

“Computational Genomics and Computer Aided Drug Design for Organics Compounds against Mutant p53 Cancer Gene”

Dissertation

Submitted to

Jaipur National University in partial fulfilment of the requirement for the

Degree of

BACHELOR OF TECHNOLOGY

IN

BIOTECHNOLOGY



Submitted By:

Kanishk Yadav

School of Engineering and Technology

Jaipur National University, Jaipur

2022

CERTIFICATE – I

This is to certify that Mr. Kanishk Yadav, a student of B. Tech Biotechnology has completed her dissertation entitled “Computational Genomics and Computer Aided Drug Design for Organics Compounds against Mutant p53 Cancer Gene”, for submission to the School of Engineering and Technology, Jaipur National University, Jaipur.

To the best of my knowledge the work has not been submitted in part or in the full to any other University or Institute for the award of any degree. The assistance and help received during the work has been duly acknowledged. Any discrepancies or irregularities related to project idea, data, data representation, writing, acknowledgements, citations or any sort of plagiarism in the thesis will be the sole responsibility of the student.

**Dr. Manish Soni
(Advisor)**

**Mr. Kanishk Yadav
(Student)**

Date:

DECLARATION

I hereby declare that the dissertation entitled “Computational Genomics and Computer Aided Drug Design for Organics Compounds against Mutant p53 Cancer Gene” submitted by me for the award of the degree of Bachelor of Technology in Biotechnology to Jaipur National University, Jaipur is a bonafide work carried out under the guidance and supervision of Dr. Vinod Kumar Gupta (Rapture Biotech International Ltd.) and Dr. Manish Soni (Jaipur National University).

I declare that the work for the dissertation has not been submitted, for the award of any other degree or diploma and the assistance received from other sources or people has been duly acknowledged. I further declare and also aware that in case of any discrepancies or irregularities related to project idea, data, data representation, writing, acknowledgements, citations or any sort of plagiarism found in the thesis will be my sole responsibility.

Kanishk Yadav
B.Tech Biotechnology VIII Semester
2018-2022

CERTIFICATE – II

This is to certify that the dissertation entitled “Computational Genomics and Computer Aided Drug Design for Organics Compounds against Mutant p53 Cancer Gene”, submitted to partial fulfilment of the requirements for the degree of B.Tech. Biotechnology, School of Engineering and Technology, Jaipur National University, Jaipur, is a record of bonafide research/ study carried out by Mr. Kanishk Yadav.

The dissertation has been approved, after presentation and viva-voce examination by internal and external examiners.

Internal	External	Dated:
	(Examiners)	

Dr. Kapilesh Jadhav
Coordinator/Head
B.Tech. Biotechnology Programme



CERTIFICATE

This is to certify that Mr. Kanishk Yadav, S/o Mr. Dashrath Yadav, student of B.Tech Biotechnology, Enrolment No.-JNU-jpr-2018/00072, from Jaipur National University, Jaipur, Rajasthan, has successfully completed his Project/Dissertation training Program of 6 Months, from 1st January 2022 to 30th June 2022, under the guidance of Dr. Vinod Kumar Gupta in our Biotechnology/R&D Division at **NOIDA Campus** with '**A+**' Grade on the topic:

"Computational Genomics and Computer aided drug design for Organic compounds against mutant p53 cancer gene"

We wish for his bright future.

Certificate number:RPTN-2206/121

For RAPTURE BIOTECH

Chief Executive Officer

Mr. Mayank Raj Bhardwaj
Chief Executive Officer
Rapture Biotech, India

Acknowledgement

My internship was highly worthwhile and I got to learn so much about bioinformatics and thesis preparation. So, I would like to thank Rapture Biotech International Ltd. and Mr. Mayank Bhardwaj (Director) along with my supervisor Dr. Vinod Kumar Gupta (Senior Scientist) and my guide Ms. Meenu Bohra (Senior Research Associate).

I would like to thank my University for making a legible and potent student in the field of biotechnology and helping me achieve greater heights. More importantly, I would like to thank Dr. Sandeep Bhakshi (Chancellor), Prof. H.N.Verma (Pro-Chancellor), Prof. R.L.Raina (Vice-Chancellor), Prof. Anjana Sharma (Rector), Prof. Rajeev Mathur (Director), Prof. Purnima Nag (Sr. Addl. Director) and Dr. Kapilesh Jadhav (Co-ordinator), Dr. Manish Soni (Assisstant Prof.) and Dr. Sunita Rao (Assisstant Prof.).

Table of Contents

• Objective	Page(s): 15
• Introduction	Page(s): 16 to 20
• Review of Literature	Page(s): 21 to 25
• Research Methodology	Page(s): 26 to 28
• Materials and Methods	Page(s): 29 to 54
• Results	Page(s): 55 to 115
• Discussion	Page(s): 116
• Conclusion	Page(s): 117
• Significance	Page(s): 118
• References	Page(s): 119 to 123

Abbreviations

- CADD: Computer Aided Drug Design
- TP52: Tumour Protein p53 gene
- WT_{p53}: Wild Type p53
- DNA: Deoxyribonucleic acid
- RNA: Ribonucleic acid
- BLAST: Basic Local Alignment Search Tool
- ADME: Absorption, Distribution, Metabolism, and Excretion
- EMBL: European Molecular Biology Laboratory
- NCBI: National Center for Biotechnology Information
- MDM2: Mouse double minute 2
- BBB: Blood Brain Barrier
- GI Absorption: Gastrointestinal Absorption
- HSP: High-scoring Segment Pairs
- SMILES: Simplified Molecular Input Line Entry System
- GPCR: G-Protein Coupled Receptor
- KEGG: Kyoto Encyclopedia of Genes and Genomes

