Advancing Personalized Therapies: An In Silico and Machine Learning Framework to Mitigate HPV-Associated Cancers

A PROJECT REPORT

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF

MASTER OF SCIENCE

BIOTECHNOLOGY

(BATCH 2021-2023)

Submitted by:

Rati Bhardwaj

2k21/MSCBIO/34

Under the Supervision of

Dr Asmita Das

DEPARTMENT OF BIOTECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi -110042

CANDIDATE'S DECLARATION

I am Rati Bhardwaj (2K21/MSCBIO/34) student of M.Sc. Biotechnology, hereby certify that the work which is presented in the dissertation titled **"Advancing Personalized Therapies: An In Silico and Machine Learning Framework to Mitigate HPV-Associated Cancers"** in fulfilment of the requirement for the award of Degree of Master of Science in Biotechnology and submitted to the Department of Biotechnology, Delhi Technological University (Formerly Delhi College of Engineering), New Delhi is an authentic record of my own carried out during a period from January 2023 to May 2023, under the supervision of Dr Asmita Das, Department of Biotechnology.

The matter presented in this report has not been submitted or previously formed the basis for the award of any Degree, Diploma, Associateship, Fellowship, or other similar title or recognition.

CERTIFICATE

I hereby certify that the Project Dissertation titled **"Advancing Personalized Therapies: An In Silico and Machine Learning Framework to Mitigate HPV-Associated Cancers"** which is submitted by Rati Bhardwaj (2K21/MSCBIO/34) Department of Biotechnology, Delhi Technological University, New Delhi in partial fulfilment of the requirement for the award of the Degree of Master of Science is a record of the project work carried out by the students under my supervision and guidance.

To the best of my knowledge, this work has not been submitted in part or full for any Degree or Diploma to this University/Institute or elsewhere.

Date:

Place:

Dr Asmita Das
Professor
Department of Biotechnology
Delhi Technological University

Prof. Pravir Kumar
Head of Department
Department of Biotechnology
Delhi Technological University

ACKNOWLEDGEMENT

I take great pleasure in expressing my profound gratitude and an overwhelming sense of pride towards Dr. Asmita Das, our mentor, whose unwavering assistance and support, full dedication, invaluable guidance and motivation, and meticulous evaluation of our findings, enabled us to successfully complete this project and acquire fundamental research knowledge. I am sincerely thankful to Dr. Das for giving us the opportunity to work under their guidance.

I would also like to seize this moment to extend my appreciation to Prof. Pravir Kumar, the Head of the Department of Biotechnology, for overseeing and facilitating the utilization of all the necessary facilities, and for going above and beyond to assist students in any capacity.

We would like to express our deepest gratitude to the faculty of the Department for their unwavering support and continuous guidance throughout the laboratory work and research involved in this project. We are grateful to our families and colleagues for their kind cooperation and encouragement, which greatly contributed to the successful completion of this project.

Without the assistance of the aforementioned individuals, this project would have been much more challenging to carry out, and we would have encountered numerous difficulties along the way.

Rati Bhardwaj

2k21/MSCBIO/34

Abstract

Human Papillomavirus (HPV) is one of the principal causes of death among females in emerging nations. It is a sexually transmitted virus that shatters the cell cycle by two of its oncoproteins i.e., E6 and E7. HPV E7 oncoprotein binds to the pRb (Retinoblastoma tumor suppressor gene). This binding hinders the association of pRb with E2F causing the release of transcription factor E2F, vital for the cell cycle to proceed for the S phase thus causing abnormal unregulated cell division and growth. Precautionary HPV Vaccines are available, however infected populations are deprived of therapeutic treatments based on drugs. To rectify this issue, we try to recognize bioactive molecule inhibitors of the HPV E7 oncoprotein using bioinformatics softwares and in-silico molecular docking. In this in-silico study, flavonoid compounds were virtually screened and observed for their ability to inhibit E7 oncoprotein. The virtual screening was conducted in several steps starting with molecular docking to find their potential binding energy with the target followed by validating results with machine learning using Python. According to the literature review Baicalein is considered as a most potent inhibitor of E7 (taken as control). In this study, we worked on approximately 50 flavonoids out of which 15 have shown significant binding energies. Usually, the bioactive compounds present in flavonoids will not significantly cause a side effect. Linarin and Afzelin has shown maximum binding energy of ΔG -10.6 kcal/mol and ΔG = -9.05 respectively comparable to that of the standard inhibitor Baicalein with ΔG -7.36 kcal/mol. Docking results were further validated by machine learning. Linarin and Afzelin can be easily extracted from the Angiosperms and used as an herbal product to inhibit HPV E7 oncoprotein. The compounds identified here contribute to fascinating beginnings for further medical research and the development of therapeutic treatment.

Keywords: Human Papilloma Virus (HPV), HPV E7 oncoprotein, cervical cancer, Linarin, Flavonoids, In-silico molecular docking.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Introduction

The two major types of HPVs i.e mucosal and cutaneous consist of nearly 130 different varieties. HPV 6 & 11 are the low-risk variants that can cause benign cellular proliferation [1] while HPV16 & 18 are the high-risk ones and the possible causative for cervical cancer [2]. Moreover, HPV infection has been linked to around 35% of penile cancers, anal and vaginal cancers, as well as nearly 20% of head and neck cancer [3]. The high-risk strains of HPV produce oncoproteins E6 and E7 that interfere with the normal functioning of the p53 and Retinoblastoma (pRB) tumor suppressor proteins, resulting in dysregulation of the cell cycle and contributing to the virus's oncogenic properties.

It has also been proposed that pRb degradation is mediated by HPV E7 [4]. Although the HPV E7 oncoprotein interacts with pRb to disrupt the regulation of genes involved in the Synthesis phase of the cell cycle, in healthy uninfected cells, pRb along with E2F/DP transcription factors to control the progression of the cell cycle into the Synthesis and Mitotic phases. P53 is often post-translationally altered and stabilized once the deregulated entry into the S phase is made, making cells to enter cell apoptotic pathway [5] However, HPV E6 picks out p53 and binds to it for deterioration through the ubiquitin-proteasomal pathway [6]. The E7 oncoprotein pathway will receive the majority of attention in our paper.

About 100 amino acids make up the small, acidic polypeptides known as E7 proteins. There may be an area of sequence similarity between the E7's amino acid terminus and the entirety of CR2 of Adenovirus (Ad) E1A. The preserved Leu-X-Cys-X-Glu (LXCXE) motif within the CR2 domain is crucial for the interaction between the E7 oncoprotein and pRb, the Retinoblastoma tumor suppressor gene. This interaction requires this specific motif and is found to be abundant [7].

After pharmacokinetics and pharmacodynamics research and limited in-silico docking analysis, the only molecule that was proven to be effective in cells blocking E7 oncoprotein at concentrations of higher micro-molarity was found to be Baicalein (Control) with a binding energy of $\Delta G = -7.36$ [8]. We conducted an in-silico docking analysis to find inhibitors of the E7 oncoprotein interaction with pRb in order to find more novel and potent small molecule inhibitors because the pharmacophore-based in-silico screening campaign produced E7 inhibitors with favourable properties, despite their low potency [9]. In order to find small chemical E7 inhibitors, we devised a high throughput

docking analysis method. Two inhibitors were identified in low-micromolar to a mid-nanomolar range of IC50 values that may precisely block pRb degradation in HPV-affected cells through the screening of about 50 different flavonoid compounds. These HPV E7 inhibitors offer a foundation for the development of potential therapeutic HPV inhibitors. After docking we performed Machine Learning to verify and validate the docking results.[10]Plant polyphenols provide a variety of biochemical and pharmaceutical benefits, including pro- and anti-oxidant potentials that are related to their structure-activity relationships. In the research and development of therapies for diseases like Parkinson's disease, cancer, Alzheimer's disease, Crohn's disease and cardiovascular diseases, flavonoids, in particular, have received considerable interest.

