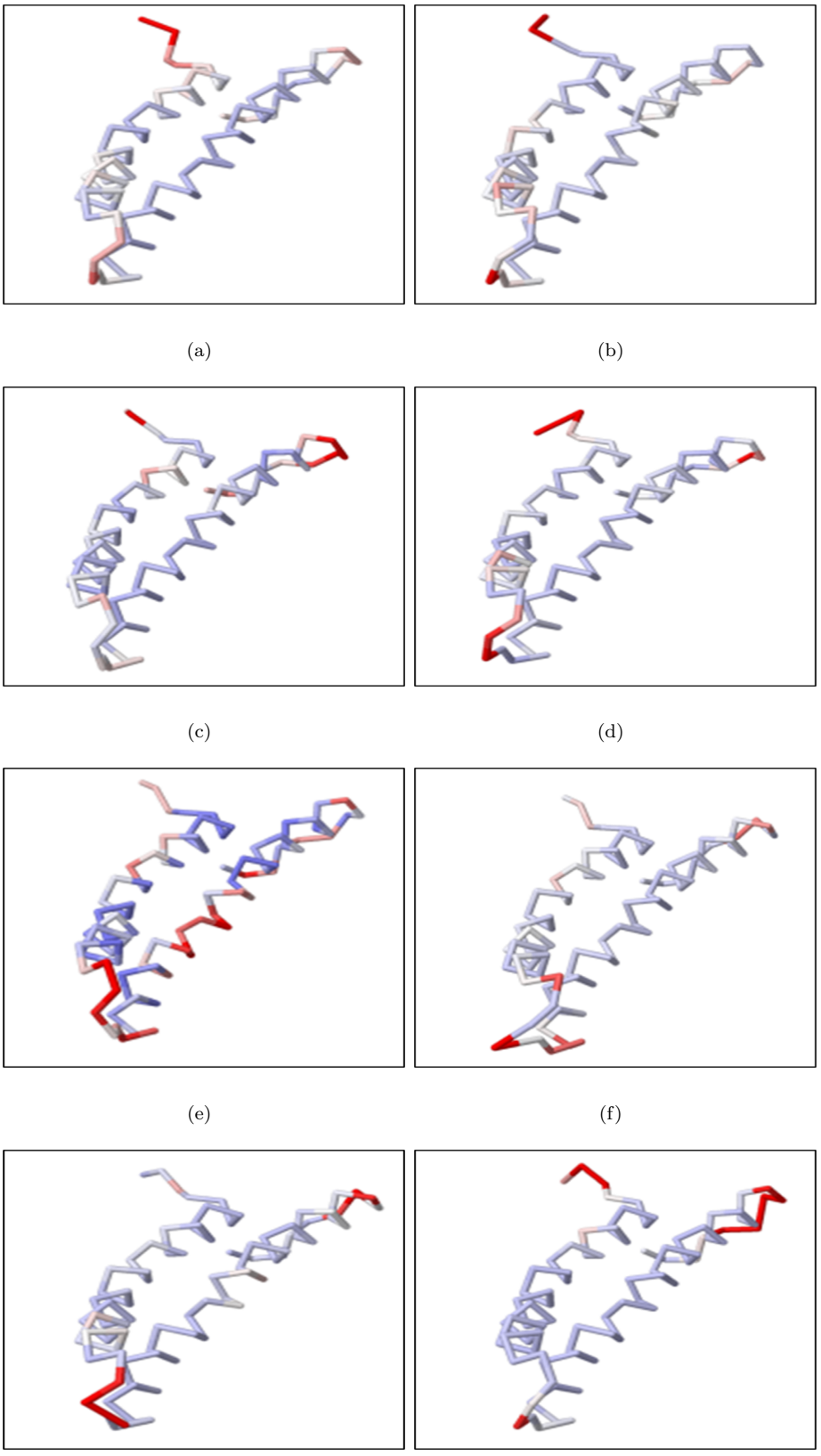
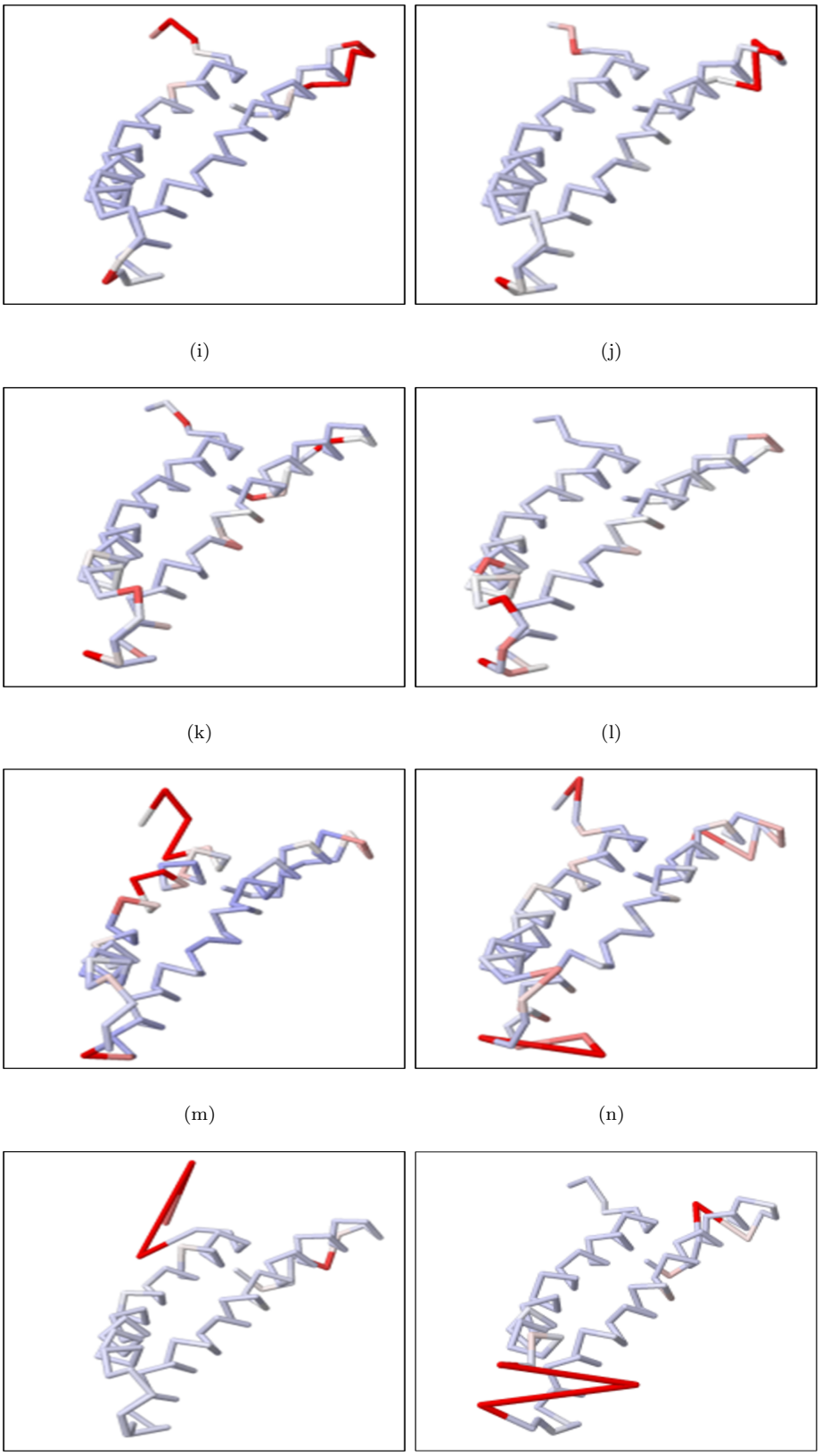


FIGURE 4.10: (A) ENMs of E5. Red color is showing the amino acid residues having highest degrees of fluctuation or elasticity in (a) Mode 1 (b) Mode 2 (c) Mode 3 (d) Mode 4



(a) (b) (c) (d) (e) (f) (g) (h)

(B) ENMs of E5. Red color is showing the amino acid residues having highest degrees of fluctuation or elasticity in (a) Mode 5 (b) Mode 6 (c) Mode 7 (d) Mode 8 (e) Mode 9 (f) Mode 10 (g) Mode 11 (h) Mode 12



(i) (j)

(k) (l)

(m) (n)

(o) (p)

(C) ENMs of E5. Red color is showing the amino acid residues having highest degrees of fluctuation or elasticity in (a) Mode 5 (b) Mode 6 (c) Mode 7 (d) Mode 8 (e) Mode 9 (f) Mode 10 (g) Mode 11 (h) Mode 12

4.3 Analysis of ENMs of E6

This study was designed to analyze the interaction of E6 with human proteins at residual level and to find hot spot residues of E6 involved in its interaction with hDlg, E6AP, TNF R1 and Paxillin, with the help of tool of Eukaryotic Linear Motifs and Elastic Network Models of E6. Two domains on human proteins were also predicted involved in stabilizing the interaction of E6 with E6AP, TNF R1 and Paxillin proteins.

The motifs reported to interact with PDZ domain in humans include LIG_PDZ_Class_1, LIG_PDZ_Class_2 and LIG_PDZ_Class_3. E6 contains sequence of LIG_PDZ_Class_1 (RRETQL) at position 153-158.

4.3.0.1 Motifs on E6

Annotation of oncoprotein E6 on eukaryotic linear motifs in humans gave us 25 different regions comprised of motifs found in eukaryotes (Table 4.18). As motifs are very important and play a major role in developing an interaction between proteins. For stabilizing these interactions the motifs develop a link with specific sequences in other proteins called domains. The motifs were selected on the basis of probability of p ¡ 0.1 [48].

Table 4.18: Motifs in E6 with counter domains

S. No.	ELM	Counter Domains	Proteins reported in literature to interact
1.	LIG_PDZ_Class_1	PDZ	HDlg
2.	CLV_NRD_NRD_1	Peptidase _M16	-
3.	CLV_PCSK_FUR_1	Peptidase _S8	-
4.	CLV_PCSK_PC1ET2_1	Peptidase _S8	-
5.	CLV_PCSK_SKI1_1	Peptidase _S8	-
6.	DOC_CYCLIN_1	Cyclin_N	-
7.	DOC_PP1_RVXF_1	Metallophos	-
8.	DOC_PP2A_B56_1	Armadillo type fold	-

9.	DOC_USP7_UBL2_3	USP7_ICP0_bdg	-
10.	LIG_eIF4E_1	IF4E	-
11.	LIG_FHA_2	FHA	-
12.	LIG_LIR_Gen_1	Atg 8	-
13.	LIG_PDZ_Class_1	PDZ	HDlg
14.	LIG_SH2_STAT3	SH2	-
15.	LIG_SH2_STAT5	SH2	-
16.	LIG_SH3_3	SH3_1	-
17.	LIG_UBA3_1	UBA3	-
18.	MOD_CK1_1	Pkinase	-
19.	MOD_CK2_1	Pkinase	-
20.	MOD_GSK3_1	Pkinase	-
21.	MOD_NEK2_1	Pkinase	-
22.	MOD_NEK2_2	Pkinase	-
23.	MOD_PKA_2	Pkinase	-
24.	TRG_ENDOCYTIC_2	Adap_comp_sub	-
25.	TRG_ER_diArg_1	WD40	-

These motifs are enlisted in Table 4.18, along with the domains with which they interact. Motifs in flexible regions of E6 exhibited by ENMs are described in Table 4.19 The amino acid sequences along with their position in E6 of motifs interacting with P kinase are found at 6 different regions as shown in table 4.20, while the motif interacting with armadillo domain is DOC_PP2A_B56_1 having sequence YSKISE at position 77-82.

TABLE 4.19: Motifs present in flexible regions of E6 in 20 Modes:

S. No.	**Modes**	**Amino acid Residue Position**	**Human Motifs**
1.	Mode 1:	152-158	LIG_PDZ_Class_1
2.	Mode 2:	1-4 152-158	CLV_PCSK_PC1ET2_1 LIG_PDZ_Class_1 CLV_NRD_NRD_1

			CLV_PCSK_PC1ET2_1
3.	Mode 3:	1-4	LIG_PDZ_Class_1
		152-158	CLV_NRD_NRD_1
4.	Mode 4:	1-7	CLV_PCSK_PC1ET2_1
			CLV_PCSK_PC1ET2_1
		1-6	LIG_SH2_STAT5
5.	Mode 5:	11-14	MOD_GSK3_1
		82-88	MOD_CK1_1
		129-130	MOD_NEK2_2
6.	Mode 6:	1-5	CLV_PCSK_PC1ET2_1
			CLV_NRD_NRD_1
			CLV_PCSK_FUR_1
			LIG_FHA_2
7.	Mode 7:	149-155	MOD_CK1_1
			MOD_CK2_1
			MOD_GSK3_1
			TRG_ER_diArg_1
			CLV_PCSK_PC1ET2_1
8.	Mode 8:	1-5	LIG_PDZ_Class_1
		153-158	CLV_NRD_NRD_1
			CLV_PCSK_PC1ET2_1
9.	Mode 9:	1-5	LIG_PDZ_Class_1
		153-158	CLV_NRD_NRD_1
			CLV_PCSK_PC1ET2_1
10.	Mode 10:	1-5	LIG_PDZ_Class_1
		153-158	CLV_NRD_NRD_1
11.	Mode 11:	1-5	CLV_PCSK_PC1ET2_1
		1-5	CLV_PCSK_PC1ET2_1
		20-31	MOD_GSK3_1
12.	Mode 12:	98-102	LIG_PDZ_Class_1
		153-158	CLV_NRD_NRD_1
13.	Mode 13:	1-3	CLV_PCSK_PC1ET2_1

14.	Mode 14:	8-9	LIG_PDZ_Class_1
		153-158	CLV_NRD_NRD_1
15.	Mode 15:	153-158	LIG_PDZ_Class_1
			CLV_NRD_NRD_1
16.	Mode 16:	149-1586-822-27	MOD_CK1_1
		86-88	MOD_CK2_1
		148-152	MOD_GSK3_1
			LIG_SH2_STAT5
			CLV_PCSK_FUR_1
			MOD_NEK2_2
			LIG_FHA_2
			TRG_ER_diArg_1
17.	Mode 17:	149-158	LIG_PDZ_Class_1
			CLV_NRD_NRD_1
			CLV_PCSK_FUR_1
			MOD_CK1_1
			MOD_CK2_1
			MOD_GSK3_1
			LIG_FHA_2
			TRG_ER_diArg_1
18.	Mode 18:	149-158	LIG_PDZ_Class_1
			CLV_NRD_NRD_1
			CLV_PCSK_FUR_1
			MOD_CK1_1
			MOD_CK2_1
			MOD_GSK3_1
			LIG_FHA_2
			TRG_ER_diArg_1
			CLV_PCSK_PC1ET2_1
19.	Mode 20:	1-7	LIG_SH2_STAT5
		84-85	MOD_GSK3_1
		99-100	MOD_CK1_1
			MOD_NEK2_2

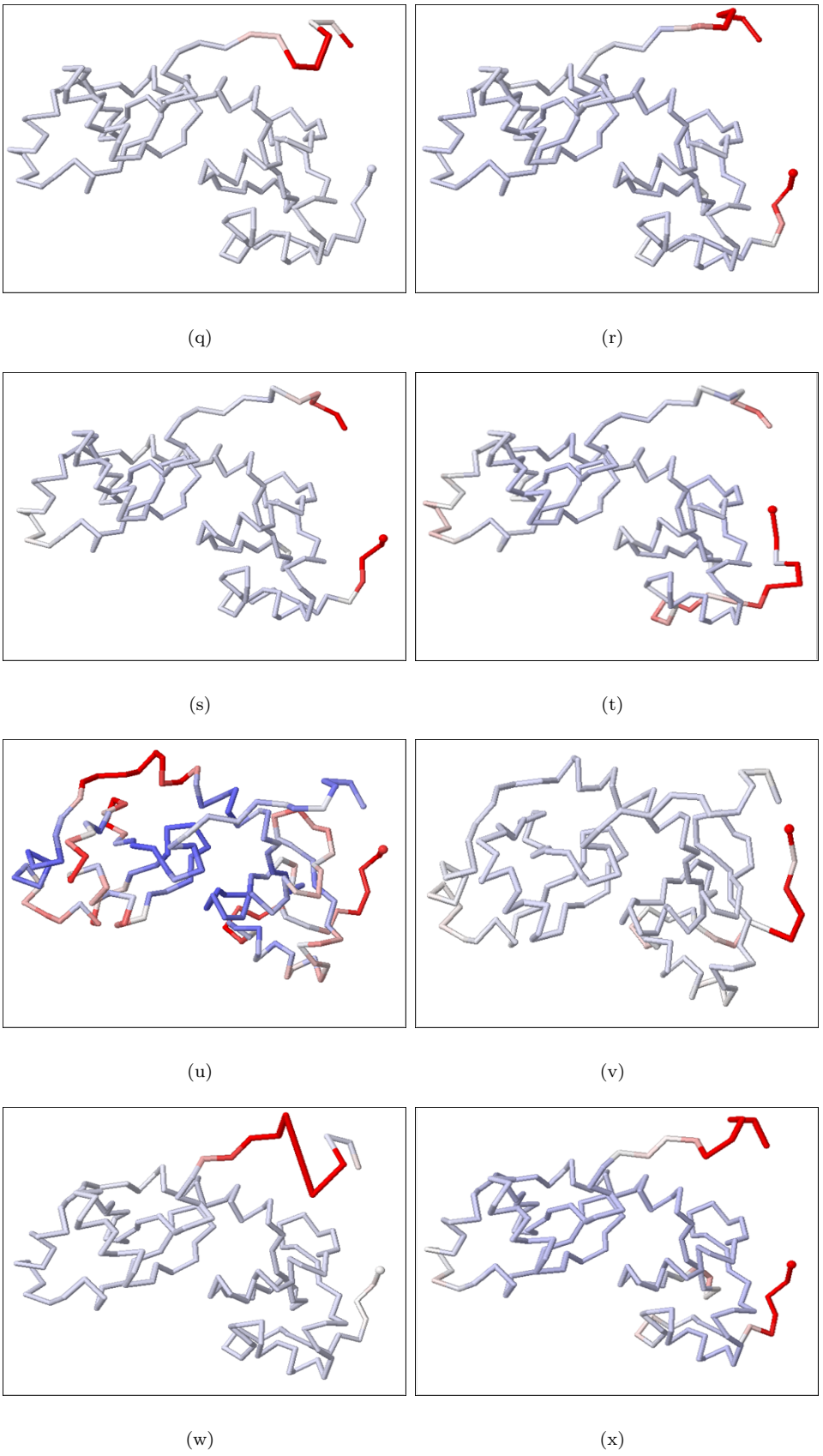
			MOD_CK1_1
			MOD_CK2_1
			MOD_GSK3_1
20.	Mode 19:	99-100	LIG_PDZ_Class_1
		149-158	CLV_NRD_NRD_1
			CLV_PCSK_FUR_1
			LIG_FHA_2
			TRG_ER_diArg_1

TABLE 4.20: Amino acid sequence of motifs on E6 interacting with P Kinase domain

S. No.	Motif	Residue Sequence	Position in E6
1.	MOD_CK1_1	SKISEYR	78-84
		SSRTRRE	149-155
2.	MOD_CK2_1	LYGTTLE	90-96
		QLCTELQT	21-28
3.	MOD_GSK3_1	YCYSLYGT	86-93
		RCMSCCRS	142-149
4.	MOD_NEK2_1	LQTTIH	26-31
5.	MOD_NEK2_2	YCYSLY	86-91
6.	MOD_PKA_2	GRWTGRC	137-143
		CRSSRTR	147-153

4.3.0.2 ENMs of E6

Elastic network models were generated in 20 different modes of protein E6. The elastic network models of E6 are shown in Figure 4.11. The protein is shown to exhibit possible movement in respective mode. Red colored regions of protein chain represent the residues which show the maximum degree of fluctuation in each mode. Elastic network models were generated in 20 different modes of protein E6. The protein is shown to exhibit possible movement in respective mode. Red colored regions of protein chain represent the residues.