List of Figures

Fig .1, Cause of cancer as depicted in the flowchart above.....	21
Fig. 2, Cancer cell cycle.....	22
Fig. 3, The above figure shows the areas of the Computer-Aided Drug Design.....	24
Fig. 4, NCBI homepage screenshot.....	29
Fig. 5, BLAST webpage containing different types of BLAST operations that can be performed.....	31
Fig. 6, Clustal Omega homepage screenshot.....	33
Fig. 7, The above screenshot represents the GeneCards homepage.....	34
Fig. 8, ORF Finder homepage screenshot.....	35
Fig. 9, Example of a six-frame translation. The nucleotide sequence is shown in the middle with forward translations above and reverse translations below. Two possible open reading frames with the sequences are highlighted.....	36
Fig. 10, Protein Data Bank homepage screenshot.....	37
Fig. 11, The PDB database, as well as its holdings list, is updated weekly (UTC+0 Wednesday). As of 1 April 2020, the PDB had the above-mentioned items in the above table.).....	38
Fig. 12, Uniprot homepage screenshot.....	39

Fig. 13, UniProt databases include the UniProt Knowledgebase (UniProtKB), UniProt Reference Clusters (UniRef), and UniProt Archive (UniParc).....	40
Fig. 14, The above screenshots represent the ProtParam homepage.....	41
Fig. 15, The above screenshot is of the STRING Database homepage. Search option is used to proceed further and make any changes.....	42
Fig. 16, The above figure is the from the KEGG Pathway webpage.....	43
Fig. 17, DrugBank homepage screenshot.....	45
Fig. 18, The above screenshot represents the PubChem homepage.....	46
Fig. 19, The above screenshot represents the Swiss Target Prediction.....	47
Fig. 20, The above figure depicts SWISS ADME webpage. The two boxes provided correspond to the areas where we can either enter the SMILE of the ligand and ‘Run!’ it or draw the entire structure of our need which will, in turn, generate a SMILE and then again ‘Run!’ it.....	48
Fig. 21, The above screenshot represents the Pre ADME homepage. Different, options can be selected on the top right-hand corner of the homepage.....	51
Fig. 22, The above screenshot is of the Molinspiration tool homepage. Different options as per the needs of the users can be accessed from the sidebar as per the user’s needs.....	52
Fig. 23, The above screenshot represents the PyRx virtual screening tool which can be download and installed for PC usage.....	53
Fig. 24, The above figure represents screenshot of the Biovia Discovery Studio which is a downloadable software and can be installed on the PC.....	54

Fig. 25, The above screen shot represents the search result of our target on NCBI.....	55
Fig. 26, The result page of the TP53 tumour protein [Homo sapiens (humans)]..	56
Fig. 27, The above screenshot represents the search result of the blast search for mutant p53 protein.....	57
Fig. 28, The above figure represents the amino acid sequence alignment.....	58
Fig. 29, The above screenshot represents the distribution of the top 100 Blast Hits on 100 subject sequences.....	59
Fig. 30, The above screenshot represents the conserved domains of the the query sequence.....	60
Fig. 31, Three different sequences for mutant p53 protein were taken and the clustal omega was used in order to check the multiple sequence alignment. ‘Show Colors’ option was selected.....	61
Fig. 32, The above screenshot is the search result of the ‘mutant p53’ on the GeneCard website.....	62
Fig. 33, The position of TP53 gene on the Chr 17.....	63
Fig. 34, Above if the screenshot of the ‘Function’ section of from the ‘Jump to section’ for TP53.....	64
Fig. 35, The above the screenshot shows the ORF’s in the mutant p53 gene nucleotide sequence.....	65
Fig. 36, The above screenshot represents the result of the Uniprot search for the mutant p53.....	67

Fig. 37, Keyword section of Uniprot result.....	68
Fig. 38, The ‘Family and Domain’ information for the uniprot has been show in the above screenshot.....	68
Fig. 39, The string database search result for p53 protein is being is visible. As we are working on Homo Sapien in terms of cancer by mutant p53 gene. We have selected the ‘Homo Sapiens’ in the ‘organisms’ section.....	71
Fig. 40, The above figure represents the result of the search for ‘p53’ on the string database. Description of various component in the figure is also available in the the figure itself.....	72
Fig. 41, This is one of the important part of the string database searches. Depending on our input (“Your Input:"), the string database has provided us with “Predicted Functional Partners” with scores.....	73
Fig. 42. It represents the search result for the p53 in the KEGG Pathway search. Here we have only shown only ‘map04115’ as it is the out preferred result for studying the p53 tumor pathway. However, the result had 66 search results for the same.....	75
Fig. 43. The above figure refers to the ‘map04115’ which is the p53 signaling pathway.....	75
Fig. 44, The AZD-1775 have the highest binding affinity among all the other ligands (drugs) as well as Ganetespib is close by with negligible difference.....	110
Fig. 45, The protein above is the prepared 2J1X protien structure in Biovia Discovery Studio.....	111
Fig. 46, In PyRx, this is the 3D structure of docked 2J1X protein. Here the thicker structure at the right hand corner of the figure represents the ligand (drug) AZD-	

1775 which is in the docked position to the 2J1X protein and is showing the best binding affinity.....	112
Fig. 47, Cleaned cellular tumour protein p53 in a docked confirmation with ligand AZD-1775.....	113
Fig. 48, Docked 2J1X protein's site. Here interactions sites with ligand are again visible like the previous figure.....	114
Fig. 49, This is a 2D – Strucutre of the same interaction as visible in the previous figures.....	115

List of Tables

Table 1, PDB Database search results for our selected protein 2J1X.....	67
Table 2, ProtParam search results for Tumour p53 cancer's amino acid sequence for Homo sapiens (Humans).....	70
Table 3, The above table contains drugs that had been selected from drug bank and various research papers for the purpose of our studies.....	78
Table 4, The above table depicts the 2D Structures and 3D Confirmations of the different drugs selected for the studies.....	80
Table 5, Different chemical and physical properties of drugs that are being used for the purpose of our studies have listed in the above table. These are some of the essential properties of the drugs that have been selected.....	83
Table 6, Docking Results of 2J1X after preparation with are chosen ligands (drugs).....	110