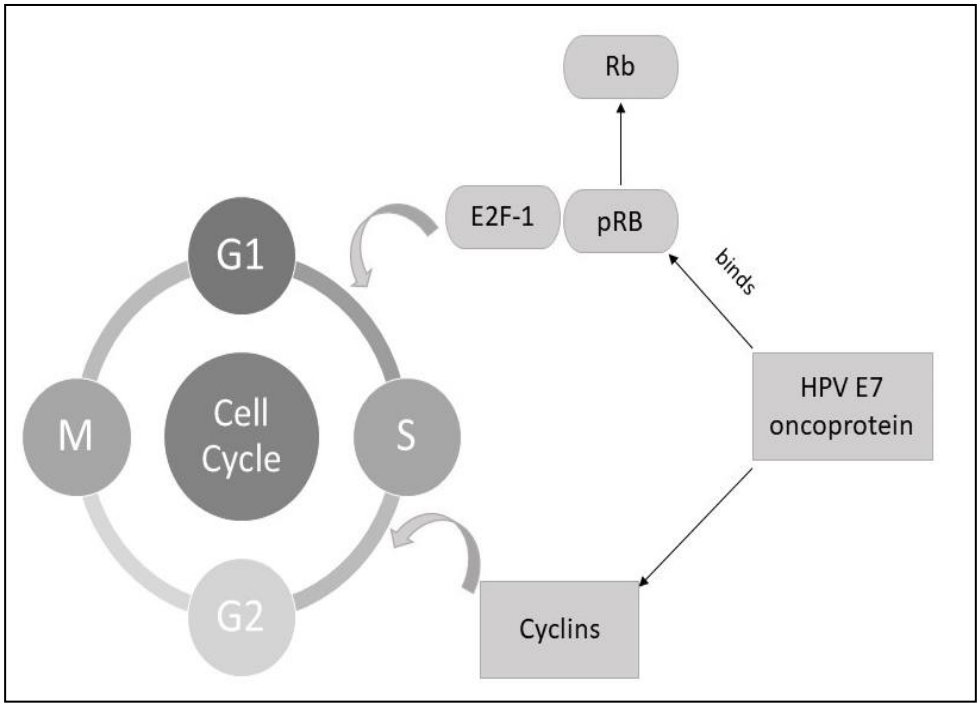


FIG. 1. Pathogenesis of HPV. The multifunctional proteins that are encoded by HPV E7 oncogenes bind predominantly to cellular pRB proteins, interfere with their regulatory cell cycle functions. pRB is tumor suppressor gene which when get activated leads to the apoptosis.

Literature Review

The third most frequent cancer overall and the seventh most common in women is cervical cancer. About 530,000 new cases were reported in 2008. The majority of these cases—more than 85%—occurred in underdeveloped countries, where cervical cancer represented 13% of all female cancers. In 2008, there were 275,000 fatalities from cervical cancer due to a mortality/incidence ratio of 52%. About 88% of the deaths took place in developing countries, which carried the bulk of this burden. Africa saw 53,000 fatalities, followed by Latin America and the Caribbean with 31,700 and Asia with 159,800. In India especially, breast carcinoma and cervical cancer continue to be the two most common cancers in females. Cervical cancer incidence in Indian women was predicted to be 27 per 100,000, whereas breast cancer incidence was 22.9 per 100,000. Around 134,420 new cases and 72,825 deaths in India in 2008 were related to cervical cancer, raising the overall mortality rate for women to above 23%.

According to a recent study on cancer mortality in India by Dikshit and colleagues, cervical cancer, which accounts for 17.1% of cases, stomach cancer, which accounts for 14.1%, and breast cancer, which accounts for 10.2%, are the top causes of fatal cancers among Indian women. [11] It is important to note that cervical cancer is regarded as the cancer that is most easily avoided.

The main cause of cervical cancer is infection with the common sexually transmitted virus human papillomavirus (HPV), which is also known as HPV. Through compelling evidence derived from molecular epidemiology studies, case-control studies, cohort studies using cervical intraepithelial neoplasia (CIN) 2 and 3 as endpoints, as well as screening studies, the correlation between HPV infection and cervical cancer has received significant recognition.[12]These many studies support the link between HPV infection and cervical cancer as a cause of the disease.[13]

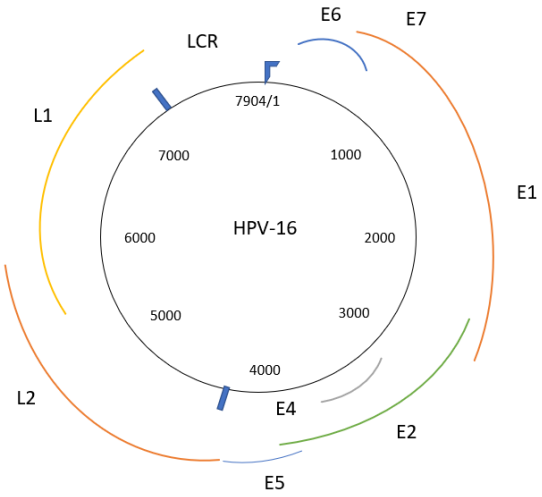


FIG. 2. HPV-16 genome

HPVs have circular, double-stranded DNA genomes that are about 8 kb in size. Eight genes make up these genomes, with the transforming abilities of E6 and E7 making them particularly important. The E6 and E7 proteins serve a variety of purposes, including transmembrane signalling, controlling the cell cycle, transforming established cell lines, immunisation of primordial cell lines, and demonstrating chromosomal integrity. These viral E6 and E7 oncoproteins are essential for malignant conversion. It has been suggested that these viral proteins cause tumours through interacting with the tumour suppressor p53 and pRB, respectively, in the case of high-risk HPV E6 and pRB, respectively. It has been suggested that these proteins cause tumours through pRB.

The 158 amino acids that make up the E6 protein are flanked by two zinc-finger binding motifs. The tumour suppressor protein p53 is thought to deteriorate as a result of the E6 protein, which is thought to promote cell proliferation. This process is carried out by a complex formed by the cellular ubiquitination enzyme E6-AP, E6, and p53. E6 causes the normal biological functions of p53 to be disrupted, which then affects the control of cell cycle progression and promotes the growth of tumour cells. Therefore, E6-mediated degradation of p53 disrupts its crucial function in regulating cell division, ultimately resulting in enhanced tumour cell proliferation.[14].

It is well known that high-risk HPV E6 has the ability to cause p53 to deteriorate, which results in the virus changing cells.Similarly, compared to E7 proteins encoded by low-risk HPV types like HPV 6 and HPV 11, those from high-risk HPV types like HPV 16 and HPV 18 show a noticeably greater affinity for binding to the Rb protein. The 'pocket domains' of the Rb protein are a specific area where E7 binds.The 'pocket domain' regions in the Rb protein are essential to the protein's capacity as a tumour suppressor.

Rb's capacity to bind to transcription factors of the E2F family and control the expression of genes involved in replication is one of its primary biological roles. Through its association with E2F, Rb functions as a repressor of replication enzyme genes. Rb's tumour suppressor function, which aids in regulating and limiting the replication process, is linked to the repression of gene expression that it causes.[15]. Rb, a tumour suppressor, can stop replication enzyme-related genes from being expressed. However, the link between Rb and E2F is disrupted when E7, a protein prevalent in high-risk HPV strains, is present. The E2F factors are consequently released and activated, encouraging cell replication and division. This discovery fits with the observation that E7-expressing keratinocytes can reproduce even after going through differentiation. [16]. As a result, the interaction between tumour suppressor genes and oncogenes plays a crucial role in the transformation of cells by interfering with the normal operations of the tumour suppressor gene products. The regulatory mechanisms that regulate cell growth, division, and other cellular functions are altered as a result of this interaction. Normal processes are subsequently interfered with, which causes uncontrollable cell growth and the development of malignant tumours.

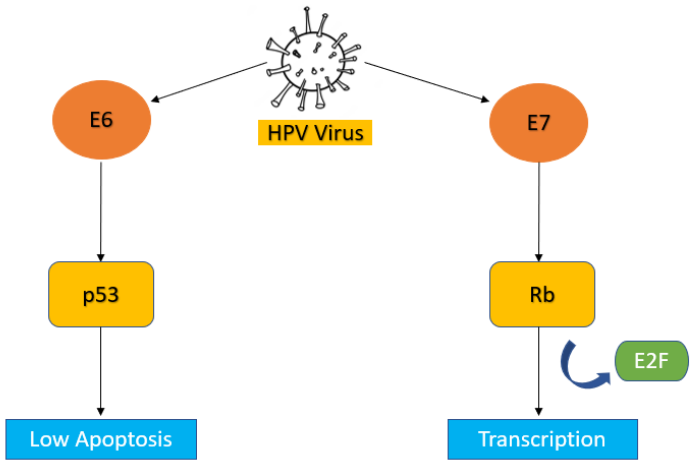


FIG. 3. Action of E6 and E7 oncoprotein on Rb and P53

HPV16 is a high-risk genotype among the many HPV genotypes. Its genomic size is 7.9 Kb, and it contains the genetic code for several viral oncoproteins, including E6 and E7. These oncoproteins are crucial for the virus's DNA insertion into the host cell and subsequent replication. The formation of new blood vessels, cell proliferation, metastasis (the spread of cancer to other parts of the body), and increased activity of telomerase (an enzyme that regulates the length of protective caps on chromosomes called telomeres) are all cancer-related markers that are primarily triggered by E6 and E7. This is significant. These cellular disruptions caused by E6 and E7's synchronised activity can hasten the onset and spread of malignancies linked to HPV.[17], and at the same time, the anti-apoptotic proteins such as BCL2, XIAP, LIVIN[18] [19], etc. upregulate in the HPV-infected cancer cells can be a targeted therapy to induce cell survival.