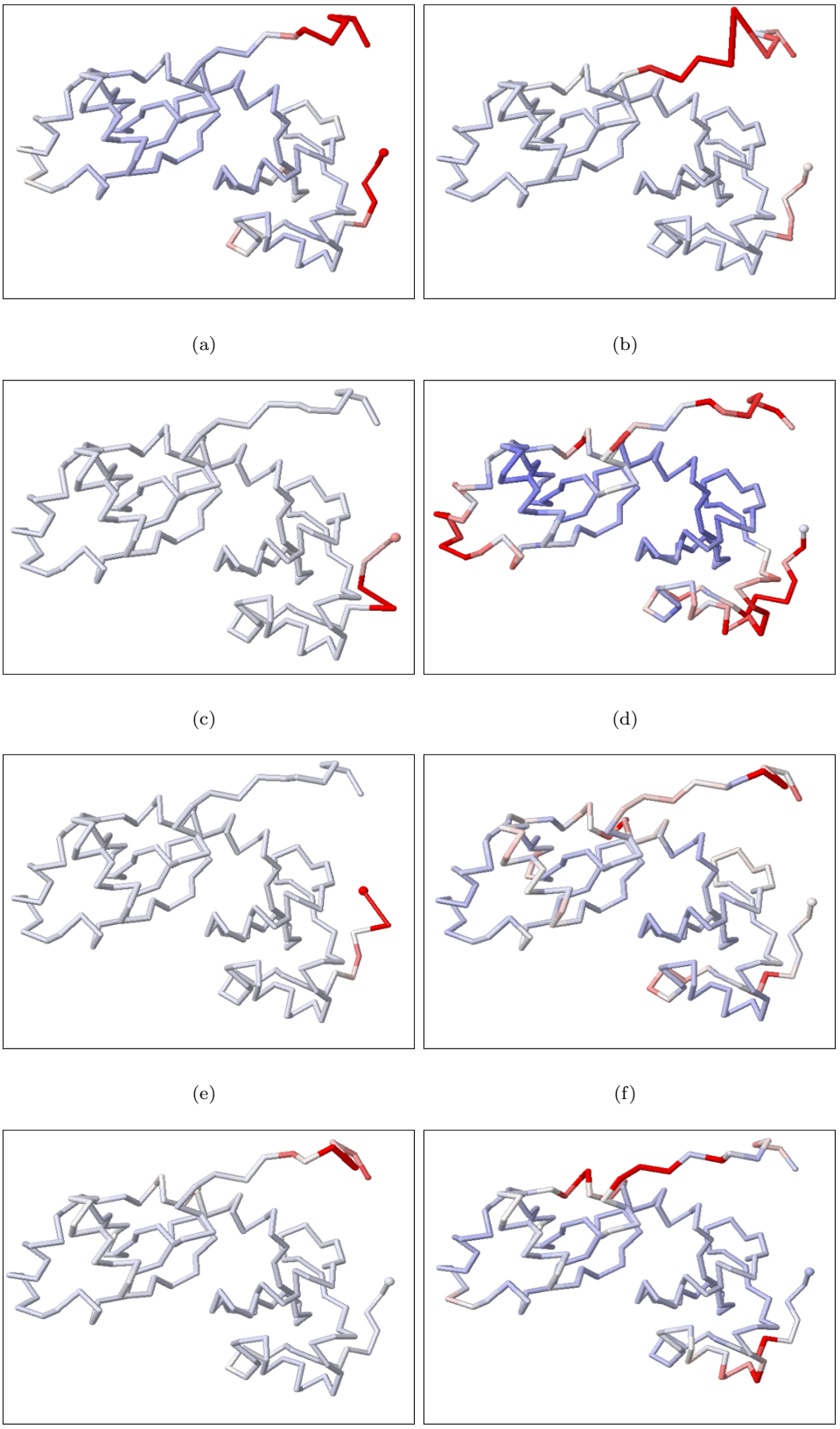
(q) (r)

(s) (t)

(u) (v)

(w) (x)

FIGURE 4.11: (A) ENMs of E6. Red color is showing the amino acid residues having highest degrees of fluctuation or elasticity in (a) Mode 1 (b) Mode 2 (c) Mode 3 (d) Mode 4 (e) Mode 5 (f) Mode 6 (g) Mode 7 (h) Mode 8



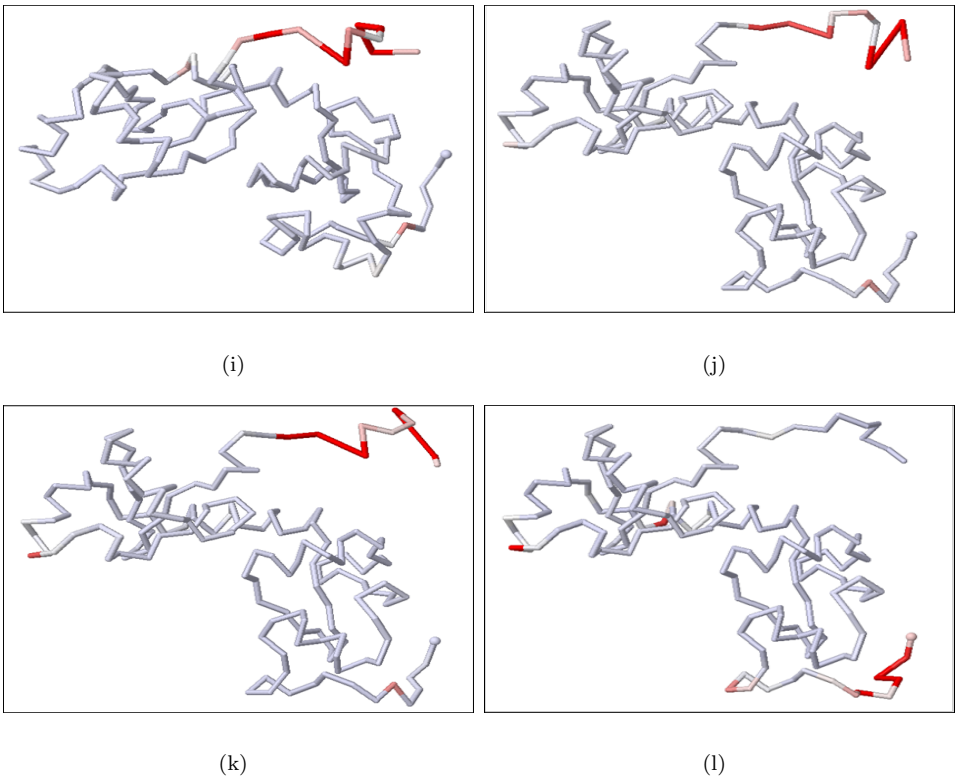
(a) (b)

(c) (d)

(e) (f)

(g) (h)

(B) ENMs of E6. Red color is showing the amino acid residues having highest degrees of fluctuation or elasticity in (a) Mode 9 (b) Mode 10 (c) Mode 11 (d) Mode 12 (e) Mode 13 (f) Mode 14 (g) Mode 15 (h) Mode 16



(i) (j)

(k) (l)

(C) ENMs of E6. Red color is showing the amino acid residues having highest degrees of fluctuation or elasticity in (i) Mode 17 (j) Mode 18 (k) Mode 19 (l) Mode 20

4.3.0.3 Observed Conformational Changes in E6 for Interaction with hDlg Protein

Mode 1, 2, 3, 8, 9, 10, 12, 14, 15, 17, 18 & 19 have shown the flexibility in the specific region which is possessed by the motifs involved in its interaction with hDlg. These modes were selected for the calculations of force constants and correlation analysis (Appendix A-33) and identification of hot spot residues involved in target protein binding. Modes 1, 2, 7, 8, 9, 10, 12, 14, 15, 17, 18 & 19 are the modes showing fluctuations in residues forming the RRETQL motif that is involved in interaction with hDlg protein, as shown in Figure 4.11, C termini of all HPV E6 proteins possess a class I PDZ binding motif (x-T/S-x-L/V) [237].

4.3.0.4 Distance Weights for Force Constants

B factors of individual amino acids residues are shown in the graph. These b-factors show the fluctuations of individual residues calculated on the criteria of

mean square fluctuations or experimental b-factors in Nuclear Magnetic Resonance structures, the calculated mean squared fluctuations, the normalized values, or on the basis of the fluctuations in every single mode individually (Appendix A-34, A-35).

The high mobility of the regions showing higher values of B-factors is requirement for not only the signaling or gating action of the binding or interacting region of the protein, but also for recognition of the specific protein. The participation of the most flexible regions of proteins emerging in the most cooperative, collective and the slowest modes for recognizing substrates is a spectacle that has been already discovered in many GNM studies for the equilibrium dynamics in many proteins as well as their complexes. The effective functioning and recognition of the functional sites of proteins depends majorly on the flexibility of specific amino acids residues in the collective modes [238].

The plot of B factors E6 are shown in Figure 4.12. Graphs of Modes 1, 2, 7, 8, 9, 10, 12, 14, 15, 17, 18 & 19 show highest values of B factors in regions having motifs XTXV/L (RRETQL), which validate its involvement in stabilizing the interaction of E6 with hDlg protein, as the regions having higher B factors are the regions with greater flexibility and are more suitable candidates to get involved in establishing interactions with other proteins. Amino acids showing higher values of B factor were found at positions 153-158, comprising of amino acids arginine, glutamate, threonine, glutamine, and leucine that is the region where XTXV/L is present.

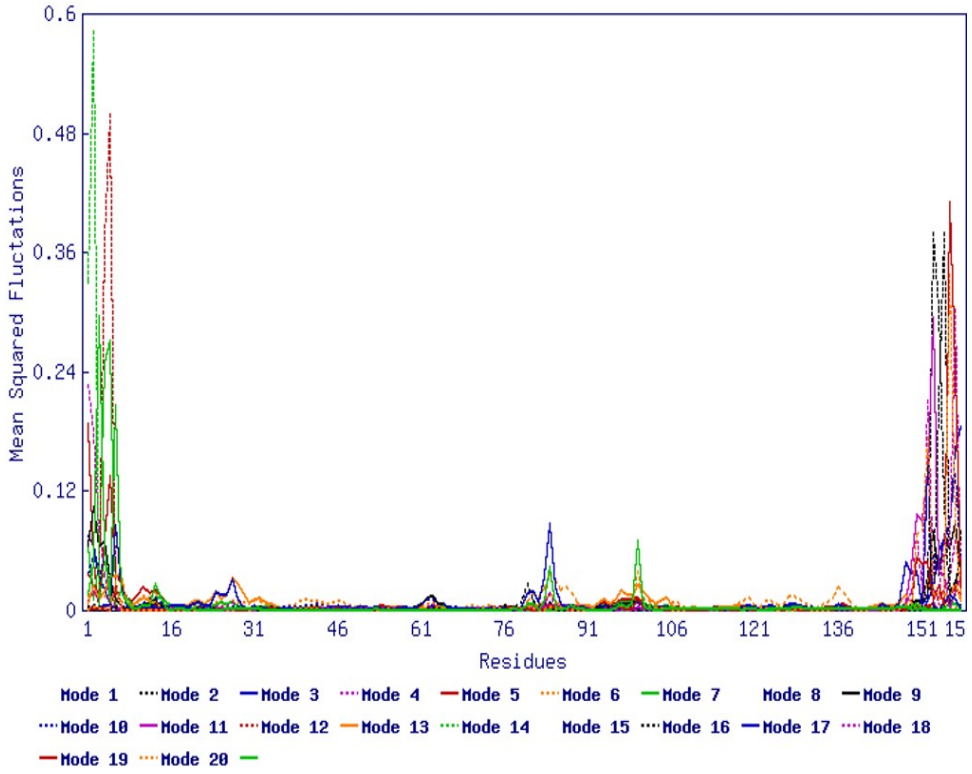


FIGURE 4.12: The plots of B factors of E6. Modes 1, 2, 7, 8, 9, 10, 12, 14, 15, 17, 18, & 19 show higher values of B factors in regions of motif XTXV/L/ RRETQL

4.3.0.5 Correlation Analysis of E6

In Appendix A-33 the correlation plots of E6 are depicted. In modes 1, 2, 7, 8, 9, 10, 12, 14, 15, 17, 18 & 19 high correlations were observed in the XTXV/L/ RRETQL motif region, giving evidence of their fluctuations and involvement in maintaining interactions with hDlg protein. In the color scheme of matrix the blue color represents large inter-residue fluctuations while red shows small fluctuations. Specific Blue regions show correlation of residues arginine, glutamate, threonine, glutamine, and leucine, with other amino acid residues.

4.3.0.6 Hot Spot Residues of E6 for Interaction with hDlg

On the basis of above findings of dynamics fluctuations of E6, for the region of hDlg binding amino acid residues, it was found that its hot-spot residues are R153, R154, E155, T156, Q157 and L158, whose agitations lead to the significant changes in fluctuations of hDlg-binding residues (Figure 4.13).

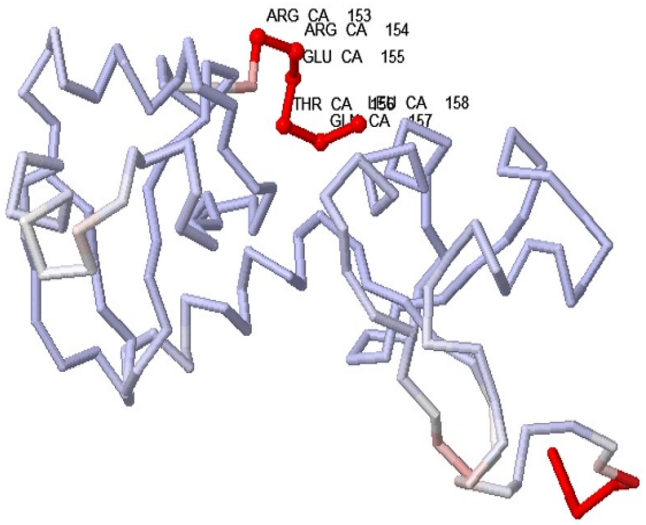


FIGURE 4.13: The hot spot residues of E6, which play a vital role in interaction of E with hDlg protein.

The important factor in E6-PDZ recognition of target and preferred interaction of various E6 proteins with different proteins having PDZ domain, is the specificity of the PDZ interaction, for example HPV-16 E6 to hDlg1 and MAGI-1 [296]. The capability of E6 for binding and promoting the degradation of p53 is not dependent on this one of the most conserved region of E6 [297], which suggests that this region may involve in regulation of some other processes important for cell transformation. Studies have shown that interaction of E6 with hDlg, that is the human homologue of the Drosophila tumor suppressor, called discs large protein (DlgA), is mediated by interaction between a conserved motif on C-terminal of oncogenic proteins of E6 and the PDZ domain of hDlg. In fact, there is a remarkable similarity between this motif (XTXV/L) of E6 and the peptide domain XS/TXV at C-terminal, that has been reported in previous studies to have interaction with this hydrophobic groove of the PDZ domain containing protein regions. PDZ domains consist of 80 $\pm$ 90 amino acid regions found in various proteins having distinct origin and function as the specific protein recognition sites. On the inner surface of the plasma membrane, Cytoplasmic proteins are targeted by these domains for the formation of complexes, on the inne surface of the plasma membrane. They help in clustering and linking of signaling molecules at specific sites in multiprotein complexes [298].

As it is very well explored that [299, 300], the dynamics of the protein structures can be explained well by NMA of an ENM. The important motions of proteins are characterized by the lowest frequency normal modes which are universally the most dominant. For further investigation, each pair of residues was considered for the correlations of the motions. Correlation analysis is able to detect and capture the motif RREQTL motions in stabilizing the interaction with hDlg protein. We used the standard ANM approach, which analyzes co-ordinately moving blocks of amino acid residues, and gives information directly about their correlations, for each mode. Force constant depends on type of amino acid which is based on nature of its R group and its behavior in that specific environment including the neighboring amino acids. If a specific amino acid/s is/are changed the force constant will also be effected and changed. Transcriptional modifications occur at RNA level usually leading to some change in amino acid sequence of synthesizing protein. As force constants depend on type of amino acids, so with any change they will also be affected [238].

PDZ domains are composed of almost 90 amino acid [301] and there are three categories of PDZ domain-containing proteins: PDZ-only proteins, PDZ proteins having other binding domains; and Membrane associated guanylate kinases (MAGUKs) [302]. Examples of PDZ domain-containing proteins are tight junctions and adherins, which usually function as platforms which gather signaling molecular complexes and accumulate these to specific cell to cell contact regions [303]. Numerous PDZ-domain containing proteins are interacted by proteins of the two significant high risk types, HPV-16 and HPV-18 through their PDZ-binding motifs (PBMs) [296, 304].

Interestingly, several potential tumor suppressors contain PDZ domain and are thus potential targets of E6 [305]. Presence of a peptide sequence homologous to carboxylic terminal of HPV18 E6 in its crystal structure and its particular mechanism of binding with PDZ domain supports the importance of the exact PDZ binding motif (PBM) sequence (x-T/S-x-L/V) in substrate recognition [306, 307]. This describes an outlandish evolutionary conservation of this phenomenon among high risk virus types to target the same control pathway of cell polarity by identical mechanisms, but with the help of different proteins [308].

The amino acids residues (R153, R154, E155, T156, Q157 and L158) found as hot spot amino acid residues spanning an extensive network which not only interacts with hDlg protein, but also stabilizes the interaction to create the cascade required in response. These residues of RREQTL (Arginine, Glutamic acid, Glutamine, Threonine & Leucine) motif are highly conserved residues of E6, which are found to be dynamically involved in interaction with PDZ domain of hDlg protein. Narayan et al., 2009 has shown in the study that serine residues of hDlg protein at positions 158 and 442 play a major role in being targeted by E6. Any change in amino acids at these sites could possibly effect its interaction with E6. Leucine is also part of the motif involved in interaction of E6 with E6AP, TNF R1 and Paxillin. These key amino acid residues can be proposed to facilitate dynamical couplings as well as mutual modulations between the E6 and hDlg and other proteins. They can be potential targets for drug designers.

The motif LXXLL on E6AP which binds with E6 has been broadly characterized. A number of other proteins like paxillin, E6AP, tuberin and IRF3, also use this motif LXXLL for binding with E6. While hDlg is the protein which has been reported to have and interaction with E6 through PDZ domain [298].