Objective

- 1) Comparison and variability of mutant p53 cancer gene sequence with the reference genome.
- 2) To analyze the local and global similarity between sequences by comparing nucleotide or protein sequence to sequence database and calculate the statistical significance of matches.
- 3) Clustal Omega multiple sequence alignment program that uses a seed guide tree and Hidden Markov Model (HMM) profiling technique.
- 4) The computational alignment method to calculate all possible parameters using the biological database.
- 5) To analyze the structural properties, domain, and function of the mutant p53 cancer gene.
- 6) CADD approach in molecular docking to find optimized confirmation between ligand and receptor that results in minimum energy that binds to a particular protein of interest.

Introduction

Cancer

Cancer is a term used to describe abnormal cell proliferation caused by a range of disorders. This aberrant cell proliferation has the potential to infiltrate or spread to other sections of the body. These are in contrast to benign tumors, which remain stationary and do not invade other areas of the body, and malignant tumors, which have a proclivity to spread to other sections of the body. Humans are affected by about 100 different types of cancer.

The bulk of cancers (90–95%) are caused by genetic changes caused by environmental and lifestyle factors. The remaining 5–10% is related to inherited genetics. Environmental variables include not only pollution, but also lifestyle, economic, and behavioural aspects that are not hereditary. Tobacco usage (25–30%), food and obesity (30–35%), infections (15–20%), radiation (both ionizing and non-ionizing, up to 10%), physical inactivity, and pollution are all common environmental variables that lead to cancer death. Psychological stress does not appear to be a risk factor for cancer development, while it may impair results in people who already have cancer.

Types of cancer:-

Cancer has been classified into several types by doctors based on its origin. Cancer is classified into four types:

Carcinomas: These cancers start in the skin or the tissue that covers the surface of internal organs and glands. Carcinomas are typically solid tumors. They are the most prevalent form of cancer. Prostate cancer, breast cancer, lung cancer, and colorectal cancer are all examples of carcinomas.

Sarcomas: These are cancerous tumors that begin in the tissues that support and connect the body. Sarcomas can occur in the fat, muscles, nerves, tendons, joints, blood vessels, lymph vessels, cartilage, or bone.

Leukemia: It is a type of blood cancer. Leukemia develops when healthy blood cells undergo uncontrollable changes and grow uncontrollably. Acute lymphocytic leukemia, chronic lymphocytic leukemia, acute myeloid leukemia, and chronic myeloid leukemia are the four main types of leukemia.

Lymphomas: These are cancers that start in the lymphatic system. The lymphatic system is a network of vessels and glands that aid in infection resistance. Lymphomas are classified into two types: non-Hodgkin lymphoma and Hodgkin lymphoma.

Mutation:

A mutation is a change in the sequence of DNA. Mutations can occur as a result of DNA copying errors during cell division, ionizing radiation exposure (such as X-rays, γ -rays, etc.), mutagen exposure, or virus infection. Somatic mutations occur in body cells and are not passed on to offspring, but germline mutations occur in eggs and sperm and can be inherited by the offspring [Cohen S. et al 2009].

Genes direct the production of proteins in your cells, which influence how they function. Proteins provide specialized tasks and serve as cell messengers. Each gene must contain the proper instructions for producing the protein. This enables the protein to carry out its intended job in the cell. When one or more genes in a cell mutate, all cancers begin. A mutation is a type of modification. It results in the production of a faulty protein. It could also stop the synthesis of a protein.

An aberrant protein differs from a normal protein in that it offers different information. This can promote uncontrollable cell proliferation, leading to cancer.

There are two kinds of genetic mutations:-

- 1) Somatic mutations: Somatic/Acquired mutations are the most common cause of cancer. They are caused by damage to genes in a specific cell during a person's existence. This might be a breast or colon cell, for example, which then divides numerous times and becomes a tumor. A tumor is a malformed mass. Cancer that develops as a result of acquired mutations is referred to as sporadic cancer. Acquired mutations do not exist in every cell of the body and are not passed down from parent to child. Tobacco, ultraviolet (UV) light, viruses, and age are all factors that contribute to these changes.
- 2) Germline mutations: These are rarer. A germline mutation arises in either a sperm or an egg cell. It is passed straight from a parent to a kid during conception. The mutation from the first sperm or egg cell gets copied into every cell in the body as the embryo develops into a kid. Because the mutation affects reproductive cells, it has the potential to be passed down from generation to generation. Inherited cancer is cancer caused by germline mutations. It accounts for around 5% to 20% of all cancers.

p53 gene:

p53 is a tumor-suppressor protein that is altered in the majority of cancers. p53 induces a number of responses, including cell-cycle arrest and apoptosis. Each of these appears to help with tumor suppression. p53 responds to both acute stresses, such as genotoxic stress and oncogene activation and constitutive stress, such as

hypoxia or starvation stress. Each of these factors can contribute to the development of a tumor. [Vousden K. et al 2007]

In addition to tumor suppression, p53 may play a role in normal development, and p53 deficiency can be deleterious at this time. Activating p53 can have negative repercussions, such as contributing to disorders like neurodegenerative syndromes and the aging process of an organism. However, it is now obvious that p53 may play a far larger function in the development, life length, and general fitness of an organism. [Vousden K. et al 2007]

p53 is the most frequently altered tumor suppressor gene in human cancer. The bulk of p53 mutations are missense mutations that cause full-length p53 mutant proteins to be expressed. [Vousden K. et al 2007]

Organic compound against mutant p53 gene:

The loss of p53 tumor-suppressor function caused by TP53 mutation/deletion or suppression of p53 permits cells harmed by stress to proliferate. This unregulated growth can result in the formation of tumors. [Nicole Zache et al 2008]