There are currently just a few therapy options for HPV-induced cervical cancer. The most promising method for avoiding HPV infection that lasts is vaccination. Nevertheless, their widespread use is hampered by their high cost and the presence of several viral variations, which present difficulties for ensuring universal vaccine coverage.[11] It is important to note that cervical cancer is regarded as the most preventable malignancy.

The development of drug discovery screening strategies has been greatly accelerated by recent advancements in computer-aided techniques. It is possible to significantly reduce expenses and speed up the process by using in silico docking investigations, which entail

virtual screening of prospective medications before wet-lab confirmation. The use of bioinformatics is essential in many areas related to drug discovery. [20]. More and more individuals are becoming aware of the safety and efficacy of natural remedies, particularly when compared to synthetic ones. Natural chemical extraction, particularly from plants, can be costly and time-consuming. In order to find new drugs, it is advantageous to screen a small number of potentially potent compounds from a vast population of molecules. This procedure can be improved by employing targeted molecular docking methods. Numerous studies have demonstrated that natural products are excellent at focusing on proteins that combat HPV. Prasasty [11] used the molecular docking technique to identify five natural compounds that are effective against HPV proteins. Recently, Kotadiya [21] analysed 17,944 natural compounds using this method to find anti-HPV E6 and E7 drugs.

Materials

This is a brief description of softwares and database that are used for the in-silico molecular docking and machine learning.

IMPPAT database

An extensive database that focuses on the medicinal plants found in India is called Indian Medicinal Plants, Phytochemistry and Therapeutics. Researchers, scientists, medical professionals, and anybody else interested in learning about or using the therapeutic characteristics of Indian plants might benefit greatly from IMPPAT. The database includes comprehensive details on a variety of plant species, such as their botanical names, common names, traditional applications, chemical components, pharmacological activity, and pertinent scientific studies. IMPPAT plays a critical role in encouraging the discovery and use of Indian medicinal plants in light of the rising interest in natural goods and conventional treatment. It makes it easier to conduct studies on the phytochemical analysis of plant extracts, the discovery of bioactive substances, and the creation of plant-based medicines.[22]

IMPPAT contributes to the in-silico drug development process by offering a comprehensive database of information on the phytochemical makeup of Indian medicinal plants. Information about the chemical components found in various plant species is included in this material. Researchers can anticipate the probable biological

functions and characteristics of these chemicals, such as their interactions with certain proteins or enzymes, by computer analysis.

Additionally, the database facilitates the computational modeling-based investigation of structure-activity correlations (SAR). Researchers can create prediction models to anticipate the biological activities of comparable substances by analysing the chemical structures of bioactive chemicals and their related biological activities. This makes it possible to find interesting chemicals and synthesise or modify them to improve their qualities and increase their efficacy and decrease their toxicity. Molecular docking studies and virtual chemical libraries are also included in IMPPAT's in silico capabilities. The database may be used as a springboard for creating digital collections of plant-derived chemical derivatives. These libraries can be utilised for virtual docking experiments where computer algorithms anticipate the binding affinity between a chemical and a target protein or for screening against certain therapeutic targets. The discovery of possible medication candidates and the improvement of their binding interactions are both facilitated by such simulations.

PDB

The protein data bank is large worldwide repository of structural data of large molecules such as nucleic acid and protein Molecules. It was created in 1971 but initially there was no sufficient data available until 1980, the number of deposited structures start to increase due to improved technology like crystallographic process and nuclear magnetic resonance (NMR), Cryoelectron microscopy, and theoretical modelling. PDB has an imperative role in structure genomics as well as structural biology and bioinformatics, there are many such as CATH and SCOP which use protein structure deposit in the PDB. PDB contain a large collection of 3D structure of nucleic acid and protein as it stores more than 180000 structure of macromolecules.

It is an invaluable resource for student, researcher and educator worldwide as it offers a pervasive and diverse range of protein structure and enabling researcher to study an interaction, function, and their folding. The data of structure has been determined and obtain from various techniques such as electron microscopy, magnetic resonance spectroscopy, crystallography etc. Data has been undergone through various rigorous validation to ensure its reliability and precision.

PDB not only contain Protein and nucleic data but also encompasses ligand and small molecules such that play a crucial role in biological process and activity. PDB is very helpful in studying protein-protein interaction, investigate complex molecular interaction and drug discovery, enzyme engineering, and fundamental studies of biological mechanism. Additionally, the PDB provides a variety of resources and tools to aid with data analysis and visualisation. PDB provides an interactive molecular viewer, search functionalities and advance query options. These tools help scientist and researchers to study a structure nuance of nucleic acid and protein molecules and compare molecule structure and gain the knowledge of their functional properties.

SwissDock

The Affinity of a Ligand to a protein molecule is done by protein-ligand docking analysis with the aim to predict the score and binding mode. SwissDock is a platform that provides tools for protein-ligand docking, which is a computational method used to predict the binding of small molecules (ligands) to target proteins. This resource is widely used in the field of structural bioinformatics and drug discovery.

SwissDock utilizes the well-known docking algorithm, AutoDock, to perform virtual screening and molecular docking simulations. It allows researchers to investigate the interactions between ligands and target proteins, providing insights into potential binding sites and the strength of the binding affinity. By predicting how a small molecule will bind to a specific protein, scientists can better understand the molecular mechanisms underlying biological processes and develop new therapeutic agents.

The SwissDock platform offers a user-friendly interface that guides researchers through the process of uploading protein and ligand structures, setting up the docking parameters, and analyzing the results. It provides tools for visualizing the predicted binding modes, calculating binding scores, and exploring the protein-ligand interactions. The output generated by SwissDock helps researchers prioritize and select potential drug candidates for further experimental validation.

One of the notable features of SwissDock is its integration with other bioinformatics resources and databases. It incorporates data from the SwissProt protein sequence database, the PDB protein structure database, and other relevant resources to enhance the accuracy and reliability of the docking predictions. This integration ensures that users have access to the latest information and resources necessary for their docking studies.

Furthermore, SwissDock offers additional features to enhance the usability and efficiency of the platform. It allows users to save and manage their docking jobs, facilitating easy retrieval and analysis of previous results. The platform also provides access to a range of tutorials, documentation, and support materials to assist users in understanding and utilizing the software effectively.

In summary, SwissDock is a web server and bioinformatics resource that specializes in protein-ligand docking. It offers a user-friendly interface, integration with relevant databases, and tools for analysis and visualization. SwissDock is a valuable resource for researchers in the field of structural biology and drug discovery, enabling them to perform virtual screening and docking simulations to explore protein-ligand interactions and aid in the development of novel therapeutics.

PLIP

Protein-Ligand Interaction Profiler (PLIP) is a powerful computational tool designed to investigate and analyze the interactions between proteins and small molecule ligands. With the rise of structural biology and the increasing availability of protein-ligand complex structures, understanding these interactions has become crucial for drug discovery, molecular biology, and bioinformatics research.

PLIP serves as a comprehensive and user-friendly platform for characterizing the intricate interplay between proteins and ligands. It incorporates various algorithms and methods to dissect the binding modes, binding sites, and binding affinities of ligands to proteins. By employing a wide range of structural analysis techniques, PLIP provides valuable insights into the three-dimensional nature of these interactions, helping researchers elucidate their functional and mechanistic implications.

One of the primary features of PLIP is its ability to automatically identify and classify ligand-binding sites on protein structures. By employing sophisticated algorithms, it scans the protein surface, identifies potential binding pockets, and assigns them to specific ligand classes. This functionality allows researchers to quickly identify potential binding sites and prioritize them for further analysis, accelerating the drug discovery process.

Furthermore, PLIP employs molecular docking algorithms to predict the binding poses of ligands within protein structures. These docking simulations provide a computational estimation of the most favorable spatial arrangement between the protein and ligand,

offering valuable insights into the binding affinity and potential interaction mechanisms. This information aids in the rational design of novel ligands and the optimization of existing drug candidates.

PLIP also offers extensive visualization capabilities, allowing researchers to explore the protein-ligand interactions in a highly intuitive manner. It generates interactive 3D visualizations of protein structures, highlighting the key residues involved in ligand binding and providing a comprehensive overview of the binding modes. These visual representations facilitate the interpretation of complex interactions and aid in the communication of findings to a broader scientific audience.

PLIP takes input as a PDB format or any structure from RCSB PDB server can be processed by conferring the free text search in ligand-protein complex or via a four letter PDB Id. Moreover, the outcome file generated by In-silico docking result could be uploaded on PLIP to check their covalent and non-covalent interaction.