Some other proteins like Bak, p53, p300/CBP, NFX1, hADA3, FADD, procaspase 8, and Gps2 are the proteins which have been reported to have interactions with several E6 proteins. These interactions does not show involvement of either LXXLL or PDZ or both domains [307]. They have the possibility of binding to E6 by some motifs which are undefined yet, or indirectly by binding to some proteins associated with E6 just like E6AP.

The LXXLL motifs are composed of leucine-rich amphipathic helices having X position with at least a negative charge, with presence of leucine as hydrophobic residues. ELQELLGE is the sequence found in E6AP interacting with E6. E6 protein of numerous HPVs is characterized by its interaction with some proteins through LXXLL motif [309]. E6 binds with TNF R1 through specific sequence containing two E/D-L-L-L/V-G motifs. The region containing a pair of this sequence (PRREATLELLGRVRDMDLLGCL), at amino acids 419-422 and 429-433 is called death domain.

As the binding of TNF R1 with E6 has been experimentally checked by using of mammalian two-hybrid system. And the protein of TNF R1 with this region removed showed significant reduction in binding with both HPV16 E6 and TRADD [310]. Thus, it is suggested by these results that the cytoplasmic tail at the C-terminal of TNF R1 is the binding site for both HPV 16 E6 and TRADD.
The E6AP, TNF R1 and Paxillin binding with E6 rely on a short sequence of amino acids, that is leucine rich, called LXXLL binding motif; when mutate and/or deleted, there occurs prevention of formation of the complex E6-E6AP, E6-TNF R1 and E6-Paxilin. Because of interaction of E6 with these several LXXLL binding motif containing proteins, we hypothesized the possibility of presence of Protein Kinase (P kinase) and Armadillo domain in E6AP, TNF R1 and Paxillin, for their interaction with E6.

All P kinase domains sequences have a common conserved LLG/D/Y/R sequence at intervals [311], as shown in the following amino acid sequence with the highlighted regions. This sequence is similar to the sequence found in conserved regions of reactive domains of E6AP, TNF R1 and Paxillin.

FILHKMLGKGSFGKVFLAEFKKTNQFFAIKALKKDVVLMDDDVECTMVEK
RVLSLAWEHPFLTHMFCTFQTKENLFFVMEYLNGGDLMYHIQSCHKFDLS
RATFYAAEIILGLQFLHSKGIVYRDLKLDN ILLD KDGHIKIADFGMCKEN
MLGDAKTNTFCGTPDYIAPE ILLG QKYNHSVDWWSFG VLLY EMLIGQSP
FHGQDEEELFHSIRMDNPFYPRWLEKEAK DLLV KLFVREPEKRLGVRGDI
RQHPLF

Thus the specific regions in E6AP, TNF R1 and Paxillin could be predicted as the Protein Kinase domains. Amino acid sequence of motifs interacting with P kinase in E6 is given in supporting material.
The transcription factor HBP1 is degraded by proteasomal mechanism of degradation, that is mediated by ubiquitination through CTLH E3 ubiquitin-protein ligase complex, which utilizes the Armadillo domains for accepting ubiquitin from UBE2H [312]. Armadillo domain, found on different proteins in humans, are also characterized by having conserved leucine rich regions, as shown in sequences below.

Arm 1: NKQKANLIVLGAV PRLLYLLQ QETSSTELKTECAVVLGSLAM

Arm 2: ENNVKSLLDCHIIP ALLQGLLS PDLKFIEACLRCLRTIFT

Arm 3: TPE ELLY TDATVIPHLM ALLS RSRYTQEYICQIFSHCCK

Arm 4: PDHQTILFNHGAVQNIA HLLT SLSYKVRMQALKCFSVLAF

Arm 9: QQLRTSFQDHAVWKPLMKVLQNAPDEILVVASSMLC NLLL

Arm 11: QKIKADILRSLSTEQLF RLLS DSDLNVLMKTLGLLR NLLS

This similarity also shows its possibility to be the domain responsible for interactivity of E6 with these proteins. Thus the Pkinase and Armadillo domains could be predicted as the conserved domains found in the conserved regions of protein E6AP, TNFR1 and Paxillin required for their interaction with E6.

4.3.1 Analysis of ENMs E7

E7 is a protein composed of 98 amino acids [313, 314]. Among its amino acids sequence there are 18 regions which possess sequences similar to the sequences of different linear motifs found in human proteins. As motifs are very important and play a major role in developing an interaction between proteins. For stabilizing these interactions, the motifs develop a link with specific sequences in other proteins called domains.

pRB is degraded by E7 and its relative family member proteins through the proteasome degradation mechanism pathway, rsulting in the release of the transcription factor E2F which controls various cellular pathways. Overexpression of p16 that is a cyclin-dependent kinase inhibitor and is involved in prevention of phosphorylation of pRb proteins, has been observed during this process of inactivation of pRb via E7. E7 can destabilize the immune system by weakening the synthesi of proteins of adhesion molecules.

4.3.1.1 Motifs on E7

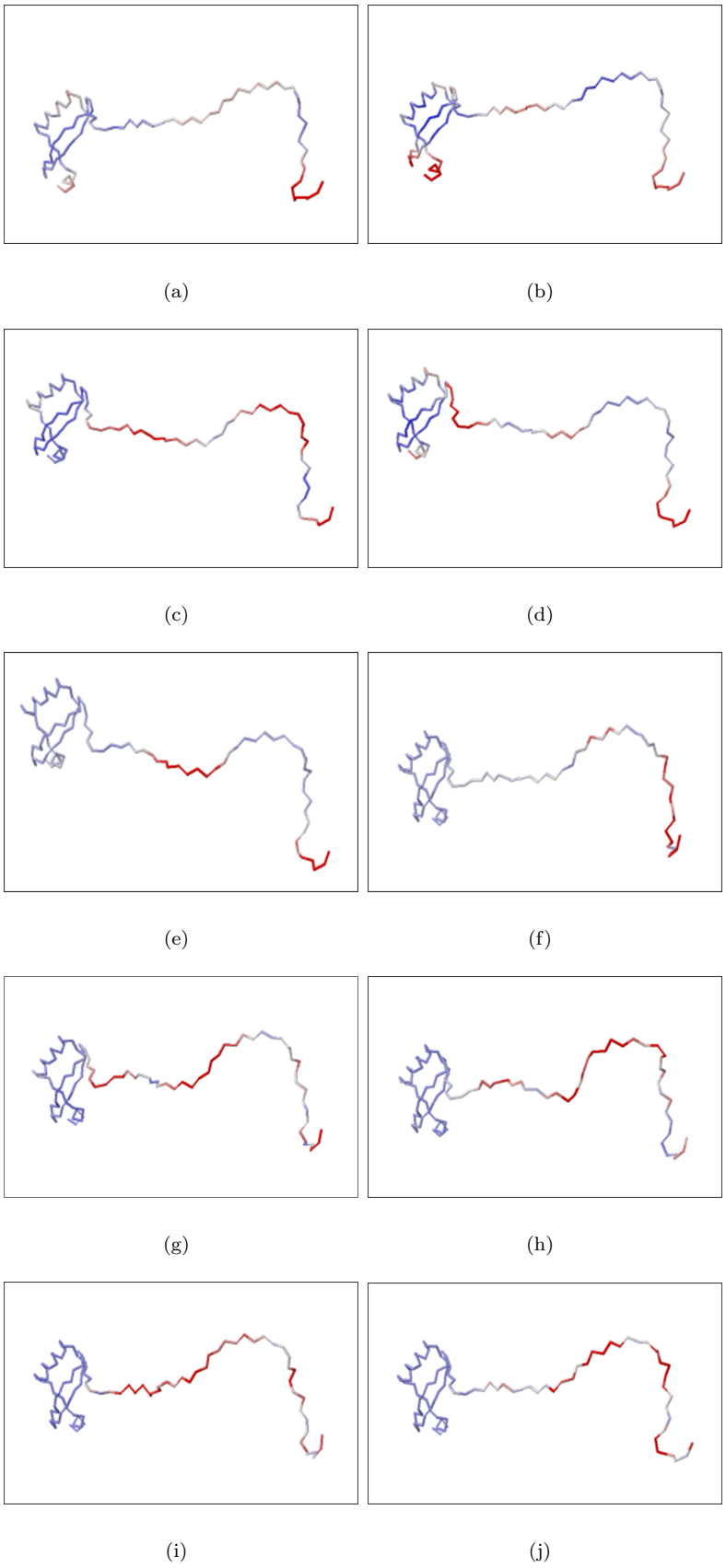
The motifs were selected on the basis of probability of p ¡ 0.1. These motifs are enlisted in Table 4.21, along with the domains with which they interact.

TABLE 4.21: Motifs in E7 with counter domains

S.No.	ELM	Counter Domains	Proteins from Literature
1.	CLV_C14_Caspase3-7	Peptidase_C14	-
2.	DOC_MAPK_MEF2A_6	P Kinase	-
3.	MOD_CK2_1	P Kinase	-
4.	MOD_NEK2_1	P Kinase	-
5.	MOD_NEK2_2	P Kinase	-
6.	MOD_ProDKin_1	P Kinase	-
7.	MOD_PLK	P Kinase	-
8.	DOC_USP7_MATH_1	MATH	-
9.	LIG_TRAF2_1	MATH	-
10.	DOC_WW_Pin1_4	WW	-
11.	LIG_FHA_1	FHA	-
12.	LIG_FHA_2	FHA	-
13.	LIG_NRBOX	Hormone_recep	-
14.	LIG_Rb_LxCxE_1	RB_B	Retinoblastoma
15.	LIG_SH2_STAT5	SH2	-
16.	MOD_N-GLC_1	STT3	-
17.	MOD_PIKK_1	PI3_PI4_kinase	-
18.	TRG_ENDOCYTIC_2	Adap_comp_sub	-

4.3.1.2 ENMs of E7

Elastic network models were generated in 20 different modes of protein E7. The elastic network models of E7 are shown in Figure 4.14 The protein is shown to exhibit possible movement in respective mode. Red colored regions of protein chain represent the residues which show the maximum degree of fluctuation in each mode. Mode 3, 5, 7, 8, 12 & 16 have shown the flexibility in the specific region which is possessed by LxCxE motif. These modes were selected for the calculations of force constants and correlation analysis and identification of hot spot residues involved in Rb binding. Motifs in flexible regions of E7 exhibited by ENMs are as follows:



(a) (b)

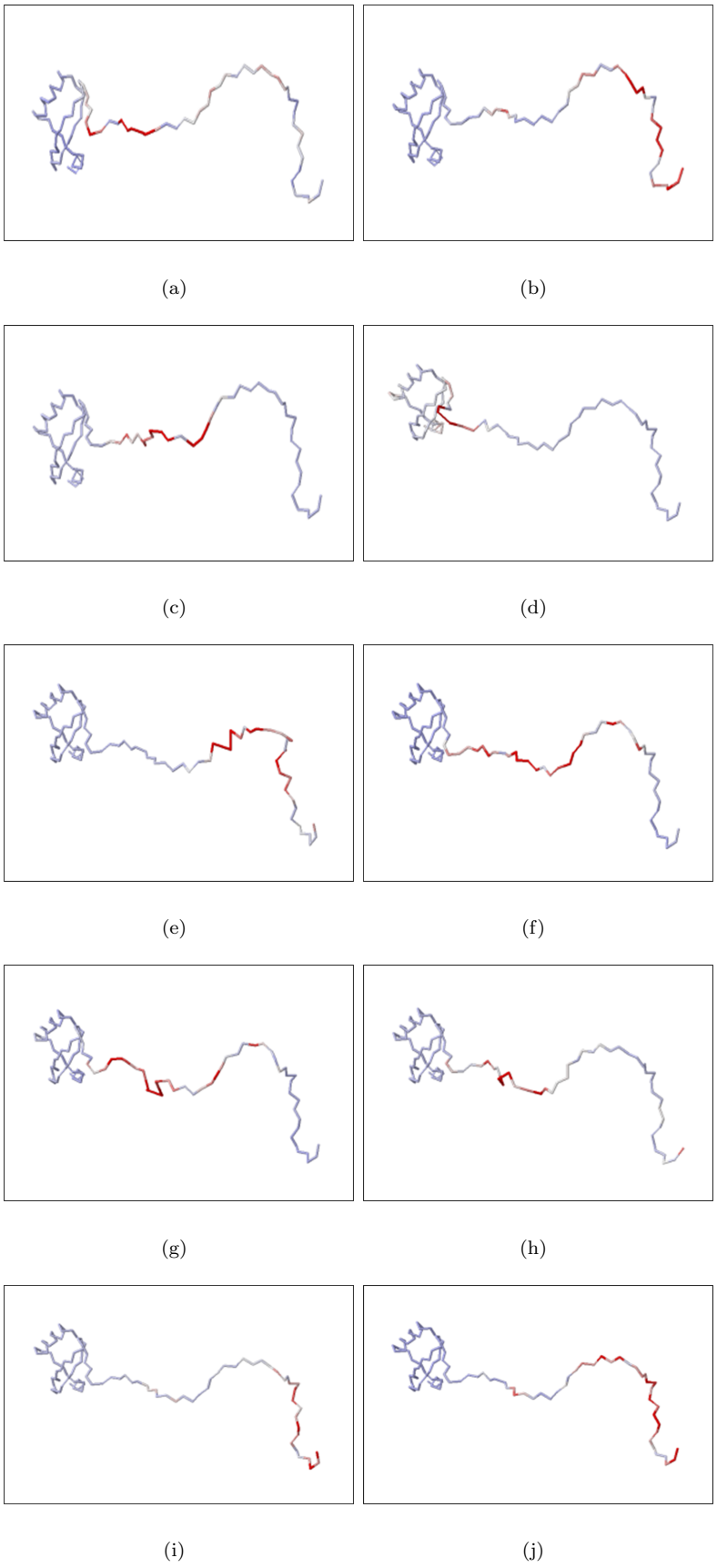
(c) (d)

(e) (f)

(g) (h)

(i) (j)

FIGURE 4.14: (A) ENMs of E5. Red color is showing the amino acid residues having highest degrees of fluctuation or elasticity (a) Mode 1 (b) Mode 2 (c) Mode 3 (d) Mode 4 (e) Mode 5 (f) Mode 6 (g) Mode 7 (h) Mode 8 (i) Mode 9 (j) Mode 10



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j)

(B) ENMs of E5. Red color is showing the amino acid residues having highest degrees of fluctuation or elasticity (a) Mode 1 (b) Mode 2 (c) Mode 3 (d) Mode 4 (e) Mode 5 (f) Mode 6 (g) Mode 7 (h) Mode 8 (i) Mode 9 (j) Mode 10

Mode 1: In mode 1 of E7 the amino acids which appear to show maximum flexibility and fluctuation are situated between position 1 and 10. This region contains the sequences similar to those of motifs MOD_ProDKin_1, DOC_WW_Pin_1_4, MOD_CK2_1 and LIG_FHA_2.

Mode 2: The regions showing maximum degree of fluctuation and flexibility in mode 2 lie between the positions 55-70 and 91-98. Motifs to be found in these regions are DOC_MAPK_MEF2A_6, LIG_FHA_1, MOD_NEK2_2, MOD_PLK, MOD_PIKK_1 AND DOC_USP7_MATH_1.

Mode 3: The fluctuating regions in mode 3 possess sequence/s for the motifs of CLV_C14_Caspase 3-7, DOC_WW_Pin1-4, LIG_FHA_1, LIG_FHA_2, LIG_Rb_ LxCxE_1, LIG_SH2_STAT5, LIG_TRAF2-1. MOD_CK2_1, MOD_N_GLC_1, MOD_ ProDkin_1 and TRG_ENDOCYTIC_2.

Mode 4: In mode 4, E7 shows 2 regions having higher degree of fluctuation, one lies in the region same as in mode 1 at positions 1-10 and other at positions 37-55 which possesses the sequences for the motifs of CLV_C14_Caspase 3-7 and TRG_ENDOCYTIC_2.

Mode 5: Mode 5 exhibits flexibility on positions 1-10 and 25-40. These regions show similarity to the sequence/s of motifs DOC_WW_Pin1_4, LIG_FHA_1, LIG_FHA_2, LIG_Rb_LxCxE_1, LIG_SH2_STAT5, LIG_TRAF2_1, MOD_CK2_1, MOD_N-GLC_1, MOD_ProDKin_1, and TRG_ENDOCYTIC_2.

Mode 6: In mode 6 of E7 the amino acids which appear to show maximum flexibility and fluctuation are situated between position 1 and 10, like that of mode 1. This region contains the sequences similar to those of motifs MOD_ProDKin_1, DOC_WW_Pin_1_4, MOD_CK2_1 and LIG_FHA_2.

Mode 7: The amino acids showing highest degree of fluctuation in mode 7 at position 19-55. The motifs lying in these regions include CLV_C14_Caspase3-7, LIG_FHA_1, LIG_Rb_LxCxE_1, LIG_SH2_STAT5, LIG_TRAF2_1, MOD_CK2_1, MOD_N-GLC_1, and TRG_ENDOCYTIC_2.

Mode 8: The mode 8 shows fluctuation in region of 19-28 positions of amino acids. These sequences match with the sequences of LIG_FHA_1, LIG_Rb_LxCxE_1, and LIG_SH2_STAT5.

Mode 9: The most flexible region of mode 9 lies in the region of amino acids situating at between positions 35 and 42. These sequences possess the similarity with motif CLV_C14_Caspase3-7.

Mode 10: The most flexible region exhibited by the amino acids in mode 10 is lying from positions 4-10. It possesses the sequences of motifs LIG_FHA_1, LIG_FHA_2 and MOD_CK2-1.

Mode 11: The amino acids showing maximum degree of movement in mode 11 are from position 32-40, with the possibility of presence of motifs CLV_C14_Caspase3-7 and LIG_TRAF2_1.

Mode 12: The mode 12 shows the amino acids from positions 14 -23 to be most flexible having the motif sequences of LIG_FHA_1 and LIG_Rb_LxCxE_1.

Mode 13: LIG_TRAF2_1 shows its presence in the mode 13 due to flexibility of the amino acids at positions 31-35.

Mode 14: No motif has been found matching the sequence of amino acids lying in the most flexible region of the mode 14 at positions 40-45. No motif has been found matching the sequence of amino acids lying in the most flexible region of the mode 14 at positions 40-45.

Mode 15: Amino acids showing the maximum degree of fluctuation in mode 15 are from position 23-28. Motifs having such sequences are LIG_SH2_STAT5 and TRG_ENDOCYTIC_2.

Mode 16: In mode 16 two regions are showing flexibility, one from position 22-26 and other 28-35. Motifs lying in these regions are LIG_Rb_LxCxE_1, LIG_SH2_STAT5,LIG_TRAF2_1, MOD_CK2-1, MOD_N-GLC_1, and TRG_ENDOCYTIC_2.

Mode 17: LIG_TRAF2_1 is the only motif showing its presence in the mode 17 due to maximum fluctuation exhibited by the amino acids of the sequence identical to it, at positions 32-37.

Mode 18: There are 2 flexible regions in the mode 18 one at position 30-32 and other at 33-37. LIG_TRAF2_1 and MOD_CK2-1 show similarity to the amino acids sequences present in these regions.

Mode 19: The mode 19 is occupied by three motifs LIG_FHA_2, MOD_CK2-1 and MOD_ProDKin_1 being similar in amino acid sequence at position 5-10. Amino acid 1 shows the highest value of fluctuation in this mode.