Increasing data suggests that mutant p53 stability in tumors is critical for its oncogenic activities, whereas mutant p53 depletion reduces the malignant features of cancer cells. As a result, mutant p53 is an appealing druggable target for cancer treatment. Small-molecule drugs that specifically target mutant p53 have been developed using a variety of techniques. Compounds that restore mutant p53's wild-type conformation and transcriptional activity, cause mutant p53 depletion, block the downstream pathways of oncogenic mutant p53, and produce synthetic lethality to mutant p53 are examples of these. [Nicole Zache et al 2008]

Computational Genomics and Computer-Aided Drug Design (CADD):

Computational genomics (sometimes spelled computer genetics) is the use of computational and statistical analysis to understand biology from genome sequences and associated data, which includes both DNA and RNA sequences as well as other "post-genomic" data (i.e., experimental data obtained with technologies that require the genome sequence, such as genomic DNA microarrays). [Koonin EV et al 2001]

CADD is a cutting-edge computational approach used in drug development to identify and develop new leads. Computational chemistry, molecular modeling, molecular design, and rational drug design are all aspects of computer-aided drug design. CADD is being utilized to improve the quality of selected leads [Koonin EV et al 2001].

Computer-aided drug design can be employed at any of the following stages of drug discovery [Singh J. et al 2003]:

- Virtual screening is used to identify hits (structure- or ligand-based design)
- Optimization of affinity and selectivity from hit to lead (structure-based design, QSAR, etc.)
- Contribute to the improvement of other pharmacological characteristics while retaining affinity

Review of Literature

Cancer

Cancer is a collection of disorders characterized by abnormal cell proliferation and the ability to infiltrate or spread to other regions of the body. These differ from benign tumors, which do not spread. A lump, unusual bleeding, a persistent cough, unexplained weight loss, and a change in bowel motions are all possible indications and symptoms. While these symptoms may signal cancer, they might possibly be caused by something else. Humans are affected by about 100 different forms of cancer.

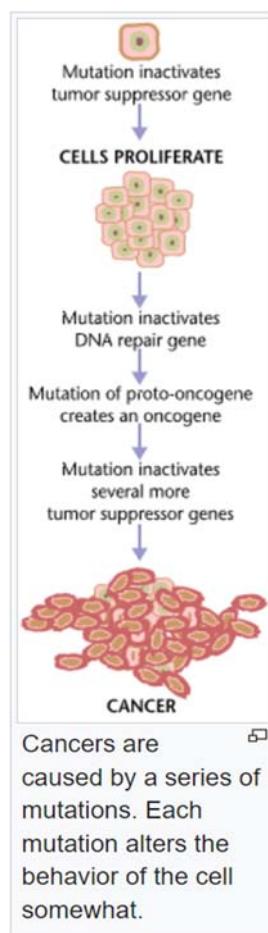


Fig. 1, Cause of cancer as depicted in the flowchart above.

CANCER LIFE CYCLE

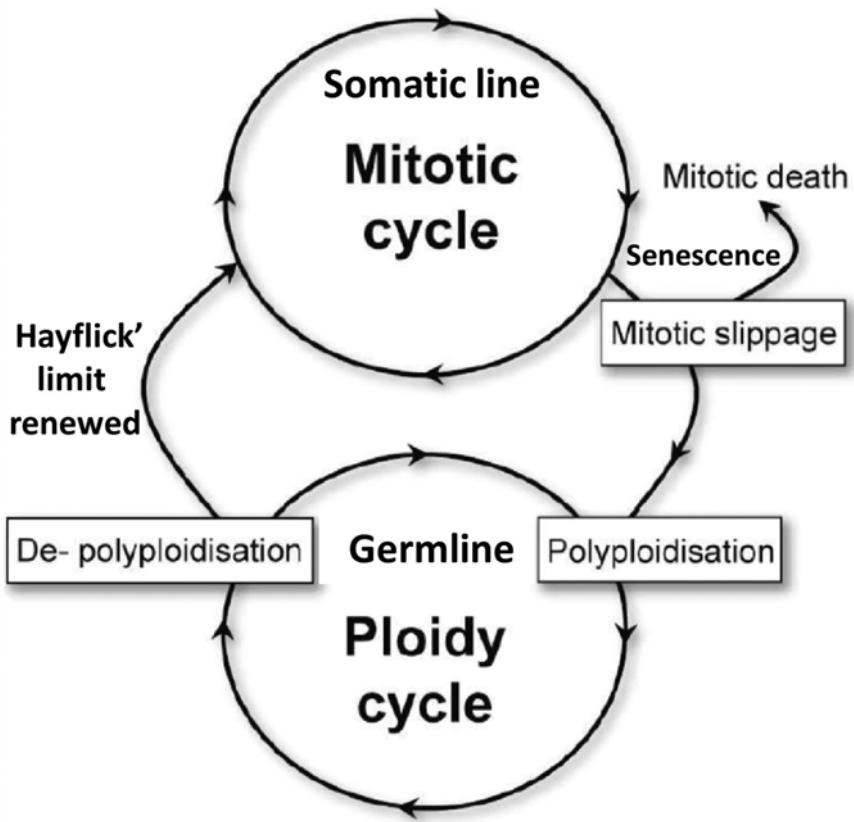


Fig. 2 Cancer cell cycle.

Cell division inside cells is tightly regulated by multiple evolutionarily conserved cell cycle control mechanisms, to ensure the production of two genetically identical cells. Cell cycle checkpoints operate as DNA surveillance mechanisms that prevent the accumulation and propagation of genetic errors during cell division. Checkpoints can delay cell cycle progression or, in response to irreparable DNA damage, induce cell cycle exit or cell death (Fig. 2). Cancer-associated mutations that perturb cell cycle control allow continuous cell division chiefly by compromising the ability of cells to exit the cell cycle [Hartwell et al 1994].