Output of PLIP shows a typical analysis of ligand-protein complex. For each interaction of ligand and protein PLIP provide a 2D and 3D interaction diagram and a table displaying the Amino acid of protein which were involving in covalent or hydrogen interaction. PLIP also allow you to download the result file as a PNG and PyMOL format. the detail of the interaction patter can be accessed by clicking on the overview of diagram.

Visual Discovery studio

Visual Discovery Studio is an innovative platform that facilitates the exploration and synthesis of art and science. It serves as a creative space where artists, designers, and scientists can collaborate to visually represent complex concepts, data, and ideas. The studio employs a range of tools and technologies, including traditional art supplies and digital design software, to create visually stunning and scientifically accurate visualizations.

Visual discovery studio has been a powerful protein-ligand interaction visualization tool. it is used to prepare the protein and ligand for docking analysis and visualize the result. Protein preparation involves the addition of poler hydrogen group and cutting un-necessary ligand or peptide chain and removing water molecules from protein molecules.

Visual discovery studio give us a coherent display of binding intricacy of protein and ligand it also provide a 2D diagram of interaction with differentiating the type of bond ligand and protein molecules forming.

Swiss ADME

In order to chemically synthesize, develop, test and optimize a drug, one has to access various parameters like biological activity, toxicity and the concentration etc. Pharmacokinetic assessment at the beginning of the discovery phase significantly lowers the likelihood of clinical-phase ADME (Absorption, Distribution, Metabolism, Excretion) -related failures. Swiss ADME is a renowned web tool not only for its reliability and robustness but also for its simplified result analysis that enables effective incorporation to drug discovery via molecular design. The tool provides a variety of input options like molecular structure and canonical smiles, analysis of a variety of molecules, the ability to save and share results and user-friendly interactive graphs including the boiled egg and the bioavailability radar. The boiled egg aids in evaluating the gastrointestinal absorption and BBB penetration. The white region or the albumin region shows the compounds that are most likely to be passively absorbed by the gastrointestinal tract whereas the yellow yolk region shows the compounds that possess high BBB permeability. The bioavailability radar on the other hand possesses a pink area which depicts the optimal range of properties like flexibility, saturation, solubility, polarity, molecular weight and lipophilicity. A compound's radar should fall in the pink area for it to be considered as a good drug candidate.

Molinspiration

A software platform termed Molinspiration provides a variety of computational tools intended to assist in molecular analysis and alteration. These tools aids in bioactivity anticipation, data visualization and virtual screening. Most important analysis carried out by Molinspiration in drug development and optimization is the bioactivity score anticipation for drug targets (GPCR, KI, ICM and NM scores). Molinspiration takes input files in the SMILES or SDfile formats to carry out these analyses. The molecular structure of a substance is represented via SMILES (Simplified Molecular Input Line Entry System), a string-based syntax, and SDfile (Structure-Data File), a file format that is frequently used to store and transmit chemical structures and related data.

Molinspiration implements its computational techniques on the provided input file to calculate the bioactivity scores for the chosen therapeutic targets. These rating systems offer information about the molecule's potential activity or affinity towards the target, assisting in the evaluation of its potential as a therapeutic candidate. The drug development process can be expedited by using this feature to prioritise molecules for additional experimental testing.

Lipinski's rule of five

The rule of five aids to access a compound's drug likeness or efficacy based on two extremely significant factors: a drug's permeability and oral bioavailability. The Lipinki's requirements base themselves on the findings of various studies that prove that easily absorbing orally administered drug attributes fall within these constraints: Mass<500 Da, H-bond donors<5, Hbond acceptors<10, LogP value<5.
The compounds that deviate from these ranges possess a greater probability to show undesirable pharmacokinetic attributes like rapid metabolism, less availability, less permeability and fail to cross the cell membranes. Hence, the rule of five is a gold standard strategy in drug development and optimization techniques that screen out and rank compounds according to their tendency to be orally active.[23]

ChEMBL Database

ChEMBL is a helpful and widely used database in the field of medicinal chemistry and drug development. It provides a comprehensive collection of information on bioactive substances, their targets, and the biological processes associated with them. This resource has substantially sped up the process of developing new medications by making it possible to conduct research on chemical compounds and their interactions with biological systems.

ChEMBL is a substantial, open-access drug database that seeks to gather data from the healthcare and pharmaceutical sectors as well as from the process of studying and creating medications. A large number of medications and ligands have been tested on a variety of proteins and biomolecules using ChEMBL. It keeps track of small molecule and medication information as well as data on their biological activities from several medical

chemistry publications. To give researchers access to comprehensive information, bioactivity data are shared with other databases including BindingDB and PubChem Bioassay

Research papers from a range of publications, such as Journal of Medicinal Chemistry, Bioorganic Medicinal Chemistry Letters, and Journal of Natural Products, are mined for vital activity data. The chosen journals have been carefully chosen to ensure the efficient use of resources while acquiring a substantial amount of reliable data, even though they do not cover every scenario. Each article's database abstracts provide details about the tested chemicals, the experiments that were conducted, and any pertinent target data.

ChEMBL provides researchers with access to a multitude of data, such as details on pharmacological profiles, binding affinities, and compound structures. The database includes data from a variety of sources, including public databases, patents, and scientific publications. With its extensive coverage and continuous updates, ChEMBL offers researchers a robust platform for identifying potential drug targets, investigating structure-activity relationships, and designing new compounds. [24]

One of ChEMBL's main benefits is its user-friendly interface, which makes it simple for researchers to search for and collect data. Users have the ability to conduct advanced searches, filter results using predetermined criteria, and obtain thorough annotations for certain substances and targets. The database's capabilities for data visualisation and analysis allow researchers to get insights from the vast amount of information available. ChEMBL has significantly aided in the advancement of drug development and research. By merging information from many sources, it has grown to be a valuable tool for researchers all around the world. The database promotes transparency and collaboration by making its contents freely accessible, allowing researchers to use the most recent information and build upon prior work.

Data access form ChEMBL

Data retrieval is made easier by the ChEMBL interface's simplicity. Users can input a keyword, protein name, ChEMBL target identifier, or UniPort accession of an interesting target for which a ligand needs to be found in the interface's search tool.

Once users have obtained a target or a number of targets of interest from the ChEMBL database, they may rapidly access the supporting bioactivity data using a drop-down menu. Using this user-friendly capability, customers may view all the data that is accessible or create filters to choose only particular activity types. Users can choose, for example, to focus on specific ADMET endpoints or only include IC50 and Ki data below a specific concentration threshold.

The following bioactivity table, which also contains information on the specific salt form used in the experiment, provides a detailed description of each studied medication. Additionally, it comprises details about the test, a description of the target (including the organism), as well as its description, units, and the type of observed activity. It should be noted that the table includes a link that points directly to the article from which the data were derived, ensuring accessibility and transparency to the original source. Researchers may immediately export the data from this view as a text file or spreadsheet to make additional research and analysis easier. Users may utilise this tool to go further into the data, conduct their own study, and extrapolate significant findings from the discovered bioactivity data.