Mode 20: In mode 20 amino acids at positions22-26 show the maximum flexibility and their sequence is similar to those of motifs LIG_SH2_STAT5 and TRG_ENDOCYTIC_2.The present study was designed to analyze the force constants, correlation analysis of amino acid pairs and hot spot residues of E7 involved in its interactivity with Rb, with the help of ENM of E7. A correlation analysis was performed, that also helped in identification of dynamically important hot-spot residues, which are strongly involved in binding with Rb. Modes 3, 5, 7, 8, 12 & 16 are the modes showing fluctuations in residues forming the LxCxE motif that is involved in interaction with Rb protein.

A correlation analysis was performed, that also helped in identification of dynamically important hot-spot residues, which are strongly involved in binding with Rb. Modes 3, 5, 7, 8, 12 & 16 are the modes showing fluctuations in residues forming the LxCxE motif that is involved in interaction with Rb protein.

4.3.1.3 Distance Weights for Force Constants

Force constants γ are the parameters in elastic network modeling, which calculate the measure of strength of intramolecular potentials which stabilize the native fold of protein. The plot of B factors of E7 is shown in Figure 4.15. Graphs of B factors for modes 3, 5, 7, 8, 12 & 16 show highest values of B factors in regions having motif LxCxE, which validate its involvement in stabilizing the interaction of E7 with Rb protein, as the regions having higher B factors are the regions with greater flexibility and are more suitable candidates to get involved in establishing interactions with other proteins. Amino acids showing higher values of B factor (Figure 4.15) were found at positions 21-37, comprising of amino acids aspartate, leucine, tyrosine, cysteine, glutamate, glutamine, asparagine and serine, that is the region where LxCxE.

The regions having higher B factors are the regions with greater flexibility and are more suitable candidates to get involved in establishing interactions with other proteins. Amino acids showing higher values of B factor were found at positions

21-37, comprising of amino acids aspartate, leucine, tyrosine, cysteine, glutamate, glutamine, asparagine and serine, that is the region where LxCxE.

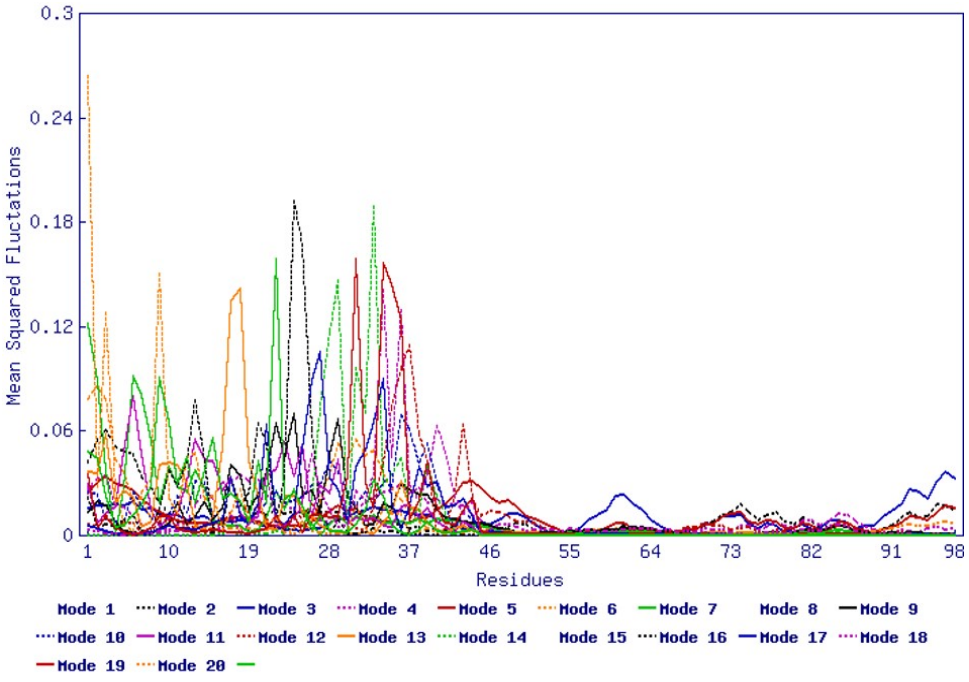


FIGURE 4.15: Graph B factors E7

4.3.1.4 Correlation Analysis of E7

Correlation analysis of E7 we again observed very high correlations of the LxCxE motif region, giving evidence of their fluctuations and involvement in maintaining interactions with Rb protein. As it is very well explored in previous studies [238, 315, 316], the dynamics of the protein structures can be explained well by NMA of an ENM. The important motions of proteins are characterized by the lowest frequency normal modes which are universally the most dominant. For further investigation, each pair of residues was considered for the correlations of the motions. Correlation analysis is able to detect and capture the motif LxCxE motions in stabilizing the interaction with Rb protein. We used the standard ANM approach, which analyzes co-ordinately moving blocks of amino acid residues, and gives information directly about their correlations, for each mode [238]. In Appendix A-36, A-37 & A-38, the correlation plots for modes 3, 5, 7, 8, 12 & 16 we again observed very high correlations of the LxCxE motif region, giving evidence of their fluctuations and involvement in maintaining interactions with Rb

protein. The retinoblastoma (Rb) protein family has been found to be interacted by multiple viral proteins through the LxCxE motif. The B domain of Rb protein possesses a highly conserved and shallow groove that is involved in its interaction with E7 mediated through the LxCxE motif. The central Cysteine is most conserved in all occurrences, however, the Leucine and Glutamate positions bear substitution of physiochemically alike amino acid residues allowing a less stringent definition of [LI]xCx[DE] [316, 317]. All these modes show maximum degree of fluctuation in the region of E7 containing sequence of motif LxCxE, as depicted by the sequence in flexible region of residues 21-37, i.e. DLYCYEQLNDSSEEEDE. The staggered but evenly spaced arrangement, and one amino acid residue apart, of the very conserved amino acid residues span one side of an stretched, beta-strand-like conformation and bind with the groove orthogonally, not through beta expansion like many other staggered motifs. The Cysteine and Leucine positions bind with a hydrophobic part of the groove having tight complementarity. One side of this binding groove forms alpha helix composed of backbone of two amide groups, that are attached with the Glutamic Acid through hydrogen bonds. This interaction is more stabilized by supplementary hydrogen bonds with the peptide backbone contributing to rigidity. Binding of Rb with LxCxE is inhibited by its Phosphorylation at Thr821 and Thr826.

4.3.1.5 Hot spot Residues of E7

The major mechanism behind the mediation of viral carcinogenesis is based on the capability of viral oncoproteins to interact and effect the activity or interactions of proteins of human cells. E7 has been considered as the major oncoprotein responsible for hijacking and changing interactivity pattern of host proteins. Recent studies have shown that E7 plays many other roles in malignant transformation of HPV infection, in addition to the inactivation of pRb. Hpv is responsible for more than 5% of the cancers found among humans. The survival of HPV in human cell is based on its genome's integrity and hypermrythylation of its DNA. These events result in overexpression of oncoproteins of HPV and causing cervical and other different types of cancers in humans.

On the basis of above findings of dynamics fluctuations of E7 (Appendix A-38, A-39), for the Rb binding amino acid residues, it was found in this study that its hot-spot residues are L22 (Leucine at 22 position), C24 (Cysteine at 24 position) , and E26 (Glutamic acid at 26 position), whose perturbations lead to a significant change in the fluctuations of Rb-binding residues. These residues span over an wide-spread network that not only interacts with Rb protein, but also stabilizes the interaction to create the cascade required in response. These residues of Lx-CxE motif are highly conserved residues of E7, which are found to be dynamically involved in interaction with Thr821 and Thr826 of Rb protein. These key amino acid residues are anticipated to facilitate dynamical couplings and shared modulations amongst the E7 and Rb protein. They can be potential targets for drug designers [318, 319].

Chapter 5

Conclusion and Recommendations

5.1 Conclusion

Insights into the mechanisms involved in development of cervical cancer from HPV pathogenesis are becoming the major focus of virologists and biochemists. Most of the work done was based on the exploration of pathways and types of proteins stabilizing these pathways. Very less work has been done on the protein-protein interactions at residual level. With the advancement of Bioinformatics, different techniques have been developed to explore the hierarchies of proteins in terms of their molecular dynamics. Elastic Network Modeling is one of those.

First Objective of this study was to predict the human proteins possibly interacted by HPV proteins by using resources of Protein Data Bank, RaptorX, Eukaryotic Linear Motifs (ELMs) and PROSITE. The human proteins were predicted, which could have the possibility of interactions with HPV proteins, on the basis of presence of Eukaryotic Linear Motifs of human proteins on HPV proteins and the counter domains of these motifs in human proteome. We predicted 12958 human proteins which could have interactions with one or more of the HPV proteins E1, E2, E4, E5, E6 and E7. Predicted proteins were statistically analyzed for

enrichment with Gene Ontology molecular functions and KEGG pathways. Predicted proteins in human having interactions with E6 were found to be largest in number having significant enrichment on KEGG Cancer Pathway, KEG Human Papillomavirus Infection Pathway, and KEGG Viral Carcinogenesis Pathway.

Second objective was to investigate and analyze HPV protein interactions with human proteins at residual level by using ENMs. The purpose was to visualize the regions of viral proteins that interact with CDs of human proteins. The objective to reveal the interactions of these domains with the human proteins was achieved by using ENMs. ENMs highlight these regions (residues) by showing fluctuations of those specific regions, that play critical role in maintaining interactions with human proteins. For example, E6 interacts with hDlg protein with the help of its ELM XTXV/L. This ELM was located at amino acids position 153-158, and that region showed flexibility in more than 50% of modes of ENMs of oncoprotein E6. E6 possesses 6 ELMs which can bind with Pkinase and 1 ELM can interact with Armadillo domains, in its flexible regions. From literature we found that E6 binds with E6AP, TNF R1 and Paxillin, through the involvement of specific conserved regions. These conserved regions had the sequence identical to Pkinase and Armadillo domain sequences. On the basis of sequence and involvement of these regions in interactions, we predicted presence of Pkinase and Armadillo domains in these human proteins.

Third Objective was to analyze the behavior of individual amino acids residues of HPV proteins with the help of Elastic Network Modeling. Amino acids of specific regions showed higher values of B-factors and exhibited fluctuations and flexibility in the regions involved in stabilizing PPIs, with human proteins. For example, in E6 the amino acids at position 153-158 involved in its interactions with hDlg (reported from literature) exhibited the higher values of B-factors.

Fourth objective of this study was prediction of motifs and domains on HPV proteins on the basis of data obtained from ELMs, and Elastic Network Modeling. Prediction of motifs, domains and hot spot residues of these proteins were other objectives of the study. We predicted these on the basis of already reported interactions and Elastic Network Models of these proteins along with insights into

the dynamics of those specific regions reported to be involved in the specific interactions. In this study motifs were predicted in particular regions of the HPV proteins, previously reported to have some involvement in specific interactions. Domains of human proteins were predicted on the basis of interactivity of the motifs with their counter domains. One third of the human proteins were found to be having Pkinase domain. The motifs DOC CYCLIN 1, MOD CDK SPK 2, and MOD CDK SPxxK 3 on N-terminal region of E1 were predicted, involved in maintaining its interaction with Cdk2 protein, that is already reported. The motifs having counter domain of Pkinase in N-terminal region of E1, were also found which helps in maintaining interaction of E1 with Histone H1 and chaperones proteins. The motifs ELMs, LIG FHA 1 and DOC MAPK gen 1 in E2 were discovered to be involved in phosphorylation interactions with human proteins. In this study already reported interaction of E6 with PDZ domain, were found with the presence of the motif LIG PDZ Class 1 on E6, this interaction was analyzed with Elastic Network Modeling. The domains Armadillo and Pkinase were predicted for interaction of E6 with E6AP, TNFR1 and Paxillin.

Our fifth and last objective was to predict the hot spot residues of HPV proteins. The amino acids proline, serine, threonine, glycine and arginine were predicted as hot spot residues of E2. This study helped to predict the hot spot residues of E6, arginine, glutamate, threonine, glutamine, and leucine. Previously reported interaction of E7 with retinoblastoma protein was analyzed and the hot spot residues of E7, Leucine, Cysteine and Glutamate, has been predicted.

The major purpose of this study was to explore the mechanism of virulence of HPV at molecular level to help in advancement of drug designing against the pathogenesis of HPV leading to cervical cancer. Protein-protein interactions are mediated and stabilized by binding of domain with motifs. Usually, the motif that is to be recognized by the domain for interaction is comprised of a simply structured region (no secondary or tertiary hierarchy) or a small segment. These motifs can either be in terminal regions or within a loop of the target protein, and they bind to the relatively flat recognition domains using a small groove. The cellular processes are mediated by these recognition mechanisms among the

interacting proteins. There are several diseases and syndromes related to the disruption of specific domain motif interactions. Domain motif interactions could therefore be an attractive group of new drug targets, because their fine modulation would allow for numerous desirable therapeutic effects.

In order to block these interactions, we need some inhibitory molecules, which could inhibit the binding pattern of interaction of domain with motif. These interactions are mediated or enhanced by many factors including recognition promiscuity of motifs and domains, structural properties of binding interface, affinity between the amino acids of these particular regions and the transient nature of these interactions. This had led to a general sense that protein–protein interactions might not be amenable to inhibition by small molecules A perhaps instructive counterpoint to this view is the case of enzymes protein kinases. A few years ago, these were the challenge for scientists to target them, in order to block their interactions with other proteins. This opinion was based on the high homology of the enzymatic site and the potent binding of the natural binder. The factors mentioned above, made it hard to explore molecules specifically enough to inhibit the kinase exclusively, involved in disease pathophysiology, with high enough affinity to compete against the ATP. Currently, there are numerous kinase inhibitors on the market. Similarly, researchers have made considerable progress over recent years in finding drug molecules that disrupt protein-protein interfaces. Typical protein-protein interaction interfaces tend to be large, flat and mainly hydrophobic, where punctual electrostatic interactions are key for the binding. Only hotspots in these interfaces are critical to the binding and recognition. They are major determinants of specificity and affinity, but at the same time allow flexibility to fit particular modifications. In general, these geometric and physiochemical properties are incompatible with the classic small molecules that satisfy Lipinski's rule of five, with good pharmacokinetics properties. In order to target these large and complex interfaces with enough specificity, we need to design larger compounds.

Classical techniques for drug development work with chemically manufactured small and active molecules. These molecules are good candidates for inhibiting interactions being having a wide range of properties desirable for drug discovery.

They are relatively easier to manipulate and synthesize, and have a cellularly good uptake. These molecules have greater and better ability to bind with deep grooves than the domain motif interactions interfaces. An evolution in old techniques is required to meet the new challenges of targeting domain motif interface with these small molecules. Efforts have been made for this by increasing the drug molecules' complexity, as well as the lead was more refined according to the complete description of structure of the naturally binding motif. While keeping the size of the compound reasonable, drug affinity and specificity could be maximized by the accurate identification of the hot spots and motif, majorly involved in the PPI.

Some examples of such discoveries include a specific inhibitor for the interaction between MDM2 and p53. Both of these proteins have been found important in carcinogenicity by HPV in our study. The molecules pyrrolidone 1 and epibestatin have been found as stabilizers of the interaction between 14–3–3 and cRaf proteins from a library of 37,000 different compounds. 14-3-3 is also among our predicted proteins having interactions with HPV proteins. By having some basic and validated background about the targeting mechanism of these proteins interactions, we can validate our study in wet lab through targeting their specific interactions with HPV proteins.

Different techniques which are in experimental phase for targeting and blocking PPIs through motifs and domains include targeting PPIs with small molecules, High Throughput Screening, Biologics, Fragment-Based Methodologies, Peptido mimetics, Stapled Peptides, and Macrocycles.

Bibliography

[1] E. Domingo, J. Holland, and S. Morse, "The evolutionary biology of viruses." *Raven Press, New York*, pp. 161–184, 1994.

[2] B. N. Fields, D. M. Knipe, P. M. Howley, K. D. Everiss, and H.-J. Kung, *Fundamental virology*. Lippincott-Raven Philadelphia^ ePA PA, 1996.

[3] J. D. Schrag, B. V. Prasad, F. J. Rixon, and W. Chiu, "Three-dimensional structure of the hsv1 nucleocapsid," *Cell*, vol. 56, no. 4, pp. 651–660, 1989.

[4] H. Zur Hausen, "Viruses in human cancers," *Science*, vol. 254, no. 5035, pp. 1167–1173, 1991.

[5] D. M. Parkin, "The global health burden of infection-associated cancers in the year 2002," *International journal of cancer*, vol. 118, no. 12, pp. 3030–3044, 2006.

[6] R. P. Beasley, C.-C. Lin, L.-Y. Hwang, and C.-S. Chien, "Hepatocellular carcinoma and hepatitis b virus: a prospective study of 22 707 men in taiwan," *The Lancet*, vol. 318, no. 8256, pp. 1129–1133, 1981.

[7] M. K. White, J. S. Pagano, and K. Khalili, "Viruses and human cancers: a long road of discovery of molecular paradigms," *Clinical microbiology reviews*, vol. 27, no. 3, pp. 463–481, 2014.

[8] Y. Chang, E. Cesarman, M. S. Pessin, F. Lee, J. Culpepper, D. M. Knowles, and P. S. Moore, "Identification of herpesvirus-like dna sequences in aids-associated kaposi's sarcoma," *Science*, vol. 266, no. 5192, pp. 1865–1869, 1994.

[9] U. N. Mui, C. T. Haley, and S. K. Tyring, "Viral oncology: molecular biology and pathogenesis," *Journal of clinical medicine*, vol. 6, no. 12, p. 111, 2017.

[10] J. Parsonnet, *Microbes and malignancy: infection as a cause of human cancers*. Oxford University Press, USA, 1999.

[11] H. Zur Hausen, "Viruses in human cancers," *Science*, vol. 254, no. 5035, pp. 1167–1173, 1991.

[12] E.-M. de Villiers, "Taxonomic classification of papillomaviruses," *Papillomavirus Report*, vol. 12, no. 3, pp. 57–63, 2001.

[13] E.-M. De Villiers, C. Fauquet, T. R. Broker, H.-U. Bernard, and H. Zur Hausen, "Classification of papillomaviruses," *Virology*, vol. 324, no. 1, pp. 17–27, 2004.

[14] J. Ferlay, M. Ervik, F. Lam, M. Colombet, L. Mery, M. Piñeros, A. Znaor, I. Soerjomataram, and F. Bray, "Global cancer observatory: cancer today," *Lyon, France: International Agency for Research on Cancer*, 2018.

[15] M. Plummer, C. de Martel, J. Vignat, J. Ferlay, F. Bray, and S. Franceschi, "Global burden of cancers attributable to infections in 2012: a synthetic analysis," *The Lancet Global Health*, vol. 4, no. 9, pp. e609–e616, 2016.

[16] C. M. De Oliveira, J. H. T. G. Fregnani, J. P. Carvalho, A. Longatto-Filho, and J. E. Levi, "Human papillomavirus genotypes distribution in 175 invasive cervical cancer cases from brazil," *BMC cancer*, vol. 13, no. 1, p. 357, 2013.