TP53 Gene and Mutations in it:

The TP53 gene codes for the production of a protein known as tumor protein p53 (or p53). This protein functions as a tumor suppressor, which means it controls cell division by preventing cells from growing and dividing (proliferating) too quickly or uncontrollably. The p53 protein is found in the nucleus of cells all over the body, where it connects (binds) to DNA. When a cell's DNA is damaged by agents such as harmful chemicals, radiation, or ultraviolet (UV) rays from sunlight, this protein plays a vital role in deciding whether the DNA is repaired or the damaged cell self-destructs (undergo apoptosis). If the damaged DNA can be repaired, p53 activates other genes to repair it. If the DNA cannot be repaired, this protein inhibits the cell from dividing and triggers apoptosis. p53 aids in the prevention of tumor formation by preventing cells with mutant or damaged DNA from proliferating. Because p53 is required for DNA repair and cell division, it has been dubbed the "guardian of the genome". [Balint E. et al 2001]

Cancer cells may grow and spread in the body as a result of mutations (changes) in the p53 gene. Cancer-causing chemicals in the environment (carcinogens) such as tobacco smoke, UV radiation, and the chemical aristolochic acid can damage (mutate) the TP53 gene (with bladder cancer). However, the toxin that causes the mutation is frequently unknown. Unlike most tumor suppressor genes, such as RB, APC, or BRCA1, which are inactivated by deletions or truncating mutations as cancer progresses, the TP53 gene in human cancers is frequently discovered to have missense mutations, in which a single nucleotide is replaced by another. The vast majority of mutations cause p53 to lose its capacity to bind DNA sequence-specifically and initiate transcription. [Hainaut P. et al 2000]

The accumulation of mutant p53 has been shown to have many consequences, including changes in DNA binding and transcriptional regulation, such as the transcription of the oncogene MYC31 inducing increased proliferation, or transcription of the gene encoding multidrug resistance, which might increase resistance to a range of cancer chemotherapeutic drugs. Computer Aided Drug Design is thus required in order to encounter such problems.[Atema A. et al 2002]

Computer-Aided Drug Design:

The use of computer techniques to locate, develop, and analyse medications and associated biological active substances is known as computer-aided drug design (CADD). CADD techniques help guide and speed up drug discovery while also speeding up the early stages of chemical development. [Bajorath J. et al 2015]

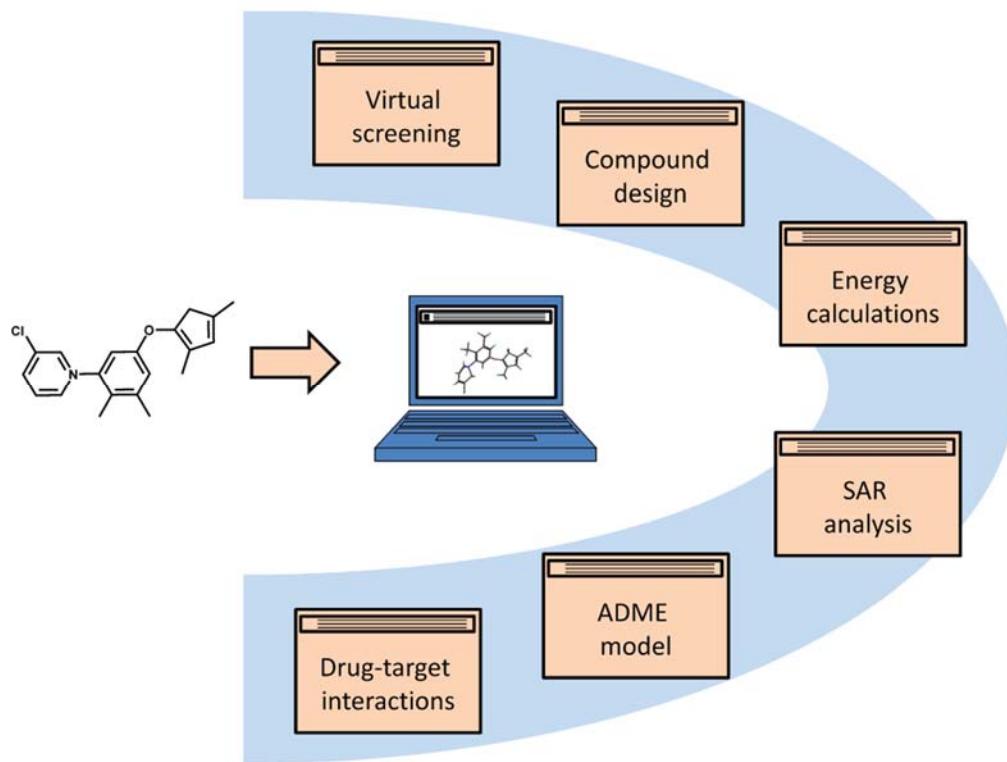


Fig. 3, The above figure shows the areas of the Computer-Aided Drug Design.

All the possible areas in the field of computer-aided drug design has been listed in the Fig. 3. There are more research areas in CADD in development that will take some time.