Methods and Methodology

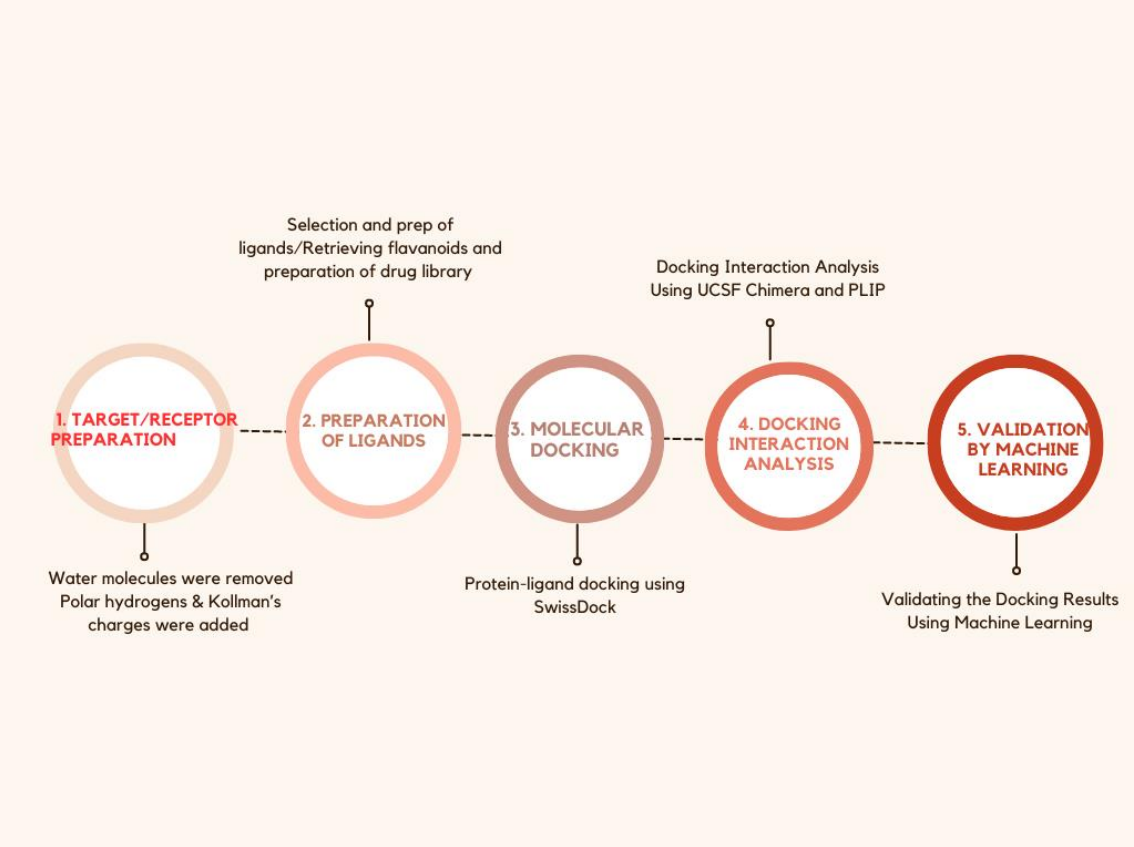


Fig. 4.: Overview of protocol followed.

A. ***Pharmacological Properties:***

The study of compound, its composition and its effects on body constitute pharmacological properties. SMILES, Molecular Formula, Molecular Weight, Number of Hydrogen Acceptors and Donors are among the Pharmacological and Physicochemical properties prediction of a potent flavonoid compound. Studying secondary qualities included Leadlikeness, solubility, Lipinski, lipophilicity, and Bioavailability score.

B. ***Ligand Receptor Preparation:***

The in-silico ligand structure data was obtained from Pubchem (https://pubchem.ncbi.nlm.nih.gov/) and the receptor studies were obtained from Protein Data Bank (PDB) https://www.rcsb.org/. The 3-D structure of target oncogene E7 i.e., 6E1I (PDB DOI: 10.2210/pdb6E1I/pdb) and 15 ligand that are flavonoids Linarin,

Epigallocatechin, Flavopiridol, Quercetin, Acacetin, Fisetin, Naringenin, Genistein, Afzelin, Apigenin, Catechin, Kaempferol, Chrysin, Wogonin and Daidzein were prepared for docking using Biovia Discovery Studio Visualizer (https://discover.3ds.com/discovery-studio-visualizer-download). Preparing the ligand and receptor requires removing hetero atoms from substances like water, non-amino acid groups, and other ligand compound.

C. ***Molecular Docking:***

In-silico Molecular Docking is usually started with the designing of ligands and receptors. Selected flavonoids were docked against the target protein using SWISSDOCK online server (http://www.swissdock.ch/docking#). The control inhibitor i.e., Baicalein was also docked against the target protein.

D. ***Docking Interaction Analysis***

Affinity binding values (△G) were saved and analysed. Flavonoid compounds and the selected target (E7 oncoprotein) interactions were analyzed by UCSF Chimera. (https://www.cgl.ucsf.edu/chimera/). UCSF Chimera offers interactive molecular structure visualisation and analysis through the study of amino acids and associated information such as density maps and sequence alignments. The binding energies of ligands and control were compared and flavonoids with high binding energies were selected.

E. ***Pharmacokinetic and Drug Likeness Screening of selected Phytochemicals***

Only flavanoids that can get past the initial round of binding energy range screening are subjected to further analysis based on their drug likeness and pharmacokinetics.

- **Lipinski's RO5 analysis:** The rule of five assesses drug likeness, or the likelihood that a molecule will be active when taken orally. The selected phytochemicals were subjected to Lipinski's RO5 to analyze their oral activity.
- **In-silico bioavailability analysis:** To comprehend the pharmacokinetics of a drug, it's vital to understand its absorption, distribution, metabolism, and excretion. In order to evaluate the bioavailability radars of the phytochemicals, SwissADME and admetSAR were employed. By providing canonical SMILES of phytochemicals as input, these values can be determined**.**
- **Bioactivity score:** To determine the druggability characteristics of ligands like NRL, PI, and EI, GPCR, ICM and KI, bioactivity score is required. The scores

can be predicted by providing canonical SMILES of phytochemicals as input to Molinspiration**.** **(**Molinspiration Cheminformatics free web services, https://www.molinspiration.com, Slovensky Grob, Slovakia)

F. ***Validating the Docking Results using Machine Learning***

A model for machine learning the Random Forest classifier was developed using 750 E7 oncoprotein inhibitors. This data was collected from the ChEMBL [25] data base by training the model with 307 molecular discriptors of above mention 750 inhibitory compounds using the python module Chembl_webresource_client (https://pypi.org/project/chembl-webresource-client/). The physical and chemical characteristics of a chemical substance are expressed as numerical numbers by molecular descriptors. Using the Python library Padelpy, the compound's molecular descriptors were produced. The model's training set consisted of these 307 molecular descriptors. 20:80 was used to divide the data into training and testing sets. Then, in order to identify the potential E7 oncoprotein inhibitor, the data was loaded into the machine learning RandomForestClassifier of the Python module scikit-learn. The model's cross-validation score ended out to be 85.6. The Python library Padelpy produced the molecular descriptors for the flavanoids. After that, the inhibitory phytochemicals were found using the trained machine learning model.

G. ***Machine Learning Analysis***

The ChEMBL database was used to estimate the inhibitory potential of Linarin, a flavanoid compound and we chose the RF technique to train the target's structural interaction fingerprints. With the purpose of creating 307 dimensional features for 730 natural substances, latent knowledge, molecular interactions, and chemical property aspects were combined. The machine learning model was trained using the created characteristics and verified data on pharmaceutical indications[26]. When the properties of natural compounds were utilised as input, the trained model effectively predicted prospective efficacies with high accuracy, sensitivity, and specificity. The RF models prioritised phytochemicals with E7 binding potential: Linarin and Afzelin.

Fig. 5.: General Work Flow of Machine Learning Mode

Results

The compounds in the Pubchem database were used for virtual screening; after that, it was filtered using Lipinski and Veber rules, physicochemical criteria, molecular characteristics, and drug similarity. In order to choose suitable molecules with membrane permeability and based on absorption in the human body, physicochemical parameters were applied (Wen et al., 2015). We selected the final 15 compounds using the aforementioned filtering criteria, and then we docked these candidate ligands using SwissDock software. Out of the 15 ligands, two hits had good docking scores, predicted IC50 values, and significant interactions with crucial residues in the target E7 oncoprotein's binding site. Figure depicts the consensus molecular docking process schematically.

Our study evaluated the potential of flavonoids as inhibitors of the HPV E7 oncoprotein by predicting the Gibbs free energy using Docking Score predicted values. The results suggest that flavonoids are likely to interact significantly with the E7 oncoprotein, with two natural inhibitors, Linarin and Afzelin, showing significant binding affinity.

Table 1. In-silico docking results of ligands that are flavonoid compounds against the HPV E7 oncoprotein target.

S.No.	**Ligands**	**ΔG**	**Amino acids**
1.	Linarin	-10.6	Phe 80, Lys 82, Phe151, Ser 81, Ala 354, Thr 376, Tyr 50, Pro 378, Leu 89, Phe 384, Ser 79, Asp 173, Asp 172, Tyr 87, Thr 174, Gly 382, Gln 175, Tyr 170, Tyr 352, Ser 76
2.	Afzelin	-9.05	Asp 122, Pro 38, Thr 333, Hsd 331, Pro 39, Arg 21, Asp 120, Trp 119, Ser 35, Trp 347, Leu 349, Hsd 37, Tyr 34, Arg 377, Gln 385
3.	Hesperetin	-7.81	Gln 175, Tyr 170, Pro 378, Phe 384, Tyr 87, Ala 354, Thr 376, Glu 375, Thr 174, Asp 173, Leu 89, Tyr 50, Ser 81, Thr 52