[17] R. Murillo and C. Ordóñez-Reyes, "Human papillomavirus (hpv) vaccination: from clinical studies to immunization programs," *International Journal of Gynecologic Cancer*, vol. 29, no. 8, pp. 1317–1326, 2019.

[18] S. A. Batool, S. Sajjad, and H. Malik, "Cervical cancer in pakistan: A review," *J Pak Med Assoc*, vol. 67, no. 7, pp. 1074–77, 2017.

[19] L. Bruni, L. Barrionuevo-Rosas, G. Albero, B. Serrano, M. Mena, D. Gómez *et al.*, "Human papillomavirus and related diseases in pakistan.[online] 2017 [cited 2017 jan 17]."

[20] I. W. G. on the Evaluation of Cancer-Preventive Strategies, I. A. for Research on Cancer, and W. H. Organization, *Cervix cancer screening*. Diamond Pocket Books (P) Ltd., 2005, vol. 10.

[21] I. In, "Iarc monograph on the evaluation of carcinogenic risks to humans," *IARC Monograph on the Evaluation of Carcinogenic Risks to Humans. Vol. Lyon: WHO International Agency for Research on Cancer*, 2002.

[22] R. A. Husain and V. Ramakrishnan, "Global variation of human papillomavirus genotypes and selected genes involved in cervical malignancies," *Annals of global health*, vol. 81, no. 5, pp. 675–683, 2015.

[23] C. S. REZAIE, A. C. S. MOHAMMAD, and M. Kamalifard, "Knowledge, attitudes and practice about pap smear among women reffering to a public hospital," 2012.

[24] H. Li, X. Wu, and X. Cheng, "Advances in diagnosis and treatment of metastatic cervical cancer," *Journal of gynecologic oncology*, vol. 27, no. 4, 2016.

[25] A. Bansal, M. P. Singh, and B. Rai, "Human papillomavirus-associated cancers: A growing global problem," *International Journal of Applied and Basic Medical Research*, vol. 6, no. 2, p. 84, 2016.

[26] K. Ganti, P. Massimi, J. Manzo-Merino, V. Tomaić, D. Pim, M. P. Playford, M. Lizano, S. Roberts, C. Kranjec, J. Doorbar *et al.*, "Interaction of the human papillomavirus e6 oncoprotein with sorting nexin 27 modulates endocytic cargo transport pathways," *PLoS pathogens*, vol. 12, no. 9, 2016.

[27] E. A. Mesri, M. A. Feitelson, and K. Munger, "Human viral oncogenesis: a cancer hallmarks analysis," *Cell host & microbe*, vol. 15, no. 3, pp. 266–282, 2014.

[28] N. S. Yeo-Teh, Y. Ito, and S. Jha, "High-risk human papillomaviral oncogenes e6 and e7 target key cellular pathways to achieve oncogenesis," *International journal of molecular sciences*, vol. 19, no. 6, p. 1706, 2018.

[29] B. Alberts, "Dna replication and recombination," *Nature*, vol. 421, no. 6921, pp. 431–435, 2003.

[30] D. Voet, J. G. Voet, and C. W. Pratt, *Fundamentals of biochemistry: life at the molecular level*, 2013, no. 577.1 VOE.

[31] G. M. Cooper and D. Ganem, "The cell: a molecular approach," *Nature Medicine*, vol. 3, no. 9, pp. 1042–1042, 1997.

[32] J. Darnell, H. Lodish, and D. Baltimore, *Molecular cell biology.*, 1990, no. QH581. 2 D22 1990.

[33] Z. Jin, M. Kotera, and S. Goto, "Virus proteins similar to human proteins as possible disturbance on human pathways," *Systems and synthetic biology*, vol. 8, no. 4, pp. 283–295, 2014.

[34] V. Bierbaum and R. Lipowsky, "Chemomechanical coupling and motor cycles of myosin v," *Biophysical journal*, vol. 100, no. 7, pp. 1747–1755, 2011.

[35] M. Karplus and J. A. McCammon, "Molecular dynamics simulations of biomolecules," *Nature structural biology*, vol. 9, no. 9, pp. 646–652, 2002.

[36] P. Joshi, T. M. Greco, A. J. Guise, Y. Luo, F. Yu, A. I. Nesvizhskii, and I. M. Cristea, "The functional interactome landscape of the human histone deacetylase family," *Molecular systems biology*, vol. 9, no. 1, 2013.

[37] P. Singh, A. Panchaud, and D. R. Goodlett, "Chemical cross-linking and mass spectrometry as a low-resolution protein structure determination technique," *Analytical chemistry*, vol. 82, no. 7, pp. 2636–2642, 2010.

[38] O. Rozenblatt-Rosen, R. C. Deo, M. Padi, G. Adelmant, M. A. Calderwood, T. Rolland, M. Grace, A. Dricot, M. Askenazi, M. Tavares *et al.*, "Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins," *Nature*, vol. 487, no. 7408, pp. 491–495, 2012.

[39] X. Yu, X. Bian, A. Throop, L. Song, L. Del Moral, J. Park, C. Seiler, M. Fiacco, J. Steel, P. Hunter *et al.*, "Exploration of panviral proteome: high-throughput cloning and functional implications in virus-host interactions," *Theranostics*, vol. 4, no. 8, p. 808, 2014.

[40] V. Tozzini, "Coarse-grained models for proteins," *Current opinion in structural biology*, vol. 15, no. 2, pp. 144–150, 2005.

[41] I. Bahar, A. R. Atilgan, and B. Erman, "Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential," *Folding and Design*, vol. 2, no. 3, pp. 173–181, 1997.

[42] T. Haliloglu, I. Bahar, and B. Erman, "Gaussian dynamics of folded proteins," *Physical review letters*, vol. 79, no. 16, p. 3090, 1997.

[43] K. Hinsen, "Analysis of domain motions by approximate normal mode calculations," *Proteins: Structure, Function, and Bioinformatics*, vol. 33, no. 3, pp. 417–429, 1998.

[44] Q. Cui and I. Bahar, *Normal mode analysis: theory and applications to biological and chemical systems*. CRC press, 2005.

[45] M. M. Tirion, "Large amplitude elastic motions in proteins from a single-parameter, atomic analysis," *Physical review letters*, vol. 77, no. 9, p. 1905, 1996.

[46] U. Karaoz, T. Murali, S. Letovsky, Y. Zheng, C. Ding, C. R. Cantor, and S. Kasif, "Whole-genome annotation by using evidence integration in functional-linkage networks," *Proceedings of the National Academy of Sciences*, vol. 101, no. 9, pp. 2888–2893, 2004.

[47] R. Sharan, I. Ulitsky, and R. Shamir, "Network-based prediction of protein function," *Molecular systems biology*, vol. 3, no. 1, 2007.

[48] H. Dinkel and H. Sticht, "A computational strategy for the prediction of functional linear peptide motifs in proteins," *Bioinformatics*, vol. 23, no. 24, pp. 3297–3303, 2007.

[49] P. Evans, W. Dampier, L. Ungar, and A. Tozeren, "Prediction of hiv-1 virus-host protein interactions using virus and host sequence motifs," *BMC medical genomics*, vol. 2, no. 1, p. 27, 2009.

[50] V. Neduva and R. B. Russell, "Peptides mediating interaction networks: new leads at last," *Current opinion in biotechnology*, vol. 17, no. 5, pp. 465–471, 2006.

[51] N. J. Krogan, G. Cagney, H. Yu, G. Zhong, X. Guo, A. Ignatchenko, J. Li, S. Pu, N. Datta, A. P. Tikuisis *et al.*, "Global landscape of protein complexes in the yeast saccharomyces cerevisiae," *Nature*, vol. 440, no. 7084, pp. 637–643, 2006.

[52] W. Dampier and A. Tozeren, "Signaling perturbations induced by invading h. pylori proteins in the host epithelial cells: A mathematical modeling approach," *Journal of theoretical biology*, vol. 248, no. 1, pp. 130–144, 2007.

[53] A. Sodhi, S. Montaner, and J. S. Gutkind, "Viral hijacking of g-protein-coupled-receptor signalling networks," *Nature Reviews Molecular Cell Biology*, vol. 5, no. 12, pp. 998–1012, 2004.

[54] J.-N. Tournier and A. Quesnel-Hellmann, "Host–pathogen interactions: a biological rendez-vous of the infectious nonself and danger models?" *PLoS pathogens*, vol. 2, no. 5, 2006.

[55] M. Ćupić, I. Lazarević, and N. Kuljić-Kapulica, "Oncogenic viruses and their role in tumor formation," *Srpski arhiv za celokupno lekarstvo*, vol. 133, no. 7-8, pp. 384–387, 2005.

[56] L. Goldman, D. A. Ausiello *et al.*, *Cecil medicine*. Saunders Elsevier Philadelphia, 2008.

[57] O. Bagci and S. Kurtgöz, "Amplification of cellular oncogenes in solid tumors," *North American journal of medical sciences*, vol. 7, no. 8, p. 341, 2015.

[58] C. M. Croce, "Oncogenes and cancer," *New England journal of medicine*, vol. 358, no. 5, pp. 502–511, 2008.

[59] J. Yokota, "Tumor progression and metastasis," *Carcinogenesis*, vol. 21, no. 3, pp. 497–503, 2000.

[60] J. S. Butel, "Viral carcinogenesis: revelation of molecular mechanisms and etiology of human disease," *Carcinogenesis*, vol. 21, no. 3, pp. 405–426, 2000.

[61] K. Takatsuki, "Discovery of adult t-cell leukemia," *Retrovirology*, vol. 2, no. 1, p. 16, 2005.

[62] M. Schiffman, P. E. Castle, J. Jeronimo, A. C. Rodriguez, and S. Wacholder, "Human papillomavirus and cervical cancer," *The Lancet*, vol. 370, no. 9590, pp. 890–907, 2007.

[63] K. Münger, A. Baldwin, K. M. Edwards, H. Hayakawa, C. L. Nguyen, M. Owens, M. Grace, and K. Huh, "Mechanisms of human papillomavirus-induced oncogenesis," *Journal of virology*, vol. 78, no. 21, pp. 11 451–11 460, 2004.

[64] H. zur Hausen, "Immortalization of human cells and their malignant conversion by high risk human papillomavirus genotypes," in *Seminars in cancer biology*, vol. 9, no. 6. Elsevier, 1999, pp. 405–411.

[65] H. Zur Hausen, "Papillomavirus infections—a major cause of human cancers," *Biochimica et biophysica acta (BBA)-reviews on cancer*, vol. 1288, no. 2, pp. F55–F78, 1996.

[66] J. Nicholas, "Human herpesvirus 8-encoded proteins with potential roles in virus-associated neoplasia," *Front Biosci*, vol. 12, pp. 265–281, 2007.

[67] G. Cathomas, "Kaposi's sarcoma-associated herpesvirus (kshv)/human herpesvirus 8 (hhv-8) as a tumour virus," *HERPES-CAMBRIDGE-*, vol. 10, no. 3, pp. 72–77, 2003.

[68] J. S. Pagano, M. Blaser, M.-A. Buendia, B. Damania, K. Khalili, N. Raab-Traub, and B. Roizman, "Infectious agents and cancer: criteria for a causal relation," in *Seminars in cancer biology*, vol. 14, no. 6. Elsevier, 2004, pp. 453–471.

[69] Q. Tao, L. S. Young, C. Woodman, and P. G. Murray, "Epstein-barr virus (ebv) and its associated human cancers–genetics, epigenetics, pathobiology and novel therapeutics," *Front Biosci*, vol. 11, no. 2, pp. 2672–2713, 2006.

[70] E. Klein, L. Kis, and G. Klein, "Epstein–barr virus infection in humans: from harmless to life endangering virus–lymphocyte interactions," *Oncogene*, vol. 26, no. 9, pp. 1297–1305, 2007.

[71] D.-Y. Jin, "Molecular pathogenesis of hepatitis c virus-associated hepatocellular carcinoma," *Front Biosci*, vol. 12, no. 1, pp. 222–233, 2007.

[72] M. Levrero, "Viral hepatitis and liver cancer: the case of hepatitis c," *Oncogene*, vol. 25, no. 27, pp. 3834–3847, 2006.

[73] N. Park, I. Song, and Y. Chung, "Chronic hepatitis b in hepatocarcinogenesis," *Postgraduate medical journal*, vol. 82, no. 970, pp. 507–515, 2006.

[74] D. Kremsdorf, P. Soussan, P. Paterlini-Brechot, and C. Brechot, "Hepatitis b virus-related hepatocellular carcinoma: paradigms for viral-related human carcinogenesis," *Oncogene*, vol. 25, no. 27, pp. 3823–3833, 2006.

[75] A. Lee and C. Lee, "Oncogenesis and transforming viruses: the hepatitis b virus and hepatocellularcarcinoma–the etiopathogenic link," *Front Biosci*, vol. 12, no. 1, pp. 234–245, 2007.

[76] C. N. Cole, "Polyomaviridae: the viruses and their replication," *Fields virology*, pp. 2141–2174, 2001.

[77] M. J. Imperiale, "Oncogenic transformation by the human polyomaviruses," *Oncogene*, vol. 20, no. 54, pp. 7917–7923, 2001.

[78] N. Bannert and R. Kurth, "Retroelements and the human genome: new perspectives on an old relation," *Proceedings of the National Academy of Sciences*, vol. 101, no. suppl 2, pp. 14 572–14 579, 2004.

[79] R. Gifford and M. Tristem, "The evolution, distribution and diversity of endogenous retroviruses," *Virus genes*, vol. 26, no. 3, pp. 291–315, 2003.

[80] J. F. Holland and B. G. Pogo, "Mouse mammary tumor virus-like viral infection and human breast cancer," *Clinical cancer research*, vol. 10, no. 17, pp. 5647–5649, 2004.

[81] T. Nishizawa, H. Okamoto, K. Konishi, H. Yoshizawa, Y. Miyakawa, and M. Mayumi, "A novel dna virus (ttv) associated with elevated transaminase levels in posttransfusion hepatitis of unknown etiology," *Biochemical and biophysical research communications*, vol. 241, no. 1, pp. 92–97, 1997.

[82] G. Klein, "Perspectives in studies of human tumor viruses." *Frontiers in bioscience: a journal and virtual library*, vol. 7, pp. d268–74, 2002.

[83] R. A. Weiss, "The discovery of endogenous retroviruses," *Retrovirology*, vol. 3, no. 1, p. 67, 2006.

[84] M. Şevik, "Oncogenic viruses and mechanisms of oncogenesis," *Turkish Journal of Veterinary and Animal Sciences*, vol. 36, no. 4, pp. 323–329, 2012.

[85] M. Matsuoka and K.-T. Jeang, "Human t-cell leukaemia virus type 1 (htlv-1) infectivity and cellular transformation," *Nature Reviews Cancer*, vol. 7, no. 4, pp. 270–280, 2007.

[86] F. A. Murphy, E. P. J. Gibbs, M. C. Horzinek, and M. J. Studdert, *Veterinary virology*. Elsevier, 1999.

[87] D. J. Griffiths, "Endogenous retroviruses in the human genome sequence," *Genome biology*, vol. 2, no. 6, pp. reviews1017–1, 2001.

[88] T. Muster, A. Waltenberger, A. Grassauer, S. Hirschl, P. Caucig, I. Romirer, D. Födinger, H. Seppele, O. Schanab, C. Magin-Lachmann *et al.*, "An endogenous retrovirus derived from human melanoma cells," *Cancer research*, vol. 63, no. 24, pp. 8735–8741, 2003.

[89] B. Dong, S. Kim, S. Hong, J. D. Gupta, K. Malathi, E. A. Klein, D. Ganem, J. L. DeRisi, S. A. Chow, and R. H. Silverman, "An infectious retrovirus susceptible to an ifn antiviral pathway from human prostate tumors," *Proceedings of the National Academy of Sciences*, vol. 104, no. 5, pp. 1655–1660, 2007.

[90] O. Hohn, H. Krause, P. Barbarotto, L. Niederstadt, N. Beimforde, J. Denner, K. Miller, R. Kurth, and N. Bannert, "Lack of evidence for xenotropic murine leukemia virus-related virus (xmrv) in german prostate cancer patients," *Retrovirology*, vol. 6, no. 1, p. 92, 2009.

[91] Z.-M. Zheng, "Viral oncogenes, noncoding rnas, and rna splicing in human tumor viruses," *International journal of biological sciences*, vol. 6, no. 7, p. 730, 2010.

[92] B. Damania, "Dna tumor viruses and human cancer," *Trends in microbiology*, vol. 15, no. 1, pp. 38–44, 2007.

[93] H. Zur Hausen, "Papillomaviruses and cancer: from basic studies to clinical application," *Nature reviews cancer*, vol. 2, no. 5, pp. 342–350, 2002.

[94] A. Felsani, A. Mileo, and M. Paggi, "Retinoblastoma family proteins as key targets of the small dna virus oncoproteins," *Oncogene*, vol. 25, no. 38, pp. 5277–5285, 2006.

[95] K. K. Lum and I. M. Cristea, "Proteomic approaches to uncovering virus–host protein interactions during the progression of viral infection," *Expert review of proteomics*, vol. 13, no. 3, pp. 325–340, 2016.

[96] H. Judson, B. Lewin, G. Stent, and J. Watson, "Basic genetic mechanisms," *Molecular Biology of the Cell. 3rd edn., Garland Science, New York*, pp. 273–287, 1994.

[97] K.-L. Wallin, F. Wiklund, T. Ångström, F. Bergman, U. Stendahl, G. Wadell, G. Hallmans, and J. Dillner, "Type-specific persistence of human papillomavirus dna before the development of invasive cervical cancer," *New England Journal of Medicine*, vol. 341, no. 22, pp. 1633–1638, 1999.

[98] L. A. Koutsky, K. A. Ault, C. M. Wheeler, D. R. Brown, E. Barr, F. B. Alvarez, L. M. Chiacchierini, and K. U. Jansen, "A controlled trial of a human papillomavirus type 16 vaccine," *New England Journal of Medicine*, vol. 347, no. 21, pp. 1645–1651, 2002.

[99] C. B. Buck, P. M. Day, and B. L. Trus, "The papillomavirus major capsid protein l1," *Virology*, vol. 445, no. 1-2, pp. 169–174, 2013.

[100] P. M. Day, R. Gambhira, R. B. Roden, D. R. Lowy, and J. T. Schiller, "Mechanisms of human papillomavirus type 16 neutralization by l2 cross-neutralizing and l1 type-specific antibodies," *Journal of virology*, vol. 82, no. 9, pp. 4638–4646, 2008.

[101] J. Doorbar, N. Egawa, H. Griffin, C. Kranjec, and I. Murakami, "Human papillomavirus molecular biology and disease association," *Reviews in medical virology*, vol. 25, pp. 2–23, 2015.

[102] M. E. Harden and K. Munger, "Human papillomavirus molecular biology," *Mutation Research/Reviews in Mutation Research*, vol. 772, pp. 3–12, 2017.

[103] M. Bergvall, T. Melendy, and J. Archambault, "The e1 proteins," *Virology*, vol. 445, no. 1-2, pp. 35–56, 2013.

[104] R. L. Garcea and M. J. Imperiale, "Simian virus 40 infection of humans," *Journal of virology*, vol. 77, no. 9, pp. 5039–5045, 2003.