CADD approaches have improved compound searching based on similarity, target identification and structure prediction, binding site/cavity prediction and validation, understanding the protein–ligand interaction, screening a large number of compounds, understanding the dynamics of protein–ligand binding under

physiological conditions, predicting ADMET properties, estimating biological activity using QSAR, and guiding the required changes. [Anshul T. et al 2022]

Scheme of Computer-Aided Drug Design:

In general, in silico approaches with utility for drug discovery can roughly be divided into three major categories. These include the following:

- First, the design, implementation, and maintenance of computational infrastructures to process, organize, analyze, and store rapidly growing amounts of drug discovery data (e.g. compound library, biological screening, pharmacological, clinical, and literature data);
- Second, methods to help identify, characterize, and prioritize biological targets and establish links between target engagement, biology, and disease (these approaches essentially fall into the domain of bioinformatics); and
- Third, methods to help make better compounds and generate drug candidates. [Bajorath J. et al 2014]

Computer-aided drug design approach can be used to identify previously unknown molecules for targeting p53 protein with anti-cancer activity and thus pave the way for the study of a therapeutic solution. [Rui P S Patricio et al 2021]

Research Methodology

In Silico Studies and Characterization of Organic Compounds

Organic Compounds are being collected and analysed as part of the current research. Various computational tools, including as PubChem, DrugBank, SWISS ADME, PreADME, and others, will be used to investigate physical, structural, and chemical features. These are bioactive compounds having drug-like characteristics that have been manually curated. It combines chemical, bioactivity, and genetic data to aid in the translation of biological data into new medications that work.

PubChem and DrugBank Library

PubChem: The world's greatest collection of freely accessible chemical knowledge is available online at PubChem. Chemicals can be found using their names, molecular formulas, structures, and other identifiers. Find information about chemical and physical qualities, biological activities, safety and toxicity, patents, and literature citations, among other things. **Drugbank:** In silico drug discovery, drug 'rejuvenation,' drug docking or screening, drug metabolism prediction, drug target prediction, and general pharmaceutical education have all used DrugBank.

Target Prediction for Active Organic Compounds with Therapeutics Relevance

Computational tools such as Swiss Target Prediction (<http://www.swisstargetprediction.ch/>) will be used to predict targets that could play a role in anti-cancer therapeutics. Biological processes related to cancer mechanisms will be further investigated using tools such as KEGG pathway databases, STRING database, and (<https://www.genome.jp/kegg/pathway>), and (<https://string-db.org/>), respectively.

CADD Approach: Molecular Interaction among Ligands and Target.

Computer-aided drug designing tools and strategy will be used in present study to identify the targets for given organic compounds. **Target Identification and Validations:** Based on the computational studies, protein structure of cellular tumour protein p53 called ‘2J1X’ will be selected for target identification and validation. It has been shown in our selected organic compounds/ligands table that all the drugs are in experimental and investigational stages. Three-dimensional structures of these target proteins will be retrieved from protein data bank (**RCSB PDB Databases**). Organic compound’s structures will be either derived from databases such as **NCBI Pubchem or DrugBank**.

Molecular Docking and Virtual Screening

PyRx and Biovia Discovery Studio Tools will be used to perform screening of drug library derived from seven organic compounds/ligands with anti-cancer activity. These tools provide a user-friendly docking platform for flexible ligand docking keeping target protein structure rigid. After importing and preparing protein structure, docking wizard provides a selection of docking option. Docking tab allows selection of a specific scoring function like binding affinity and root mean square values.

Computer aided prediction of ADME, activity and Toxicity Studies.

The pharmaco-kinetics properties of given drugs will be studied on SWISS ADME. This website allows us to compute physicochemical descriptors as well as to predict ADME parameters, pharmacokinetic properties, drug-like nature and medicinal chemistry friendliness of one or multiple small molecules to support drug discovery. Toxicity studies and prediction will be performed on PreADME tools.

Outline of the work

- *In Silico* Structural studies and characterization of 2J1X protein.
- 2D and 3D structure prediction and analysis of 2J1X protein.
- Protein-Protein intercation and Molecular function mediated by 2J1X protein.
- Ligand identification and optimization.
- Target prediction for active organic compounds/ligands/drugs having therapeutics relevance.
- Molecular docking and interaction studies among target (2J1X) and ligands.
- CADD approach to understand the molecular interaction among ligands and target with refrence to cancer.
- Computer aided prediction of ADME, activity and toxicity studies.
- Prediction of mode of action and effects on molecular pathways and biological process for therapeutic significance
- Cleaning and Docking of the ligands and target to find out the best ligand for suppression of mutant p53's oncogenic activity