4.	Epigallactechin	-7.77	Glu 375, Gln 53, Thr 52, Ser 81, Tyr 50, Pro 378, Tyr 170, Asp 173, Asp 172, Tyr 87, Phe 284, Ala 354, Thr 376
5.	Flavopiridol	-7.77	Pro 193, Ala 192, Arg 191, Gln 175, Asp 173, Tyr 352, Phe 384, Thr 383, Pro 378, Gly 382, Gly 380, Thr 174, Asp 381, Ala 379, Leu 89, Thr 321, Arg 324
6.	Acacetin	-7.70	Gln 175, Pro 378, Tyr 170, Phe 384, Asp 173, Ala 354, Thr 376, Tyr 87, Tyr 50, Ser 81, Thr 52, Glu 375, Gln 53
7.	Fisetin	-7.63	Gln175, Thr 174, Tyr 87, Phe 151, Asp 173, Tyr 170, Asp 172, Arg 149, Lys 82, Phe 384, Tyr 50, Pro 378, Leu 89
8.	Naringenin	-7.62	Gln 175, Thr 174, Thr 52, Pro 378, Ser 81, Phe 384, Tyr 87, Tyr 170, Asp 173, Glu 375, Gln 53, Tyr 50, Ala 354, Thr 376
9.	Genistein	-7.52	Asp 173, Pro 378, Ala 354, Tyr 50, Thr 376, Glu 375, Gln 53, Thr 52, Ser 81, Tyr 87, Phe 384, Tyr 170, Thr 174, Gln 175
10.	Apigenin	-7.51	Ala 354, Thr 376, Tyr 170, Pro 378, Phe 384, Tyr 87, Asp 173, Tyr 50, Ser 81, Thr 52, Glu 375, Gln 53
11.	Catechin	-7.49	Phe 384, Tyr 170, Thr 174, Gln 175, Ser 76, Arg 178, Asp 77, Arg 191, Met 75, Glu 74, HSe 91, Leu 89, Gly 380, Tyr 87, Pro 378, Asp 173
12.	Kaempferol	-7.48	Gly 380, Leu 89, Pro 378, Ser 76, Tyr 50, Phe 384, Ser 81, Lys 82, Phe 151, Arg 149, Asp 172, Tyr 170, Asp 173, Tyr 87, Gln 175
13.	Chrysin	-7.45	Tyr 170, Phe 384, Asp 173, Tyr 87, Tyr 50, Ser 81, Thr 52, Glu 375, Gln 53, Thr 376, Ala 354, Pro 378
14.	Wogonnin	-7.27	Thr 376, Tyr 50, Leu 89, Pro 378, Phe 384, Gln, Thr 174, Tyr 87, Tyr 170, Asp 172, Asp 173, Phe 151, Lys 82, Ser 81
15	Daidzein	-7.23	Glu 375, Lys 82, Ser 81, Asp 173, Tyr 170, Ala 354, Phe 384, Thr 376, Tyr 87, Tr 50, Phe 86, Thr 52, Ser 85, Gln 53

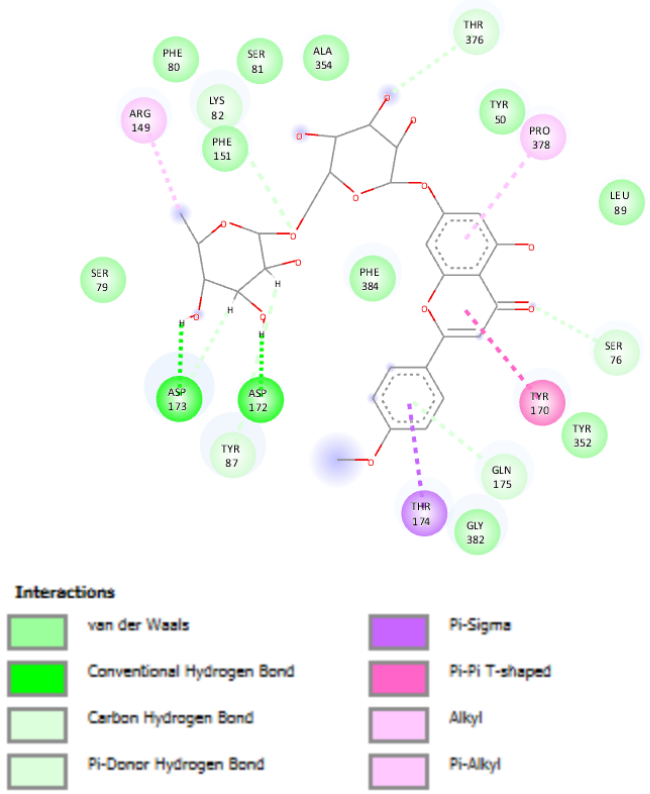
Linarin demonstrated the highest binding affinity among all 15 ligands tested, with a ΔG value of -10.6 kcal/mol, furthermore Afzelin has shown ΔG value of -9.05. This indicates that Linarin and Afzelin has a strong binding affinity with the E7 oncoprotein and is a promising inhibitor of the target. The ligand-target complex involved several important amino acids, including Phe384, Ala354, Ser81, Thr174, Leu89, and Asp involved binding with Linarin and Asp 122, Pro 38, Thr 333, Hsd 331, Pro 39, Arg 21, Asp 120, Trp 119, Ser 35, Trp 347, Leu 349, Hsd 37, Tyr 34, Arg 377, Gln 385 involved with Afzelin. The obtained binding energies of Linarin and Afzelin indicates the strength of the interactions between each flavanoid and the target protein. Lower binding energies generally suggest stronger binding affinity and a higher likelihood of forming stable complexes.

Following a successful docking analysis, this study evaluated the drugs further to determine their pharmacodynamics and pharmacokinetics characteristics utilising bioinformatics software tools including swisADME and Molinspiration. In accordance with the Lipinski rule of five, which assesses drug-likeness based on physicochemical parameters, Linarin and Afzelin showed no deviations.

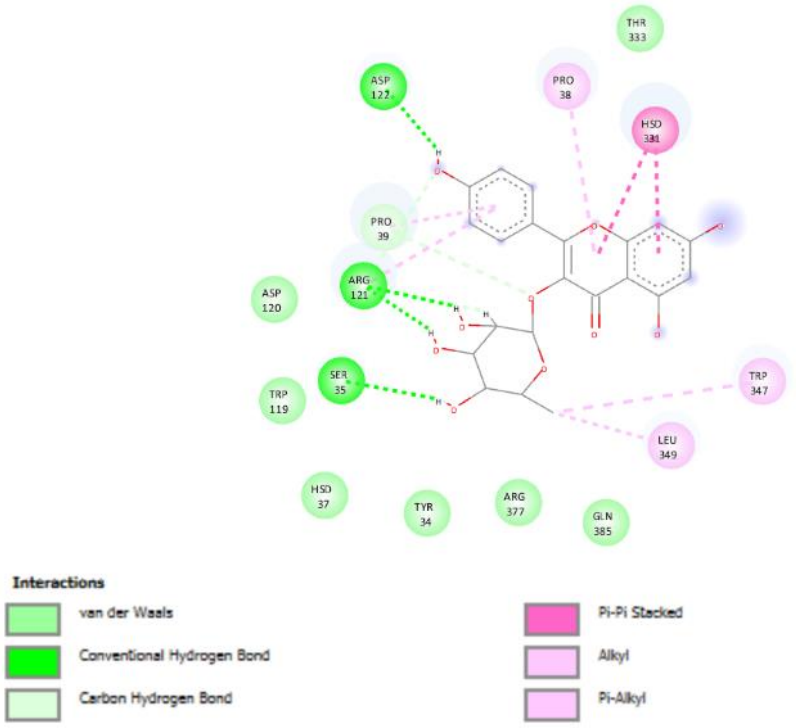
In addition, the MOLinspiration software's pharmacokinetics analysis included a number of aspects, including GPCR, kinase inhibitor, enzyme inhibitor, NRL (Nuclear Receptor Ligand), and protease inhibitor. These investigations shed light on the inhibitory activity of the medications as well as any potential interactions between them and particular biological targets. These bioinformatics analysis' results have provided important information about the pharmacokinetics and pharmacodynamics of Linarin and Afzelin. Understanding the potential efficacy and safety profiles of these compounds is made simpler due to this thorough assessment, which also serves to direct future research and development efforts in the creation and improvement of new drugs.

A machine learning study was done on the selected phytochemicals to verify our docking results. The RandomForestClassifier was chosen among the many machine learning models because it demonstrated the best accuracy and F1 value, showing its usefulness in predicting the activity of the chemicals.