[105] N. Egawa, T. Nakahara, S.-i. Ohno, M. Narisawa-Saito, T. Yugawa, M. Fujita, K. Yamato, Y. Natori, and T. Kiyono, "The e1 protein of human papillomavirus type 16 is dispensable for maintenance replication of the viral genome," *Journal of virology*, vol. 86, no. 6, pp. 3276–3283, 2012.

[106] E. A. Abbate, J. M. Berger, and M. R. Botchan, "The x-ray structure of the papillomavirus helicase in complex with its molecular matchmaker e2," *Genes & development*, vol. 18, no. 16, pp. 1981–1996, 2004.

[107] A. A. McBride, "The papillomavirus e2 proteins," *Virology*, vol. 445, no. 1-2, pp. 57–79, 2013.

[108] T. Nakahara, W. L. Peh, J. Doorbar, D. Lee, and P. F. Lambert, "Human papillomavirus type 16 e1 ˆ e4 contributes to multiple facets of the papillomavirus life cycle," *Journal of virology*, vol. 79, no. 20, pp. 13 150–13 165, 2005.

[109] R. Wilson, F. Fehrmann, and L. A. Laimins, "Role of the e1 ^ e4 protein in the differentiation-dependent life cycle of human papillomavirus type 31," *Journal of virology*, vol. 79, no. 11, pp. 6732–6740, 2005.

[110] R. Wilson, G. B. Ryan, G. L. Knight, L. A. Laimins, and S. Roberts, "The full-length e1^ e4 protein of human papillomavirus type 18 modulates differentiation-dependent viral dna amplification and late gene expression," *Virology*, vol. 362, no. 2, pp. 453–460, 2007.

[111] W. L. Peh, J. L. Brandsma, N. D. Christensen, N. M. Cladel, X. Wu, and J. Doorbar, "The viral e4 protein is required for the completion of the cottontail rabbit papillomavirus productive cycle in vivo," *Journal of virology*, vol. 78, no. 4, pp. 2142–2151, 2004.

[112] G. L. Knight, J. R. Grainger, P. H. Gallimore, and S. Roberts, "Cooperation between different forms of the human papillomavirus type 1 e4 protein to block cell cycle progression and cellular dna synthesis," *Journal of virology*, vol. 78, no. 24, pp. 13 920–13 933, 2004.

[113] G. H. Ashrafi, M. R. Haghshenas, B. Marchetti, P. M. O'Brien, and M. S. Campo, "E5 protein of human papillomavirus type 16 selectively downregulates surface hla class i," *International journal of cancer*, vol. 113, no. 2, pp. 276–283, 2005.

[114] M. Campo, S. Graham, M. Cortese, G. Ashrafi, E. Araibi, E. Dornan, K. Miners, C. Nunes, and S. Man, "Hpv-16 e5 down-regulates expression of surface hla class i and reduces recognition by cd8 t cells," *Virology*, vol. 407, no. 1, pp. 137–142, 2010.

[115] G. H. Ashrafi, M. Haghshenas, B. Marchetti, and M. S. Campo, "E5 protein of human papillomavirus 16 downregulates hla class i and interacts with the heavy chain via its first hydrophobic domain," *International journal of cancer*, vol. 119, no. 9, pp. 2105–2112, 2006.

[116] M. S. Cortese, G. H. Ashrafi, and M. S. Campo, "All 4 di-leucine motifs in the first hydrophobic domain of the e5 oncoprotein of human papillomavirus

type 16 are essential for surface mhc class i downregulation activity and e5 endomembrane localization," *International journal of cancer*, vol. 126, no. 7, pp. 1675–1682, 2010.

[117] J. E. Johnson and J. A. Speir, "Quasi-equivalent viruses: a paradigm for protein assemblies," *Journal of molecular biology*, vol. 269, no. 5, pp. 665–675, 1997.

[118] F. Schapiro, J. Sparkowski, A. Adduci, F. Suprynowicz, R. Schlegel, and S. Grinstein, "Golgi alkalinization by the papillomavirus e5 oncoprotein," *The Journal of cell biology*, vol. 148, no. 2, pp. 305–316, 2000.

[119] D. DiMaio and L. M. Petti, "The e5 proteins," *Virology*, vol. 445, no. 1-2, pp. 99–114, 2013.

[120] S. M. G. Williams, G. L. Disbrow, R. Schlegel, D. Lee, D. W. Threadgill, and P. F. Lambert, "Requirement of epidermal growth factor receptor for hyperplasia induced by e5, a high-risk human papillomavirus oncogene," *Cancer research*, vol. 65, no. 15, pp. 6534–6542, 2005.

[121] D. Mattoon, K. Gupta, J. Doyon, P. J. Loll, and D. DiMaio, "Identification of the transmembrane dimer interface of the bovine papillomavirus e5 protein," *Oncogene*, vol. 20, no. 29, pp. 3824–3834, 2001.

[122] H. L. Howie, R. A. Katzenellenbogen, and D. A. Galloway, "Papillomavirus e6 proteins," *Virology*, vol. 384, no. 2, pp. 324–334, 2009.

[123] D. Patel, S.-M. Huang, L. A. Baglia, and D. J. McCance, "The e6 protein of human papillomavirus type 16 binds to and inhibits co-activation by cbp and p300," *The EMBO journal*, vol. 18, no. 18, pp. 5061–5072, 1999.

[124] H. Zimmermann, R. Degenkolbe, H.-U. Bernard, and M. J. O'Connor, "The human papillomavirus type 16 e6 oncoprotein can down-regulate p53 activity by targeting the transcriptional coactivator cbp/p300," *Journal of virology*, vol. 73, no. 8, pp. 6209–6219, 1999.

[125] R. A. Katzenellenbogen, E. M. Egelkrout, P. Vliet-Gregg, L. C. Gewin, P. R. Gafken, and D. A. Galloway, "Nfx1-123 and poly (a) binding proteins

synergistically augment activation of telomerase in human papillomavirus type 16 e6-expressing cells," *Journal of virology*, vol. 81, no. 8, pp. 3786–3796, 2007.

[126] V. Tomaić, "Functional roles of e6 and e7 oncoproteins in hpv-induced malignancies at diverse anatomical sites," *Cancers*, vol. 8, no. 10, p. 95, 2016.

[127] J. Gage, C. Meyers, and F. Wettstein, "The e7 proteins of the nononcogenic human papillomavirus type 6b (hpv-6b) and of the oncogenic hpv-16 differ in retinoblastoma protein binding and other properties." *Journal of virology*, vol. 64, no. 2, pp. 723–730, 1990.

[128] N. Dyson, "The regulation of e2f by prb-family proteins," *Genes & development*, vol. 12, no. 15, pp. 2245–2262, 1998.

[129] J. W. Harbour and D. C. Dean, "Chromatin remodeling and rb activity," *Current opinion in cell biology*, vol. 12, no. 6, pp. 685–689, 2000.

[130] S. Chellappan, V. B. Kraus, B. Kroger, K. Munger, P. M. Howley, W. Phelps, and J. Nevins, "Adenovirus e1a, simian virus 40 tumor antigen, and human papillomavirus e7 protein share the capacity to disrupt the interaction between transcription factor e2f and the retinoblastoma gene product." *Proceedings of the National Academy of Sciences*, vol. 89, no. 10, pp. 4549–4553, 1992.

[131] J. Doorbar, "The papillomavirus life cycle," *Journal of clinical virology*, vol. 32, pp. 7–15, 2005.

[132] R. L. Finnen, K. D. Erickson, X. S. Chen, and R. L. Garcea, "Interactions between papillomavirus l1 and l2 capsid proteins," *Journal of virology*, vol. 77, no. 8, pp. 4818–4826, 2003.

[133] J. G. Joyce, J.-S. Tung, C. T. Przysiecki, J. C. Cook, E. D. Lehman, J. A. Sands, K. U. Jansen, and P. M. Keller, "The l1 major capsid protein of human papillomavirus type 11 recombinant virus-like particles interacts with heparin and cell-surface glycosaminoglycans on human keratinocytes," *Journal of Biological Chemistry*, vol. 274, no. 9, pp. 5810–5822, 1999.

[134] T. Giroglou, L. Florin, F. Schäfer, R. E. Streeck, and M. Sapp, "Human papillomavirus infection requires cell surface heparan sulfate," *Journal of virology*, vol. 75, no. 3, pp. 1565–1570, 2001.

[135] A. L. Combita, A. Touzé, L. Bousarghin, P.-Y. Sizaret, N. Muñoz, and P. Coursaget, "Gene transfer using human papillomavirus pseudovirions varies according to virus genotype and requires cell surface heparan sulfate," *FEMS microbiology letters*, vol. 204, no. 1, pp. 183–188, 2001.

[136] P. Drobni, N. Mistry, N. McMillan, and M. Evander, "Carboxy-fluorescein diacetate, succinimidyl ester labeled papillomavirus virus-like particles fluoresce after internalization and interact with heparan sulfate for binding and entry," *Virology*, vol. 310, no. 1, pp. 163–172, 2003.

[137] M. Sapp and P. M. Day, "Structure, attachment and entry of polyoma-and papillomaviruses," *Virology*, vol. 384, no. 2, pp. 400–409, 2009.

[138] S. Shafti-Keramat, A. Handisurya, E. Kriehuber, G. Meneguzzi, K. Slupetzky, and R. Kirnbauer, "Different heparan sulfate proteoglycans serve as cellular receptors for human papillomaviruses," *Journal of virology*, vol. 77, no. 24, pp. 13 125–13 135, 2003.

[139] L. T. Chow, T. R. Broker, and B. M. Steinberg, "The natural history of human papillomavirus infections of the mucosal epithelia," *Apmis*, vol. 118, no. 6-7, pp. 422–449, 2010.

[140] M. Conway and C. Meyers, "Replication and assembly of human papillomaviruses," *Journal of dental research*, vol. 88, no. 4, pp. 307–317, 2009.

[141] G. A. Maglennon, P. McIntosh, and J. Doorbar, "Persistence of viral dna in the epithelial basal layer suggests a model for papillomavirus latency following immune regression," *Virology*, vol. 414, no. 2, pp. 153–163, 2011.

[142] H.-C. Selinka, T. Giroglou, T. Nowak, N. D. Christensen, and M. Sapp, "Further evidence that papillomavirus capsids exist in two distinct conformations," *Journal of virology*, vol. 77, no. 24, pp. 12 961–12 967, 2003.

[143] J. L. Smith, S. K. Campos, and M. A. Ozbun, "Human papillomavirus type 31 uses a caveolin 1-and dynamin 2-mediated entry pathway for infection of human keratinocytes," *Journal of virology*, vol. 81, no. 18, pp. 9922–9931, 2007.

[144] N. Egawa, T. Nakahara, S.-i. Ohno, M. Narisawa-Saito, T. Yugawa, M. Fujita, K. Yamato, Y. Natori, and T. Kiyono, "The e1 protein of human papillomavirus type 16 is dispensable for maintenance replication of the viral genome," *Journal of virology*, vol. 86, no. 6, pp. 3276–3283, 2012.

[145] I. H. Frazer, "Prevention of cervical cancer through papillomavirus vaccination," *Nature Reviews Immunology*, vol. 4, no. 1, pp. 46–55, 2004.

[146] M. Stanley, "Pathology and epidemiology of hpv infection in females," *Gynecologic oncology*, vol. 117, no. 2, pp. S5–S10, 2010.

[147] P. González, A. Hildesheim, A. C. Rodríguez, M. Schiffman, C. Porras, S. Wacholder, A. G. Piñeres, L. A. Pinto, R. D. Burk, and R. Herrero, "Behavioral/lifestyle and immunologic factors associated with hpv infection among women older than 45 years," *Cancer Epidemiology and Prevention Biomarkers*, vol. 19, no. 12, pp. 3044–3054, 2010.

[148] A. F. Rositch, A. E. Burke, R. P. Viscidi, M. I. Silver, K. Chang, and P. E. Gravitt, "Contributions of recent and past sexual partnerships on incident human papillomavirus detection: acquisition and reactivation in older women," *Cancer research*, vol. 72, no. 23, pp. 6183–6190, 2012.

[149] T. Cheunim, J. Zhang, S. G. Milligan, M. G. McPhillips, and S. V. Graham, "The alternative splicing factor hnrnp a1 is up-regulated during virus-infected epithelial cell differentiation and binds the human papillomavirus type 16 late regulatory element," *Virus research*, vol. 131, no. 2, pp. 189–198, 2008.

[150] B. Collier, L. Goobar-Larsson, M. Sokolowski, and S. Schwartz, "Translational inhibition in vitro of human papillomavirus type 16 l2 mrna mediated

through interaction with heterogenous ribonucleoprotein k and poly (rc)-binding proteins 1 and 2," *Journal of Biological Chemistry*, vol. 273, no. 35, pp. 22 648–22 656, 1998.

[151] S. Mole, S. G. Milligan, and S. V. Graham, "Human papillomavirus type 16 e2 protein transcriptionally activates the promoter of a key cellular splicing factor, sf2/asf," *Journal of virology*, vol. 83, no. 1, pp. 357–367, 2009.

[152] S. G. Milligan, T. Veerapraditsin, B. Ahamet, S. Mole, and S. V. Graham, "Analysis of novel human papillomavirus type 16 late mrnas in differentiated w12 cervical epithelial cells," *Virology*, vol. 360, no. 1, pp. 172–181, 2007.

[153] C. Johansson, M. Somberg, X. Li, E. B. Winquist, J. Fay, F. Ryan, D. Pim, L. Banks, and S. Schwartz, "Hpv-16 e2 contributes to induction of hpv-16 late gene expression by inhibiting early polyadenylation," *The EMBO journal*, vol. 31, no. 14, pp. 3212–3227, 2012.

[154] S. D. Roth, M. Sapp, R. E. Streeck, and H.-C. Selinka, "Characterization of neutralizing epitopes within the major capsid protein of human papillomavirus type 33," *Virology journal*, vol. 3, no. 1, p. 83, 2006.

[155] P. M. Day, R. B. Roden, D. R. Lowy, and J. T. Schiller, "The papillomavirus minor capsid protein, l2, induces localization of the major capsid protein, l1, and the viral transcription/replication protein, e2, to pml oncogenic domains," *Journal of virology*, vol. 72, no. 1, pp. 142–150, 1998.

[156] C. S. Swindle and J. A. Engler, "Association of the human papillomavirus type 11 e1 protein with histone h1," *Journal of virology*, vol. 72, no. 3, pp. 1994–2001, 1998.

[157] H.-K. Wang, A. A. Duffy, T. R. Broker, and L. T. Chow, "Robust production and passaging of infectious hpv in squamous epithelium of primary human keratinocytes," *Genes & development*, vol. 23, no. 2, pp. 181–194, 2009.

[158] C. B. Buck, C. D. Thompson, Y.-Y. S. Pang, D. R. Lowy, and J. T. Schiller, "Maturation of papillomavirus capsids," *Journal of virology*, vol. 79, no. 5, pp. 2839–2846, 2005.

[159] H.-C. Selinka, L. Florin, H. D. Patel, K. Freitag, M. Schmidtke, V. A. Makarov, and M. Sapp, "Inhibition of transfer to secondary receptors by heparan sulfate-binding drug or antibody induces noninfectious uptake of human papillomavirus," *Journal of virology*, vol. 81, no. 20, pp. 10 970–10 980, 2007.

[160] M. Sapp and M. Bienkowska-Haba, "Viral entry mechanisms: human papillomavirus and a long journey from extracellular matrix to the nucleus," *The FEBS journal*, vol. 276, no. 24, pp. 7206–7216, 2009.

[161] P. M. Day and J. T. Schiller, "The role of furin in papillomavirus infection," *Future microbiology*, vol. 4, no. 10, pp. 1255–1262, 2009.

[162] J. Bordeaux, S. Forte, E. Harding, M. Darshan, K. Klucevsek, and J. Moroianu, "The l2 minor capsid protein of low-risk human papillomavirus type 11 interacts with host nuclear import receptors and viral dna," *Journal of virology*, vol. 80, no. 16, pp. 8259–8262, 2006.

[163] M. A. Schneider, G. A. Spoden, L. Florin, and C. Lambert, "Identification of the dynein light chains required for human papillomavirus infection," *Cellular microbiology*, vol. 13, no. 1, pp. 32–46, 2011.

[164] C. B. Buck, N. Cheng, C. D. Thompson, D. R. Lowy, A. C. Steven, J. T. Schiller, and B. L. Trus, "Arrangement of l2 within the papillomavirus capsid," *Journal of virology*, vol. 82, no. 11, pp. 5190–5197, 2008.

[165] J. Lowe, D. Panda, S. Rose, T. Jensen, W. A. Hughes, P. C. Angeletti *et al.*, "Evolutionary and structural analyses of alpha-papillomavirus capsid proteins yields novel insights into l2 structure and interaction with l1," *Virology journal*, vol. 5, no. 1, p. 150, 2008.

[166] M. B. Marušič, N. Mencin, M. Ličen, L. Banks, and H. Š. Grm, "Modification of human papillomavirus minor capsid protein l2 by sumoylation," *Journal of virology*, vol. 84, no. 21, pp. 11 585–11 589, 2010.

[167] X. S. Chen, R. L. Garcea, I. Goldberg, G. Casini, and S. C. Harrison, "Structure of small virus-like particles assembled from the l1 protein of human papillomavirus 16," *Molecular cell*, vol. 5, no. 3, pp. 557–567, 2000.

[168] A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman, "Global cancer statistics," *CA: a cancer journal for clinicians*, vol. 61, no. 2, pp. 69–90, 2011.

[169] C. Hughes, "Cervical cancer: prevention, diagnosis, treatment and nursing care," *Nursing Standard (through 2013)*, vol. 23, no. 27, p. 48, 2009.

[170] L. X. Clegg, M. E. Reichman, B. A. Miller, B. F. Hankey, G. K. Singh, Y. D. Lin, M. T. Goodman, C. F. Lynch, S. M. Schwartz, V. W. Chen *et al.*, "Impact of socioeconomic status on cancer incidence and stage at diagnosis: selected findings from the surveillance, epidemiology, and end results: National longitudinal mortality study," *Cancer causes & control*, vol. 20, no. 4, pp. 417–435, 2009.

[171] M. L. Brown, G. F. Riley, N. Schussler, and R. Etzioni, "Estimating health care costs related to cancer treatment from seer-medicare data," *Medical care*, pp. IV104–IV117, 2002.

[172] G. A. Boulet, C. A. Horvath, S. Berghmans, and J. Bogers, "Human papillomavirus in cervical cancer screening: important role as biomarker," *Cancer Epidemiology and Prevention Biomarkers*, vol. 17, no. 4, pp. 810–817, 2008.

[173] Y. Xu, Q. Wang, Y. Han, G. Song, and X. Xu, "Type-specific and cross-reactive antibodies induced by human papillomavirus 31 l1/l2 virus-like particles," *Journal of medical microbiology*, vol. 56, no. 7, pp. 907–913, 2007.

[174] L. S. Massad, M. Einstein, E. Myers, C. M. Wheeler, N. Wentzensen, and D. Solomon, "The impact of human papillomavirus vaccination on cervical cancer prevention efforts," *Gynecologic oncology*, vol. 114, no. 2, pp. 360–364, 2009.