Materials and Methods

NCBI

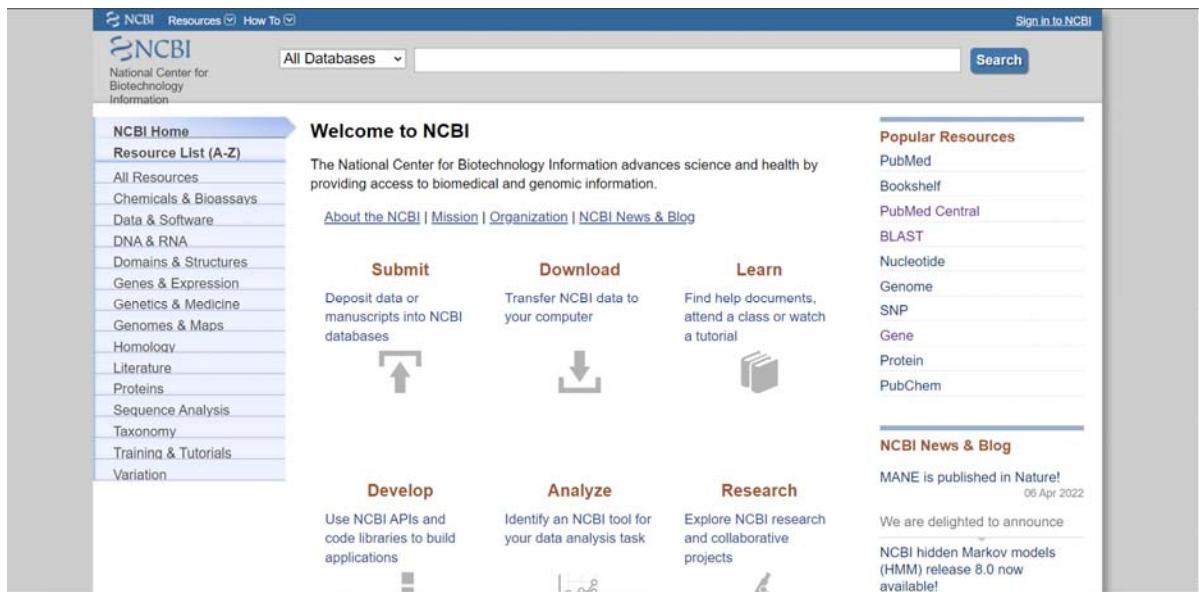


Fig. 4, NCBI homepage screenshot

The National Center for Biotechnology Information (NCBI) is a component of the National Institutes of Health's National Library of Medicine (NLM) (NIH). The NCBI holds a number of databases related to biotechnology and biomedicine, and it is a valuable resource for bioinformatics tools and services. GenBank, a database for DNA sequences, and PubMed, a bibliographic database for biomedical literature, are two major databases. Among the other datasets is the NCBI Epigenomics database. All of these databases are accessible online using the Entrez search engine (Fig. 4).

‘Popular Resources’ as listed on the NCBI website are:-

- **PubMed:** PubMed is a free search engine accessing primarily the MEDLINE database of references and abstracts on life sciences and biomedical topics.
- **Bookshelf:** The books division of the NLM Literature Archive (LitArch) at the National Center for Biotechnology Information (NCBI), is an online

searchable collection of books, reports, databases, and other scholarly literature in biology, medicine, and the life sciences.

- PubMed Central: PubMed Central is a free digital repository that houses open access full-text scholarly papers from biomedical and life sciences publications.
- BLAST: It is a bioinformatics technique and tool for comparing primary biological sequence information, such as amino acid sequences of proteins or nucleotides of DNA and/or RNA sequences.
- Nucleotide: The Nucleotide database is a compilation of sequences from several databases such as GenBank, RefSeq, TPA, and PDB.
- Genome: This site organises genomic information such as sequences, maps, chromosomes, assemblies, and annotations.
- SNP: The Single Nucleotide Polymorphism Database is a free public database for genetic diversity within and across species that was created.
- Gene: Gene combines information from a broad variety of animals. Nomenclature, Reference Sequences (RefSeqs), maps, pathways, variants, phenotypes, and linkages to genome-, phenotype-, and locus-specific resources from around the world may be included in a record.
- Protein: The Protein database contains translations from annotated coding sections in GenBank, RefSeq, and TPA, as well as records from SwissProt, PIR, PRF, and PDB. Protein sequences are key determinants of biological structure and function.
- PubChem: PubChem is a database of chemical compounds and their biological test activities.

BLAST

The screenshot shows the BLAST homepage. At the top, there's a banner with the NIH logo and the text "An official website of the United States government Here's how you know". Below the banner, the "National Library of Medicine" and "National Center for Biotechnology Information" are mentioned. A "Log in" button is also present. The main navigation bar includes "BLAST®", "Home", "Recent Results", "Saved Strategies", and "Help". A "NEWS" section highlights "BLAST+ 2.13.0 is here!" with the message: "Starting with this release, we are including the blastn_vdb and tblastn_vdb executables in the BLAST+ distribution." It also shows the date "Thu. 17 Mar 2022 12:00:00 EST" and a link to "More BLAST news...". Below this, there's a section titled "Web BLAST" featuring four options: "Nucleotide BLAST" (nucleotide & nucleotide), "blastx" (translated nucleotide > protein), "tblastn" (protein > translated nucleotide), and "Protein BLAST" (protein > protein). Each option has a small icon related to its function.

Fig. 5, BLAST webpage containing different types of BLAST operations that can be performed

BLAST (basic local alignment search tool) is a bioinformatics method and program that compares fundamental biological sequence information, such as amino-acid sequences of proteins or nucleotides of DNA and/or RNA sequences. A BLAST search allows a researcher to compare a subject protein or nucleotide sequence (referred to as a query) to a library or database of sequences and find database sequences that resemble the query sequence beyond a specified threshold. For example, following the discovery of a previously unknown gene in the mouse, a scientist will typically perform a BLAST search of the human genome to see if humans carry a similar gene; BLAST will identify sequences in the human genome that resemble the mouse gene based on similarity of sequence (Fig. 5).

BLAST requires a query sequence to search for and a sequence to search against (also known as the target sequence) or a sequence database containing many such sequences in order to execute. BLAST will search the database for subsequences that are comparable to sub-sequences in the query. In most cases, the query

sequence is substantially shorter than the database sequence; for example, the query may be one thousand nucleotides long whereas the database may be several billion nucleotides long.

The primary notion behind BLAST is that statistically important alignments frequently contain High-scoring Segment Pairs (HSP). BLAST uses a heuristic technique that approximates the Smith-Waterman algorithm to look for high-scoring sequence alignments between the query sequence and the existing sequences in the database. The exhaustive Smith-Waterman technique, on the other hand, is too slow for scanning big genomic databases like GenBank. Therefore, the BLAST algorithm uses a heuristic approach that is less accurate than the Smith-Waterman algorithm but over 50 times faster. The speed and relatively good accuracy of BLAST are among the key technical innovations of the BLAST programs. [Oehmen C. et al 2006]

5 Major categories in BLAST are:-

- 1) BLASTN: The "blastn" software is a sensitive general-purpose nucleotide search and alignment tool that may be used to match tRNA or rRNA sequences, as well as mRNA or genomic DNA sequences with a mix of coding and noncoding regions.
- 2) BLASTP: The standard protein-protein BLAST "blastp" algorithm is used to identify a query amino acid sequence as well as detect comparable sequences in protein databases. Blastp, like other BLAST algorithms, is intended to detect local areas of similarity.
- 3) BLASTX: "blastx" is a nucleotide 6-frame translation-protein. This software compares the six-frame conceptual translation products of a nucleotide query sequence (both strands) to a protein sequence database in order to identify a protein-coding gene in a genomic sequence or to determine if the cDNA corresponds to a known protein.
- 4) TBLASTN: The "tblastn" algorithm works by converting database nucleotide sequences to amino acid sequences in all six reading frames,

then aligning the amino acid sequences to the query. Associating proteins with chromosomes or mRNAs is important in many scientific research, hence TBLASTN is commonly utilized.