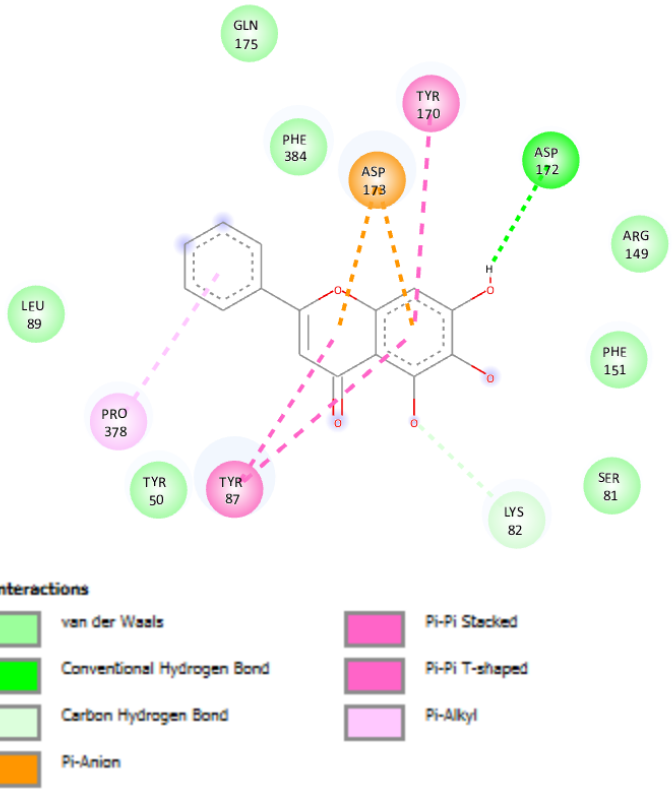
The study revealed two flavanoids as possible active compounds for inhibiting the E7 oncoprotein using the RandomForestClassifier model. They were Linarin and Afzelin, two phytochemicals. The machine learning algorithm made predictions about the compounds' possible inhibitory activity against the target protein using a variety of the compounds' properties and characteristics. The potential therapeutic use of these flavanoids to target the E7 oncoprotein has been demonstrated by this machine learning-based technique. The active substances discovered might make excellent candidates for further experimental validation and development as potential E7 inhibitors.



a. LInarin interaction with E7



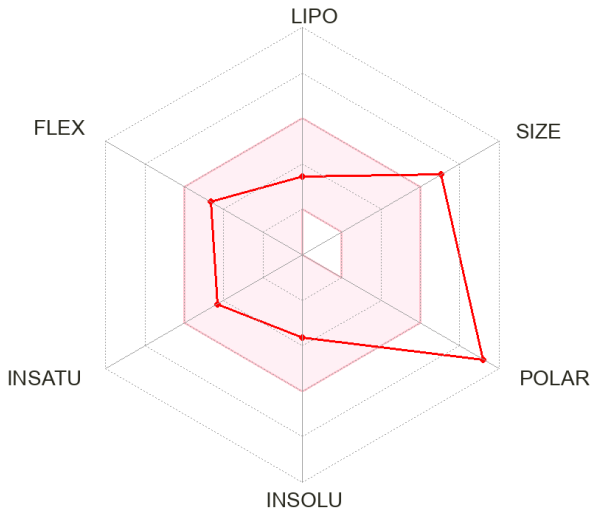
b. Interaction of Afzelin with E7



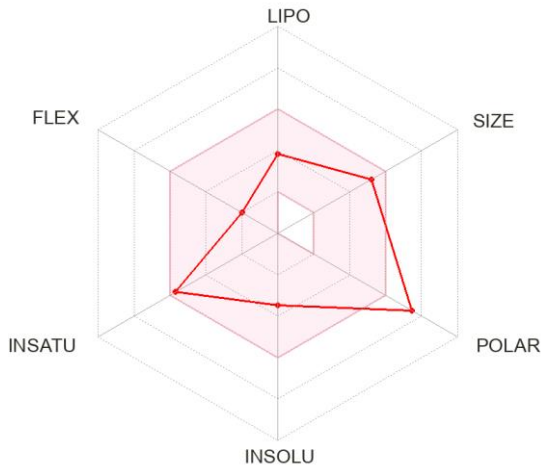
c. Baicalein (Standard) interaction with E7

ADMET Analysis

Admet analysis" refers to the evaluation of a drug candidate's absorption, distribution, metabolism, excretion, and toxicity properties. ADME stands for absorption, distribution, metabolism, and excretion.



Radar graph of Linarin



Radar graph of Afzelin

Flavanoids	Mass	Hydrogen bond donor	Hydrogen bond acceptor	LogP value	Molar refractivity	No. of violations
Linarin	492.5	5	10	-2.76	141.80	1
Afzelin	432.38	6	10	-1.34	106.97	1

Table 2: *Lipinski's Rule of Five Analysis (RO5) Mass<500, H-bond donors<5, Hbond acceptors<10, LogP value<5*

Molinspiration Results

Bioactivity Scores of Phytochemicals retrieved from Molinspiration. GPCR (G-Protein coupled receptor ligand), PI (Protease Inhibitor), EI (Enzyme Inhibitor), NRL (Nuclear Receptor Ligand), KI (Kinase Inhibitor), ICM (Ion Channel Modulator)

Phytochemical	**GPCR**	**PI**	**EI**	**NRL**	**KI**	**ICM**
Linarin	-0.01	-0.02	0.15	0.15	0.09	-0.44
Afzelin	-0.01	-0.05	0.36	0.16	0.05	-0.36

Table 3: *Bioactivity Scores of Phytochemicals retrieved from Molinspiration. GPCR (G-Protein coupled receptor ligand), PI (Protease Inhibitor), EI (Enzyme Inhibitor), NRL (Nuclear Receptor Ligand), KI (Kinase Inhibitor), ICM (Ion Channel Modulator).*

Flavonoids	Activity
Linarin	Positive
Afzelin	Positive
Baicalein **(Standard inhibitor)**	Positive

Table 4: *Machine Learning Result*

Conclusion

The death rate for HPV has significantly increased worldwide during the past 20 years. The overexpression of E6 and E7 oncoprotein is one of the factors that can be associated with it. HPV has one of the highest fatality rates of all malignancies. The crucial targets E6 and E7 are those for which there are numerous medications on the market. Recent research has shown that manufactured drugs used to treat cancer frequently have side effects and toxicity, which has shifted focus to naturally occurring substances that are non-toxic and readily available and can be used as anti-cancer agents, such as flavonoids. The current in-silico investigation provides information on oncoprotein inhibition. Here, we've demonstrated how a strategic and systematic virtual screening method that includes molecular docking, ADMET studies, toxicity studies, pharmacokinetics, and molecular dynamic investigations can assist in reaching a meaningful conclusion.

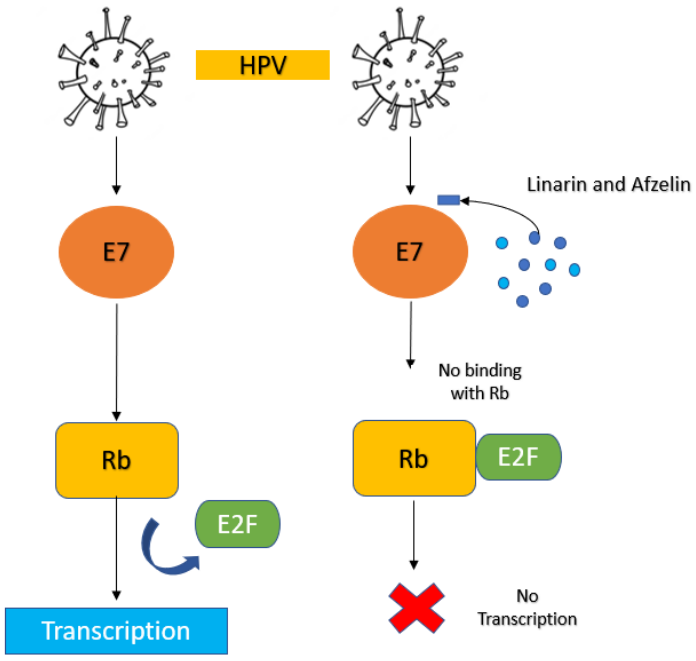


Fig. 6 Linarin and Afzelin blocking the cell cycle by interacting with E7 oncoprotein of HPV

Linarin and Afzelin is a flavonoid compound found abundantly in angiosperms and is known for its numerous health benefits. In particular, Linarin and Afzelin is found in the roots of Valeriana Jatamnsi and leaf of Argyreia nervosa, former is a rhizome herb and later is perennial climbing vine, both are readily available in the market. This in-silico study aimed to explore the potential of Linarin and Afzelin as an inhibitor of the activity of the HPV E7 oncoprotein, a protein involved in HPV pathogenesis. The study utilized bioactive compounds derived from Linarin and Afzelin, found that Linarin exhibited a maximum binding affinity with a ΔG of -10.6 while Afzelin with a binding affinity of ΔG= -9.05 , which is higher than the standard Baicalein with a ΔG of -7.36.