[175] L. P. Keegan, A. Gallo, and M. A. O'Connell, "The many roles of an rna editor," *Nature reviews Genetics*, vol. 2, no. 11, pp. 869–878, 2001.

[176] T. Palmer, L. Wallace, K. G. Pollock, K. Cuschieri, C. Robertson, K. Kavanagh, and M. Cruickshank, "Prevalence of cervical disease at age 20 after immunisation with bivalent hpv vaccine at age 12-13 in scotland: retrospective population study," *bmj*, vol. 365, p. l1161, 2019.

[177] I. Ezkurdia, D. Juan, J. M. Rodriguez, A. Frankish, M. Diekhans, J. Harrow, J. Vazquez, A. Valencia, and M. L. Tress, "Multiple evidence strands suggest that there may be as few as 19 000 human protein-coding genes," *Human molecular genetics*, vol. 23, no. 22, pp. 5866–5878, 2014.

[178] R. Milo, "What is the total number of protein molecules per cell volume? a call to rethink some published values," *Bioessays*, vol. 35, no. 12, pp. 1050–1055, 2013.

[179] T. A. Down and T. J. Hubbard, "Computational detection and location of transcription start sites in mammalian genomic dna," *Genome research*, vol. 12, no. 3, pp. 458–461, 2002.

[180] D. L. Black, "Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology," *Cell*, vol. 103, no. 3, pp. 367–370, 2000.

[181] J. F. Cáceres and A. R. Kornblihtt, "Alternative splicing: multiple control mechanisms and involvement in human disease," *TRENDS in Genetics*, vol. 18, no. 4, pp. 186–193, 2002.

[182] A. C. Goldstrohm, A. L. Greenleaf, and M. A. Garcia-Blanco, "Co-transcriptional splicing of pre-messenger rnas: considerations for the mechanism of alternative splicing," *Gene*, vol. 277, no. 1-2, pp. 31–47, 2001.

[183] B. R. Graveley, "Alternative splicing: increasing diversity in the proteomic world," *TRENDS in Genetics*, vol. 17, no. 2, pp. 100–107, 2001.

[184] D. Gautheret, O. Poirot, F. Lopez, S. Audic, and J.-M. Claverie, "Alternate polyadenylation in human mrnas: a large-scale analysis by est clustering," *Genome research*, vol. 8, no. 5, pp. 524–530, 1998.

[185] R. E. Banks, M. J. Dunn, D. F. Hochstrasser, J.-C. Sanchez, W. Blackstock, D. J. Pappin, and P. J. Selby, "Proteomics: new perspectives, new biomedical opportunities," *The Lancet*, vol. 356, no. 9243, pp. 1749–1756, 2000.

[186] T. W. Nilsen and B. R. Graveley, "Expansion of the eukaryotic proteome by alternative splicing," *Nature*, vol. 463, no. 7280, pp. 457–463, 2010.

[187] D. Brett, H. Pospisil, J. Valcárcel, J. Reich, and P. Bork, "Alternative splicing and genome complexity," *Nature genetics*, vol. 30, no. 1, pp. 29–30, 2002.

[188] D. M. Creasy and J. S. Cottrell, "Unimod: Protein modifications for mass spectrometry," *Proteomics*, vol. 4, no. 6, pp. 1534–1536, 2004.

[189] R. F. Kalejta, "Tegument proteins of human cytomegalovirus," *Microbiol. Mol. Biol. Rev.*, vol. 72, no. 2, pp. 249–265, 2008.

[190] N. Philippe, M. Legendre, G. Doutre, Y. Couté, O. Poirot, M. Lescot, D. Arslan, V. Seltzer, L. Bertaux, C. Bruley *et al.*, "Pandoraviruses: amoeba viruses with genomes up to 2.5 mb reaching that of parasitic eukaryotes," *Science*, vol. 341, no. 6143, pp. 281–286, 2013.

[191] P. M. Ojala, B. Sodeik, M. W. Ebersold, U. Kutay, and A. Helenius, "Herpes simplex virus type 1 entry into host cells: reconstitution of capsid binding and uncoating at the nuclear pore complex in vitro," *Molecular and cellular biology*, vol. 20, no. 13, pp. 4922–4931, 2000.

[192] B. A. Diner, K. K. Lum, A. Javitt, and I. M. Cristea, "Interactions of the antiviral factor interferon gamma-inducible protein 16 (ifi16) mediate immune signaling and herpes simplex virus-1 immunosuppression," *Molecular & Cellular Proteomics*, vol. 14, no. 9, pp. 2341–2356, 2015.

[193] M. H. Orzalli, N. A. DeLuca, and D. M. Knipe, "Nuclear ifi16 induction of irf-3 signaling during herpesviral infection and degradation of ifi16 by the viral icp0 protein," *Proceedings of the National Academy of Sciences*, vol. 109, no. 44, pp. E3008–E3017, 2012.

[194] I. García-Dorival, W. Wu, S. Dowall, S. Armstrong, O. Touzelet, J. Wastling, J. N. Barr, D. Matthews, M. Carroll, R. Hewson *et al.*, "Elucidation of the ebola virus vp24 cellular interactome and disruption of virus biology through targeted inhibition of host-cell protein function," *Journal of proteome research*, vol. 13, no. 11, pp. 5120–5135, 2014.

[195] T. Watanabe, E. Kawakami, J. E. Shoemaker, T. J. Lopes, Y. Matsuoka, Y. Tomita, H. Kozuka-Hata, T. Gorai, T. Kuwahara, E. Takeda *et al.*, "Influenza virus-host interactome screen as a platform for antiviral drug development," *Cell host & microbe*, vol. 16, no. 6, pp. 795–805, 2014.

[196] W. Roos, R. Bruinsma, and G. Wuite, "Physical virology," *Nature physics*, vol. 6, no. 10, pp. 733–743, 2010.

[197] A. B. Pawar and I. Kretzschmar, "Fabrication, assembly, and application of patchy particles," *Macromolecular rapid communications*, vol. 31, no. 2, pp. 150–168, 2010.

[198] A. Arkhipov, P. L. Freddolino, and K. Schulten, "Stability and dynamics of virus capsids described by coarse-grained modeling," *Structure*, vol. 14, no. 12, pp. 1767–1777, 2006.

[199] M. F. Hagan and D. Chandler, "Dynamic pathways for viral capsid assembly," *Biophysical journal*, vol. 91, no. 1, pp. 42–54, 2006.

[200] H. D. Nguyen, V. S. Reddy, and C. L. Brooks, "Deciphering the kinetic mechanism of spontaneous self-assembly of icosahedral capsids," *Nano Letters*, vol. 7, no. 2, pp. 338–344, 2007.

[201] D. Rapaport, "Self-assembly of polyhedral shells: a molecular dynamics study," *Physical Review E*, vol. 70, no. 5, p. 051905, 2004.

[202] ——, "Role of reversibility in viral capsid growth: a paradigm for self-assembly," *Physical Review Letters*, vol. 101, no. 18, p. 186101, 2008.

[203] I. G. Johnston, A. A. Louis, and J. P. Doye, "Modelling the self-assembly of virus capsids," *Journal of Physics: Condensed Matter*, vol. 22, no. 10, p. 104101, 2010.

[204] D. Rapaport, "Modeling capsid self-assembly: design and analysis," *Physical biology*, vol. 7, no. 4, p. 045001, 2010.

[205] A. W. Wilber, J. P. Doye, and A. A. Louis, "Self-assembly of monodisperse clusters: Dependence on target geometry," *The Journal of chemical physics*, vol. 131, no. 17, p. 11B601, 2009.

[206] A. W. Wilber, J. P. Doye, A. A. Louis, and A. C. Lewis, "Monodisperse self-assembly in a model with protein-like interactions," *The Journal of chemical physics*, vol. 131, no. 17, p. 11B602, 2009.

[207] P. Moisant, H. Neeman, and A. Zlotnick, "Exploring the paths of (virus) assembly," *Biophysical journal*, vol. 99, no. 5, pp. 1350–1357, 2010.

[208] A. Zlotnick, J. M. Johnson, P. W. Wingfield, S. J. Stahl, and D. Endres, "A theoretical model successfully identifies features of hepatitis b virus capsid assembly," *Biochemistry*, vol. 38, no. 44, pp. 14 644–14 652, 1999.

[209] A. Zlotnick, P. Ceres, S. Singh, and J. M. Johnson, "A small molecule inhibits and misdirects assembly of hepatitis b virus capsids," *Journal of virology*, vol. 76, no. 10, pp. 4848–4854, 2002.

[210] D. Endres and A. Zlotnick, "Model-based analysis of assembly kinetics for virus capsids or other spherical polymers," *Biophysical journal*, vol. 83, no. 2, pp. 1217–1230, 2002.

[211] J. M. Johnson, J. Tang, Y. Nyame, D. Willits, M. J. Young, and A. Zlotnick, "Regulating self-assembly of spherical oligomers," *Nano letters*, vol. 5, no. 4, pp. 765–770, 2005.

[212] A. Zlotnick and S. Mukhopadhyay, "Virus assembly, allostery and antivirals," *Trends in microbiology*, vol. 19, no. 1, pp. 14–23, 2011.

[213] H. D. Nguyen and C. L. Brooks III, "Generalized structural polymorphism in self-assembled viral particles," *Nano letters*, vol. 8, no. 12, pp. 4574–4581, 2008.

[214] O. M. Elrad and M. F. Hagan, "Mechanisms of size control and polymorphism in viral capsid assembly," *Nano letters*, vol. 8, no. 11, pp. 3850–3857, 2008.

[215] H. D. Nguyen, V. S. Reddy, and C. L. Brooks Iii, "Invariant polymorphism in virus capsid assembly," *Journal of the American Chemical Society*, vol. 131, no. 7, pp. 2606–2614, 2009.

[216] L. Cardarelli, K. L. Maxwell, and A. R. Davidson, "Assembly mechanism is the key determinant of the dosage sensitivity of a phage structural protein," *Proceedings of the National Academy of Sciences*, vol. 108, no. 25, pp. 10 168–10 173, 2011.

[217] S. Flasinski, A. Dzianott, J. A. Speir, J. E. Johnson, and J. J. Bujarski, "Structure-based rationale for the rescue of systemic movement of brome mosaic virus by spontaneous second-site mutations in the coat protein gene." *Journal of virology*, vol. 71, no. 3, pp. 2500–2504, 1997.

[218] A. Zlotnick, R. Aldrich, J. M. Johnson, P. Ceres, and M. J. Young, "Mechanism of capsid assembly for an icosahedral plant virus," *Virology*, vol. 277, no. 2, pp. 450–456, 2000.

[219] D. Willits, X. Zhao, N. Olson, T. Baker, A. Zlotnick, J. Johnson, T. Douglas, and M. Young, "Effects of the cowpea chlorotic mottle bromovirus β-hexamer structure on virion assembly," *Virology*, vol. 306, no. 2, pp. 280–288, 2003.

[220] S. J. Hanslip, N. R. Zaccai, A. P. Middelberg, and R. J. Falconer, "Assembly of human papillomavirus type-16 virus-like particles: Multifactorial study of assembly and competing aggregation," *Biotechnology progress*, vol. 22, no. 2, pp. 554–560, 2006.

[221] A. Oppenheim, O. Ben-Nun-Shaul, S. Mukherjee, and M. Abd-El-Latif, "Sv40 assembly in vivo and in vitro," *Computational and Mathematical Methods in Medicine*, vol. 9, no. 3-4, pp. 265–276, 2008.

[222] S. N. Fejer, T. R. James, J. Hernandez-Rojas, and D. J. Wales, "Energy landscapes for shells assembled from pentagonal and hexagonal pyramids," *Physical Chemistry Chemical Physics*, vol. 11, no. 12, pp. 2098–2104, 2009.

[223] M. H. Kim and M. K. Kim, "Elastic network model for protein structural dynamics," *JSM Enzymol. Protein Sci*, vol. 1, p. 1001, 2014.

[224] A. R. Atilgan, S. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar, "Anisotropy of fluctuation dynamics of proteins with an elastic network model," *Biophysical journal*, vol. 80, no. 1, pp. 505–515, 2001.

[225] J. Kuriyan, G. A. Petsko, R. M. Levy, and M. Karplus, "Effect of anisotropy and anharmonicity on protein crystallographic refinement: an evaluation by molecular dynamics," *Journal of molecular biology*, vol. 190, no. 2, pp. 227–254, 1986.

[226] G. Zaccai, "How soft is a protein? a protein dynamics force constant measured by neutron scattering," *Science*, vol. 288, no. 5471, pp. 1604–1607, 2000.

[227] P. Doruker, A. R. Atilgan, and I. Bahar, "Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: Application to α-amylase inhibitor," *Proteins: Structure, Function, and Bioinformatics*, vol. 40, no. 3, pp. 512–524, 2000.

[228] S. Pundir, M. J. Martin, and C. O'Donovan, "Uniprot protein knowledgebase," in *Protein Bioinformatics*. Springer, 2017, pp. 41–55.

[229] M. Källberg, H. Wang, S. Wang, J. Peng, Z. Wang, H. Lu, and J. Xu, "Template-based protein structure modeling using the raptorx web server," *Nature protocols*, vol. 7, no. 8, p. 1511, 2012.

[230] E. Eyal, G. Lum, and I. Bahar, "The anisotropic network model web server at 2015 (anm 2.0)," *Bioinformatics*, vol. 31, no. 9, pp. 1487–1489, 2015.

[231] M. Gouw, S. Michael, H. Sámano-Sánchez, M. Kumar, A. Zeke, B. Lang, B. Bely, L. B. Chemes, N. E. Davey, Z. Deng *et al.*, "The eukaryotic linear

motif resource–2018 update," *Nucleic acids research*, vol. 46, no. D1, pp. D428–D434, 2018.

[232] H. Mi, A. Muruganujan, D. Ebert, X. Huang, and P. D. Thomas, "Panther version 14: more genomes, a new panther go-slim and improvements in enrichment analysis tools," *Nucleic acids research*, vol. 47, no. D1, pp. D419–D426, 2019.

[233] M. Kanehisa and Y. Sato, "Kegg mapper for inferring cellular functions from protein sequences," *Protein Science*, vol. 29, no. 1, pp. 28–35, 2020.

[234] C. J. Sigrist, E. De Castro, L. Cerutti, B. A. Cuche, N. Hulo, A. Bridge, L. Bougueleret, and I. Xenarios, "New and continuing developments at prosite," *Nucleic acids research*, vol. 41, no. D1, pp. D344–D347, 2012.

[235] N. E. Davey, G. Travé, and T. J. Gibson, "How viruses hijack cell regulation," *Trends in biochemical sciences*, vol. 36, no. 3, pp. 159–169, 2011.

[236] B. Uyar, R. J. Weatheritt, H. Dinkel, N. E. Davey, and T. J. Gibson, "Proteome-wide analysis of human disease mutations in short linear motifs: neglected players in cancer?" *Molecular BioSystems*, vol. 10, no. 10, pp. 2626–2642, 2014.

[237] Z. Songyang, A. Fanning, C. Fu, J. Xu, S. Marfatia, A. Chishti, A. Crompton, A. Chan, J. Anderson, and L. Cantley, "Recognition of unique carboxyl-terminal motifs by distinct pdz domains," *Science*, vol. 275, no. 5296, pp. 73–77, 1997.

[238] A. R. Atilgan, S. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar, "Anisotropy of fluctuation dynamics of proteins with an elastic network model," *Biophysical journal*, vol. 80, no. 1, pp. 505–515, 2001.

[239] L. Wei, M. Surma, S. Shi, N. Lambert-Cheatham, and J. Shi, "Novel insights into the roles of rho kinase in cancer," *Archivum immunologiae et therapiae experimentalis*, vol. 64, no. 4, pp. 259–278, 2016.

[240] S. I. Chiosea, J. R. Grandis, V. W. Lui, B. Diergaarde, J. H. Maxwell, R. L. Ferris, S. W. Kim, A. Luvison, M. Miller, and M. N. Nikiforova, "Pik3ca,

hras and pten in human papillomavirus positive oropharyngeal squamous cell carcinoma," *BMC cancer*, vol. 13, no. 1, p. 602, 2013.

[241] R. B. Birge, C. Kalodimos, F. Inagaki, and S. Tanaka, "Crk and crkl adaptor proteins: networks for physiological and pathological signaling," *Cell Communication and Signaling*, vol. 7, no. 1, p. 13, 2009.

[242] C. V. Ramana, M. Chatterjee-Kishore, H. Nguyen, and G. R. Stark, "Complex roles of stat1 in regulating gene expression," *Oncogene*, vol. 19, no. 21, pp. 2619–2627, 2000.

[243] T. Watanabe, N. Shinohara, K. Moriya, A. Sazawa, Y. Kobayashi, Y. Ogiso, M. Takiguchi, J. Yasuda, T. Koyanagi, N. Kuzumaki *et al.*, "Significance of the grb2 and son of sevenless (sos) proteins in human bladder cancer cell lines," *IUBMB life*, vol. 49, no. 4, pp. 317–320, 2000.

[244] J. L. Marín-Rubio, L. Vela-Martín, J. Fernández-Piqueras, and M. Villa-Morales, "Fadd in cancer: Mechanisms of altered expression and function, and clinical implications," *Cancers*, vol. 11, no. 10, p. 1462, 2019.

[245] K. Tazat, M. Harsat, A. Goldshmid-Shagal, M. Ehrlich, and Y. I. Henis, "Dual effects of ral-activated pathways on p27 localization and tgf-β signaling," *Molecular biology of the cell*, vol. 24, no. 11, pp. 1812–1824, 2013.

[246] S. P. Malkoski, S. M. Haeger, T. G. Cleaver, K. J. Rodriguez, H. Li, S.-L. Lu, W. J. Feser, A. E. Barón, D. Merrick, J. G. Lighthall *et al.*, "Loss of transforming growth factor beta type ii receptor increases aggressive tumor behavior and reduces survival in lung adenocarcinoma and squamous cell carcinoma," *Clinical Cancer Research*, vol. 18, no. 8, pp. 2173–2183, 2012.

[247] Y. Yonemura, S. Fushida, E. Bando, K. Kinoshita, K. Miwa, Y. Endo, K. Sugiyama, T. Partanen, H. Yamamoto, and T. Sasaki, "Lymphangiogenesis and the vascular endothelial growth factor receptor (vegfr)-3 in gastric cancer," *European Journal of cancer*, vol. 37, no. 7, pp. 918–923, 2001.

[248] M. R. Tavares, I. C. Pavan, C. L. Amaral, L. Meneguello, A. D. Luchessi, and F. M. Simabuco, "The s6k protein family in health and disease," *Life sciences*, vol. 131, pp. 1–10, 2015.

[249] T. Zhang, X. Song, X. Liao, X. Wang, G. Zhu, C. Yang, and X. Xie, "Distinct prognostic values of phospholipase c beta family members for non-small cell lung carcinoma," *BioMed research international*, vol. 2019, 2019.

[250] S. Sigismund, D. Avanzato, and L. Lanzetti, "Emerging functions of the egfr in cancer," *Molecular oncology*, vol. 12, no. 1, pp. 3–20, 2018.

[251] F. Nurwidya, S. Andarini, F. Takahashi, E. Syahruddin, and K. Takahashi, "Implications of insulin-like growth factor 1 receptor activation in lung cancer," *The Malaysian journal of medical sciences: MJMS*, vol. 23, no. 3, p. 9, 2016.

[252] K. Stankov, S. Popovic, and M. Mikov, "C-kit signaling in cancer treatment," *Current pharmaceutical design*, vol. 20, no. 17, pp. 2849–2880, 2014.

[253] R. Garg, L. G. Benedetti, M. B. Abera, H. Wang, M. Abba, and M. G. Kazanietz, "Protein kinase c and cancer: what we know and what we do not," *Oncogene*, vol. 33, no. 45, p. 5225, 2014.

[254] C. Romei and R. Elisei, "Ret/ptc translocations and clinico-pathological features in human papillary thyroid carcinoma," *Frontiers in endocrinology*, vol. 3, p. 54, 2012.

[255] J. H. Cho and J.-S. Han, "Phospholipase d and its essential role in cancer," *Molecules and cells*, vol. 40, no. 11, p. 805, 2017.

[256] R. Grose and C. Dickson, "Fibroblast growth factor signaling in tumorigenesis," *Cytokine & growth factor reviews*, vol. 16, no. 2, pp. 179–186, 2005.

[257] P. A. Muller and K. H. Vousden, "Mutant p53 in cancer: new functions and therapeutic opportunities," *Cancer cell*, vol. 25, no. 3, pp. 304–317, 2014.

[258] Q. Wei, Y. Zhao, Z.-Q. Yang, Q.-Z. Dong, X.-J. Dong, Y. Han, C. Zhao, and E.-H. Wang, "Dishevelled family proteins are expressed in non-small cell

lung cancer and function differentially on tumor progression," *Lung cancer*, vol. 62, no. 2, pp. 181–192, 2008.

[259] H. J. Kim, N. Hawke, and A. S. Baldwin, "Nf-κb and ikk as therapeutic targets in cancer," *Cell Death & Differentiation*, vol. 13, no. 5, pp. 738–747, 2006.

[260] L. García-Gutiérrez, S. McKenna, W. Kolch, and D. Matallanas, "Rassf1a tumour suppressor: Target the network for effective cancer therapy," *Cancers*, vol. 12, no. 1, p. 229, 2020.

[261] Y. Li, D.-F. Xu, D. Jiang, J. Zhao, J.-F. Ge, and S.-Y. Zheng, "Significance of fas and fasl protein expression in cardiac carcinoma and local lymph node tissues," *International journal of clinical and experimental pathology*, vol. 8, no. 9, p. 11915, 2015.

[262] S. Igelmann, H. A. Neubauer, and G. Ferbeyre, "Stat3 and stat5 activation in solid cancers," *Cancers*, vol. 11, no. 10, p. 1428, 2019.

[263] D. Q. Nguyen, D. H. Hoang, T. T. V. Nguyen, H. D. Ho, V. Huynh, J. H. Shin, Q. T. Ly, D. D. Thi Nguyen, L. Ghoda, G. Marcucci *et al.*, "Ebp1 p48 promotes oncogenic activities in human colon cancer cells through regulation of tif-90-mediated ribosomal rna synthesis," *Journal of cellular physiology*, vol. 234, no. 10, pp. 17 612–17 621, 2019.

[264] C. C. Spriggs and L. A. Laimins, "Human papillomavirus and the dna damage response: exploiting host repair pathways for viral replication," *Viruses*, vol. 9, no. 8, p. 232, 2017.

[265] J. Doorbar, N. Egawa, H. Griffin, C. Kranjec, and I. Murakami, "Human papillomavirus molecular biology and disease association," *Reviews in medical virology*, vol. 25, pp. 2–23, 2015.

[266] M. Thomas, N. Narayan, D. Pim, V. Tomaić, P. Massimi, K. Nagasaka, C. Kranjec, N. Gammoh, and L. Banks, "Human papillomaviruses, cervical cancer and cell polarity," *Oncogene*, vol. 27, no. 55, pp. 7018–7030, 2008.

[267] S. Herbster, A. Paladino, S. de Freitas, and E. Boccardo, "Alterations in the expression and activity of extracellular matrix components in hpv-associated infections and diseases," *Clinics*, vol. 73, 2018.

[268] F. A. Venning, L. Wullkopf, and J. T. Erler, "Targeting ecm disrupts cancer progression," *Frontiers in oncology*, vol. 5, p. 224, 2015.

[269] W. Zhang, H. Chen, Y. Chen, J. Liu, X. Wang, X. Yu, J. J. Chen, and W. Zhao, "Cancerous inhibitor of protein phosphatase 2a contributes to human papillomavirus oncoprotein e7-induced cell proliferation via e2f1," *Oncotarget*, vol. 6, no. 7, p. 5253, 2015.

[270] F. Facciuto, M. B. Valdano, F. Marziali, P. Massimi, L. Banks, A. L. Cavatorta, and D. Gardiol, "Human papillomavirus (hpv)-18 e6 oncoprotein interferes with the epithelial cell polarity par3 protein," *Molecular oncology*, vol. 8, no. 3, pp. 533–543, 2014.

[271] S. Cao, F. Cong, M. Tan, G. Ding, J. Liu, L. Li, Y. Zhao, S. Liu, and Y. Xiao, "14-3-3ε acts as a proviral factor in highly pathogenic porcine reproductive and respiratory syndrome virus infection," *Veterinary research*, vol. 50, no. 1, p. 16, 2019.

[272] G. Akoz, G. Diniz, S. Ekmekci, Z. Y. Ekin, M. Uncel *et al.*, "Evaluation of human epididymal secretory protein 4 expression according to the molecular subtypes (luminal a, luminal b, human epidermal growth factor receptor 2-positive, triple-negative) of breast cancer," *Indian Journal of Pathology and Microbiology*, vol. 61, no. 3, p. 323, 2018.

[273] M. O. Krisenko and R. L. Geahlen, "Calling in syk: Syk's dual role as a tumor promoter and tumor suppressor in cancer," *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, vol. 1853, no. 1, pp. 254–263, 2015.

[274] J. Li, J.-Z. Gao, J.-L. Du, Z.-X. Huang, and L.-X. Wei, "Increased cdc20 expression is associated with development and progression of hepatocellular carcinoma," *International journal of oncology*, vol. 45, no. 4, pp. 1547–1555, 2014.

[275] I. I. C. Chio, M. Sasaki, D. Ghazarian, J. Moreno, S. Done, T. Ueda, S. Inoue, Y.-L. Chang, N. J. Chen, and T. W. Mak, "Tradd contributes to tumour suppression by regulating ulf-dependent p19 arf ubiquitylation," *Nature cell biology*, vol. 14, no. 6, pp. 625–633, 2012.

[276] S. Huard, R. T. Elder, D. Liang, G. Li, and R. Y. Zhao, "Human immunodeficiency virus type 1 vpr induces cell cycle g2 arrest through srk1/mk2-mediated phosphorylation of cdc25," *Journal of virology*, vol. 82, no. 6, pp. 2904–2917, 2008.

[277] S. Stindt, S. Spitzley, R. Bartenschlager, D. Häussinger, and J. Bode, "Hepatitis c virus inhibits il-1β-induced iκbζ expression in a calpain-and mk2-dependent manner resulting in lcn2 suppression," *Zeitschrift für Gastroenterologie*, vol. 53, no. 12, p. A5_13, 2015.

[278] M. Bergvall, T. Melendy, and J. Archambault, "The e1 proteins," *Virology*, vol. 445, no. 1-2, pp. 35–56, 2013.

[279] N. Freitas and C. Cunha, "Mechanisms and signals for the nuclear import of proteins," *Current genomics*, vol. 10, no. 8, p. 550, 2009.

[280] M. Lehoux, A. Fradet-Turcotte, M. Lussier-Price, J. G. Omichinski, and J. Archambault, "Inhibition of human papillomavirus dna replication by an e1-derived p80/uaf1-binding peptide," *Journal of virology*, vol. 86, no. 7, pp. 3486–3500, 2012.

[281] C. S. Swindle and J. A. Engler, "Association of the human papillomavirus type 11 e1 protein with histone h1," *Journal of virology*, vol. 72, no. 3, pp. 1994–2001, 1998.

[282] D. Lee, H. Sohn, G. V. Kalpana, and J. Choe, "Interaction of e1 and hsnf5 proteins stimulates replication of human papillomavirus dna," *Nature*, vol. 399, no. 6735, pp. 487–491, 1999.

[283] S. Sammak, M. D. Allen, N. Hamdani, M. Bycroft, and G. Zinzalla, "The structure of ini 1/hsnf 5 rpt 1 and its interactions with the c-myc: Max heterodimer provide insights into the interplay between myc and the swi/snf

chromatin remodeling complex," *The FEBS journal*, vol. 285, no. 22, pp. 4165–4180, 2018.

[284] E. Yung, M. Sorin, E.-J. Wang, S. Perumal, D. Ott, and G. V. Kalpana, "Specificity of interaction of ini1/hsnf5 with retroviral integrases and its functional significance," *Journal of virology*, vol. 78, no. 5, pp. 2222–2231, 2004.

[285] B. Y. Lin, A. M. Makhov, J. D. Griffith, T. R. Broker, and L. T. Chow, "Chaperone proteins abrogate inhibition of the human papillomavirus (hpv) e1 replicative helicase by the hpv e2 protein," *Molecular and cellular biology*, vol. 22, no. 18, pp. 6592–6604, 2002.

[286] Y.-M. Loo and T. Melendy, "Recruitment of replication protein a by the papillomavirus e1 protein and modulation by single-stranded dna," *Journal of virology*, vol. 78, no. 4, pp. 1605–1615, 2004.

[287] Y. Zhang, A. G. Baranovskiy, T. H. Tahirov, and Y. I. Pavlov, "The c-terminal domain of the dna polymerase catalytic subunit regulates the primase and polymerase activities of the human dna polymerase α-primase complex," *Journal of Biological Chemistry*, vol. 289, no. 32, pp. 22 021–22 034, 2014.

[288] A. A. McBride, "The papillomavirus e2 proteins," *Virology*, vol. 445, no. 1-2, pp. 57–79, 2013.

[289] J. Doorbar, "The e4 protein; structure, function and patterns of expression," *Virology*, vol. 445, no. 1-2, pp. 80–98, 2013.

[290] C. E. Davy, D. J. Jackson, K. Raj, W. L. Peh, S. A. Southern, P. Das, R. Sorathia, P. Laskey, K. Middleton, T. Nakahara *et al.*, "Human papillomavirus type 16 e1ˆe4-induced g2 arrest is associated with cytoplasmic retention of active cdk1/cyclin b1 complexes," *Journal of virology*, vol. 79, no. 7, pp. 3998–4011, 2005.

[291] Q. Ding, L. Li, and P. Whyte, "Human papillomavirus 18 e1^ e4 protein interacts with cyclin a/cdk 2 through an rxl motif," *Molecular and cellular biochemistry*, vol. 373, no. 1-2, pp. 29–40, 2013.

[292] G. L. Knight, A. G. Pugh, E. Yates, I. Bell, R. Wilson, C. A. Moody, L. A. Laimins, and S. Roberts, "A cyclin-binding motif in human papillomavirus type 18 (hpv18) e1^ e4 is necessary for association with cdk–cyclin complexes and g2/m cell cycle arrest of keratinocytes, but is not required for differentiation-dependent viral genome amplification or l1 capsid protein expression," *Virology*, vol. 412, no. 1, pp. 196–210, 2011.

[293] Q. Wang, A. Kennedy, P. Das, P. B. McIntosh, S. A. Howell, E. R. Isaacson, S. A. Hinz, C. Davy, and J. Doorbar, "Phosphorylation of the human papillomavirus type 16 e1^ e4 protein at t57 by erk triggers a structural change that enhances keratin binding and protein stability," *Journal of virology*, vol. 83, no. 8, pp. 3668–3683, 2009.

[294] S. Roberts, S. R. Kingsbury, K. Stoeber, G. L. Knight, P. H. Gallimore, and G. H. Williams, "Identification of an arginine-rich motif in human papillomavirus type 1 e1^ e4 protein necessary for e4-mediated inhibition of cellular dna synthesis in vitro and in cells," *Journal of virology*, vol. 82, no. 18, pp. 9056–9064, 2008.

[295] D. DiMaio and L. M. Petti, "The e5 proteins," *Virology*, vol. 445, no. 1-2, pp. 99–114, 2013.

[296] D. Pim, M. Bergant, S. S. Boon, K. Ganti, C. Kranjec, P. Massimi, V. K. Subbaiah, M. Thomas, V. Tomaić, and L. Banks, "Human papillomaviruses and the specificity of pdz domain targeting," *The FEBS journal*, vol. 279, no. 19, pp. 3530–3537, 2012.

[297] D. Martinez-Zapien, F. X. Ruiz, J. Poirson, A. Mitschler, J. Ramirez, A. Forster, A. Cousido-Siah, M. Masson, S. V. Pol, A. Podjarny *et al.*, "Structure of the e6/e6ap/p53 complex required for hpv-mediated degradation of p53," *Nature*, vol. 529, no. 7587, pp. 541–545, 2016.

[298] V. Tomaić, "Functional roles of e6 and e7 oncoproteins in hpv-induced malignancies at diverse anatomical sites," *Cancers*, vol. 8, no. 10, p. 95, 2016.

[299] S.-C. Chen and I. Bahar, "Mining frequent patterns in protein structures: a study of protease families," *Bioinformatics*, vol. 20, no. suppl_1, pp. i77–i85, 2004.

[300] T. Z. Sen, Y. Feng, J. V. Garcia, A. Kloczkowski, and R. L. Jernigan, "The extent of cooperativity of protein motions observed with elastic network models is similar for atomic and coarser-grained models," *Journal of chemical theory and computation*, vol. 2, no. 3, pp. 696–704, 2006.

[301] R. Javier, "Cell polarity proteins: common targets for tumorigenic human viruses," *Oncogene*, vol. 27, no. 55, pp. 7031–7046, 2008.

[302] M. van Ham and W. Hendriks, "Pdz domains–glue and guide," *Molecular biology reports*, vol. 30, no. 2, pp. 69–82, 2003.

[303] M. Genera, D. Samson, B. Raynal, A. Haouz, B. Baron, C. Simenel, R. Guerois, N. Wolff, and C. Caillet-Saguy, "Structural and functional characterization of the pdz domain of the human phosphatase ptpn3 and its interaction with the human papillomavirus e6 oncoprotein," *Scientific reports*, vol. 9, no. 1, pp. 1–12, 2019.

[304] S. S. Boon, V. Tomaić, M. Thomas, S. Roberts, and L. Banks, "Cancer-causing human papillomavirus e6 proteins display major differences in the phospho-regulation of their pdz interactions," *Journal of virology*, vol. 89, no. 3, pp. 1579–1586, 2015.

[305] L. Zhan, A. Rosenberg, K. C. Bergami, M. Yu, Z. Xuan, A. B. Jaffe, C. Allred, and S. K. Muthuswamy, "Deregulation of scribble promotes mammary tumorigenesis and reveals a role for cell polarity in carcinoma," *Cell*, vol. 135, no. 5, pp. 865–878, 2008.

[306] Y. Liu, G. D. Henry, R. S. Hegde, and J. D. Baleja, "Solution structure of the hdlg/sap97 pdz2 domain and its mechanism of interaction with hpv-18

papillomavirus e6 protein," *Biochemistry*, vol. 46, no. 38, pp. 10 864–10 874, 2007.

[307] Y. Zhang, J. Dasgupta, R. Z. Ma, L. Banks, M. Thomas, and X. S. Chen, "Structures of a human papillomavirus (hpv) e6 polypeptide bound to maguk proteins: mechanisms of targeting tumor suppressors by a high-risk hpv oncoprotein," *Journal of virology*, vol. 81, no. 7, pp. 3618–3626, 2007.

[308] V. Tomaić, D. Gardiol, P. Massimi, M. Ozbun, M. Myers, and L. Banks, "Human and primate tumour viruses use pdz binding as an evolutionarily conserved mechanism of targeting cell polarity regulators," *Oncogene*, vol. 28, no. 1, pp. 1–8, 2009.

[309] H. L. Howie, R. A. Katzenellenbogen, and D. A. Galloway, "Papillomavirus e6 proteins," *Virology*, vol. 384, no. 2, pp. 324–334, 2009.

[310] M. Filippova, H. Song, J. L. Connolly, T. S. Dermody, and P. J. Duerksen-Hughes, "The human papillomavirus 16 e6 protein binds to tumor necrosis factor (tnf) r1 and protects cells from tnf-induced apoptosis," *Journal of Biological Chemistry*, vol. 277, no. 24, pp. 21 730–21 739, 2002.

[311] F. S. Oppermann, F. Gnad, J. V. Olsen, R. Hornberger, Z. Greff, G. Kéri, M. Mann, and H. Daub, "Large-scale proteomics analysis of the human kinome," *Molecular & Cellular Proteomics*, vol. 8, no. 7, pp. 1751–1764, 2009.

[312] F. Lampert, D. Stafa, A. Goga, M. V. Soste, S. Gilberto, N. Olieric, P. Picotti, M. Stoffel, and M. Peter, "The multi-subunit gid/ctlh e3 ubiquitin ligase promotes cell proliferation and targets the transcription factor hbp1 for degradation," *Elife*, vol. 7, p. e35528, 2018.

[313] N. Dyson, P. M. Howley, K. Munger, and E. Harlow, "The human papilloma virus-16 e7 oncoprotein is able to bind to the retinoblastoma gene product," *Science*, vol. 243, no. 4893, pp. 934–937, 1989.

[314] K. Münger, B. Werness, N. Dyson, W. Phelps, E. Harlow, and P. Howley, "Complex formation of human papillomavirus e7 proteins with the

retinoblastoma tumor suppressor gene product." *The EMBO journal*, vol. 8, no. 13, pp. 4099–4105, 1989.

[315] T. Z. Sen, Y. Feng, J. V. Garcia, A. Kloczkowski, and R. L. Jernigan, "The extent of cooperativity of protein motions observed with elastic network models is similar for atomic and coarser-grained models," *Journal of chemical theory and computation*, vol. 2, no. 3, pp. 696–704, 2006.

[316] S.-C. Chen and I. Bahar, "Mining frequent patterns in protein structures: a study of protease families," *Bioinformatics*, vol. 20, no. suppl_1, pp. i77–i85, 2004.

[317] X. Liu and R. Marmorstein, "Structure of the retinoblastoma protein bound to adenovirus e1a reveals the molecular basis for viral oncoprotein inactivation of a tumor suppressor," *Genes & development*, vol. 21, no. 21, pp. 2711–2716, 2007.

[318] A. L. Brass, D. M. Dykxhoorn, Y. Benita, N. Yan, A. Engelman, R. J. Xavier, J. Lieberman, and S. J. Elledge, "Identification of host proteins required for hiv infection through a functional genomic screen," *science*, vol. 319, no. 5865, pp. 921–926, 2008.

[319] I. L. Alberts, N. P. Todorov, and P. M. Dean, "Receptor flexibility in de novo ligand design and docking," *Journal of medicinal chemistry*, vol. 48, no. 21, pp. 6585–6596, 2005.

Appendix A

1. Online Link for Appendix https://tinyurl.com/ccdmjvmag

2. Also Available in CD Attached