The study also considered additional parameters such as minimal cytotoxicity, maximum bioavailability, and bioactivity to evaluate Linarin and Afzelin as a potential therapeutic drug. Based on these parameters, both is found to be a promising therapeutic drug candidate for inhibiting not only HPV E7 oncoprotein but also other proteins involved in HPV pathogenesis such as HPV E6 oncoprotein.

In summary, this study highlights the potential of natural compounds such as Linarin and Afzelin as an effective and non-toxic alternative to synthetic drugs for the treatment of cancer. The use of in-silico approaches to identify potential drug candidates can significantly reduce the time and cost required for drug development.

References

[1] H. Zur Hausen, "Papillomaviruses and cancer: From basic studies to clinical application," *Nature Reviews Cancer*, vol. 2, no. 5. European Association for Cardio-Thoracic Surgery, pp. 342–350, 2002. doi: 10.1038/nrc798.

[2] A. R. Kreimer *et al.*, "HPV16 semiquantitative viral load and serologic biomarkers in oral and oropharyngeal squamous cell carcinomas," *Int J Cancer*, vol. 115, no. 2, pp. 329–332, Jun. 2005, doi: 10.1002/ijc.20872.

[3] H. De Vuyst, G. M. Clifford, M. C. Nascimento, M. M. Madeleine, and S. Franceschi, "Prevalence and type distribution of human papillomavirus in carcinoma and intraepithelial neoplasia of the vulva, vagina and anus: A meta-analysis," *Int J Cancer*, vol. 124, no. 7, pp. 1626–1636, Apr. 2009, doi: 10.1002/ijc.24116.

[4] S. N. Boyer, D. E. Wazer, and V. Band, "E7 protein of human papilloma virus-16 induces degradation of retinoblastoma protein through the ubiquitin-proteasome pathway.," *Cancer Res*, vol. 56, no. 20, pp. 4620–4, Oct. 1996.

[5] M. Thomas, D. Pim, and L. Banks, "The role of the E6-p53 interaction in the molecular pathogenesis of HPV." [Online]. Available: http://www.stockton-press.co.uk/onc

[6] H. Zur Hausen, "Papillomaviruses and cancer: From basic studies to clinical application," *Nature Reviews Cancer*, vol. 2, no. 5. European Association for Cardio-Thoracic Surgery, pp. 342–350, 2002. doi: 10.1038/nrc798.

[7] K. Munger,', C. L. Yee, W. C. Phelps, J. A. Pietenpol, H. L. Moses, and P. M. Howley', "Biochemical and Biological Differences between E7 Oncoproteins of the High-and Low-Risk Human Papillomavirus Types Are Determined by Amino-Terminal Sequences," 1991. [Online]. Available: https://journals.asm.org/journal/jvi

[8] Q. Wang, H. Xu, and X. Zhao, "Baicalin inhibits human cervical cancer cells by suppressing protein kinase c/signal transducer and activator of transcription (PKC/STAT3) signaling pathway," *Medical Science Monitor*, vol. 24, pp. 1955–1961, Apr. 2018, doi: 10.12659/MSM.909640.

[9] J. J. Cherry *et al.*, "Structure based identification and characterization of flavonoids that disrupt human papillomavirus-16 E6 function," *PLoS One*, vol. 8, no. 12, Dec. 2013, doi: 10.1371/journal.pone.0084506.

[10] R. Zhang, X. Li, X. Zhang, H. Qin, and W. Xiao, "Machine learning approaches for elucidating the biological effects of natural products," *Natural Product Reports*, vol. 38, no. 2. Royal Society of Chemistry, pp. 346–361, Feb. 01, 2021. doi: 10.1039/d0np00043d.

[11] R. Dikshit *et al.*, "Cancer mortality in India: a nationally representative survey," *The Lancet*, vol. 379, no. 9828, pp. 1807–1816, May 2012, doi: 10.1016/S0140-6736(12)60358-4.

[12] F. X. Bosch, A. Lorincz, N. Munoz, C. J. L. M. Meijer, and K. V Shah, “The causal relation between human papillomavirus and cervical cancer,” *J Clin Pathol*, vol. 55, no. 4, pp. 244–265, Apr. 2002, doi: 10.1136/jcp.55.4.244.

[13] F. X. Bosch, “Human papillomavirus: science and technologies for the elimination of cervical cancer,” *Expert Opin Pharmacother*, vol. 12, no. 14, pp. 2189–2204, Oct. 2011, doi: 10.1517/14656566.2011.596527.

[14] D. L. Jones and K. Münger, “Interactions of the human papillomavirus E7 protein with cell cycle regulators,” *Semin Cancer Biol*, vol. 7, no. 6, pp. 327–337, Dec. 1996, doi: 10.1006/scbi.1996.0042.

[15] M. M. Lipinski and T. Jacks, “The retinoblastoma gene family in differentiation and development,” *Oncogene*, vol. 18, no. 55, pp. 7873–7882, Dec. 1999, doi: 10.1038/sj.onc.1203244.

[16] S. Cheng, D. C. Schmidt-Grimminger, T. Murant, T. R. Broker, and L. T. Chow, “Differentiation-dependent up-regulation of the human papillomavirus E7 gene reactivates cellular DNA replication in suprabasal differentiated keratinocytes.,” *Genes Dev*, vol. 9, no. 19, pp. 2335–2349, Oct. 1995, doi: 10.1101/gad.9.19.2335.

[17] A. Pal and R. Kundu, “Human Papillomavirus E6 and E7: The Cervical Cancer Hallmarks and Targets for Therapy,” *Front Microbiol*, vol. 10, Jan. 2020, doi: 10.3389/fmicb.2019.03116.

[18] S. Singh, R. R. Chhipa, M. V. Vijayakumar, and M. K. Bhat, “DNA damaging drugs-induced down-regulation of Bcl-2 is essential for induction of apoptosis in high-risk HPV-positive HEp-2 and KB cells,” *Cancer Lett*, vol. 236, no. 2, pp. 213–221, May 2006, doi: 10.1016/j.canlet.2005.05.024.

[19] R. Xiao *et al.*, “Dual Antagonist of cIAP/XIAP ASTX660 Sensitizes HPV− and HPV+ Head and Neck Cancers to TNFα, TRAIL, and Radiation Therapy,” *Clinical Cancer Research*, vol. 25, no. 21, pp. 6463–6474, Nov. 2019, doi: 10.1158/1078-0432.CCR-18-3802.

[20] R. Li, C. Guo, Y. Li, Z. Qin, and W. Huang, “Therapeutic targets and signaling mechanisms of vitamin C activity against sepsis: a bioinformatics study,” *Brief Bioinform*, vol. 22, no. 3, May 2021, doi: 10.1093/bib/bbaa079.

[21] R. Kotadiya, A. Johnson, and J. J. Georrge, “Natural Products as Inhibitors of E6 and E7 Proteins of Human Papillomavirus (HPV): An In Silico Approach,” *SSRN Electronic Journal*, 2020, doi: 10.2139/ssrn.3574478.

[22] K. Mohanraj *et al.*, “IMPPAT: A curated database of Indian Medicinal Plants, Phytochemistry and Therapeutics,” *Sci Rep*, vol. 8, no. 1, Dec. 2018, doi: 10.1038/s41598-018-22631-z.

[23] C. A. Lipinski, “Lead- and drug-like compounds: the rule-of-five revolution,” *Drug Discov Today Technol*, vol. 1, no. 4, pp. 337–341, Dec. 2004, doi: 10.1016/j.ddtec.2004.11.007.

[24] D. Mendez *et al.*, “ChEMBL: Towards direct deposition of bioassay data,” *Nucleic Acids Res*, vol. 47, no. D1, pp. D930–D940, Jan. 2019, doi: 10.1093/nar/gky1075.

[25] A. Gaulton *et al.*, “ChEMBL: A large-scale bioactivity database for drug discovery,” *Nucleic Acids Res*, vol. 40, no. D1, Jan. 2012, doi: 10.1093/nar/gkr777.

[26] S. Krishna, S. Kumar, D. K. Singh, A. D. Lakra, D. Banerjee, and M. I. Siddiqi, “Multiple Machine Learning Based-Chemoinformatics Models for Identification of Histone Acetyl Transferase Inhibitors,” *Mol Inform*, vol. 37, no. 8, Aug. 2018, doi: 10.1002/minf.201700150.

WORD COUNT

8818 Words

CHARACTER COUNT

50920 Characters

PAGE COUNT

38 Pages

FILE SIZE

926.0KB

SUBMISSION DATE

May 29, 2023 4:18 PM GMT+5:30

REPORT DATE

May 29, 2023 4:18 PM GMT+5:30

18% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 13% Internet database
- 11% Publications database
- Crossref database
- Crossref Posted Content database
- 12% Submitted Works database

Excluded from Similarity Report

- Bibliographic material