- 5) TBLASTX: The NCBI's new blast+ package includes the "tblastx" software. tblastx compares the six-frame translations of a nucleotide query sequence to the six-frame translations of a nucleotide sequence database. To receive help with the different options, type `tblastx -help`.

Clustal Omega (Ω)

The screenshot shows the official website for Clustal Omega. The top navigation bar includes links for Home, Webservers, Download, Documentation, Contact, and News. The 'Webservers' button is highlighted. The 'Download Clustal Omega' button is also visible. The 'Source code' link is located just below the 'Download' button. The main content area features a brief introduction to the tool, noting its scalability and superior quality compared to previous versions. It also mentions that it is a command line-only tool and provides a link to a scientific paper describing the algorithms used.

Fig. 6, Clustal Omega homepage screenshot.

Clustal is a collection of frequently used computer tools for multiple sequence alignment in bioinformatics. There have been various versions of Clustal during the course of the algorithm's development, which are mentioned here. Each tool's and algorithm's analysis is also explained in their respective areas. The operating systems available in the sidebar are a combination of software availability and may not be supported for all current versions of the Clustal tools. Clustal Omega supports the most operating systems of any Clustal utility (Fig. 6).

Clustal Omega is the newest member of the Clustal family. It is far more scalable than earlier versions, allowing hundreds of thousands of sequences to be aligned in just a few hours. It will also make advantage of several processors if they are available. Furthermore, alignment quality is higher to prior versions, as determined by a variety of prominent benchmarks.

ClustalΩ (also known as Clustal O and Clustal Omega) is a fast and scalable C and C++ tool used for multiple sequence alignment. To produce these alignments, it employs seeded guide trees and a novel HMM engine that focuses on two profiles. The software requires three or more sequences to calculate multiple sequence alignment; for two sequences, pairwise sequence alignment techniques should be used (EMBOSS, LALIGN). [Chenna R. et al 2003]

GeneCard

Fig. 7, The above screenshot represents the GeneCards homepage.

GeneCards is a human gene database that contains information on all known and expected human genes, including genomic, proteomic, transcriptomic, genetic, and functional information. The Crown Human Genome Center at the Weizmann Institute of Science is in charge of developing and maintaining it (Fig. 7).

Open Reading Frame (ORF) Finder

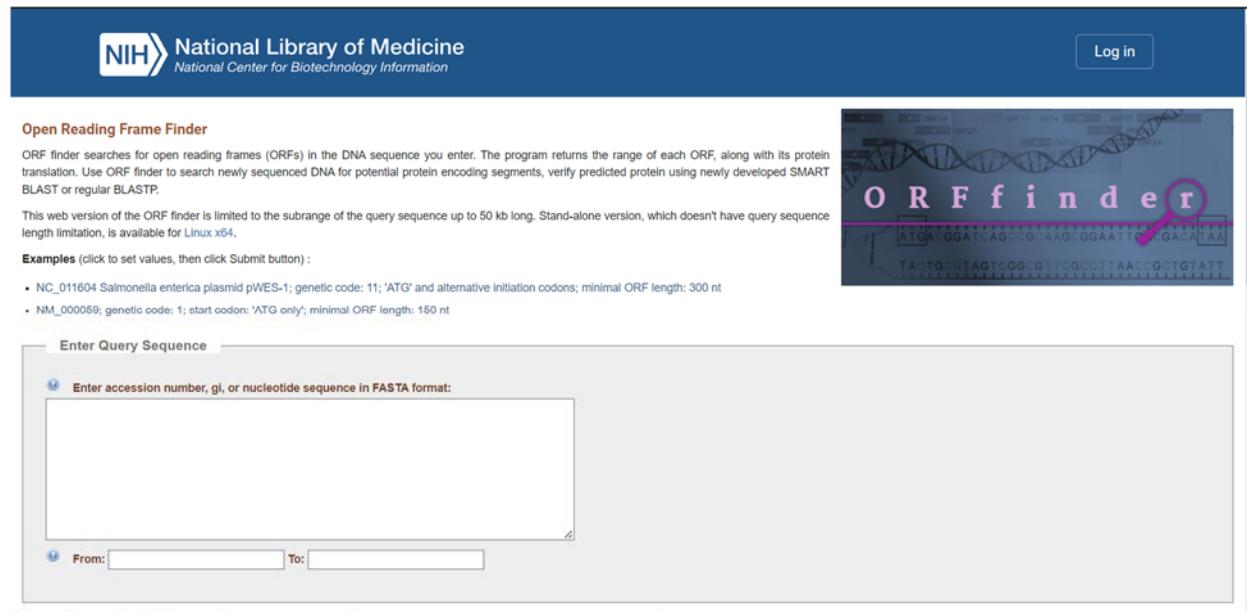


Fig. 8, ORF Finder homepage screenshot

ORF finder scans the DNA sequence you enter for open reading frames (ORFs). The software gives the ORF range as well as the protein translation for each ORF. Use the ORF finder to explore newly sequenced DNA for probable protein-encoding regions, then use the newly built SMART BLAST or standard BLASTP to confirm the predicted protein (Fig. 8).

This online version of the ORF finder is limited to a query sequence subrange of up to 50 kb in length. For Linux x64, there is a stand-alone version that does not have a query sequence length constraint.

A DNA strand contains three different reading frames because DNA is interpreted in groups of three nucleotides (codons). A DNA molecule's double helix

comprises two anti-parallel strands; with each strand having three reading frames, there are six potential frame translations.

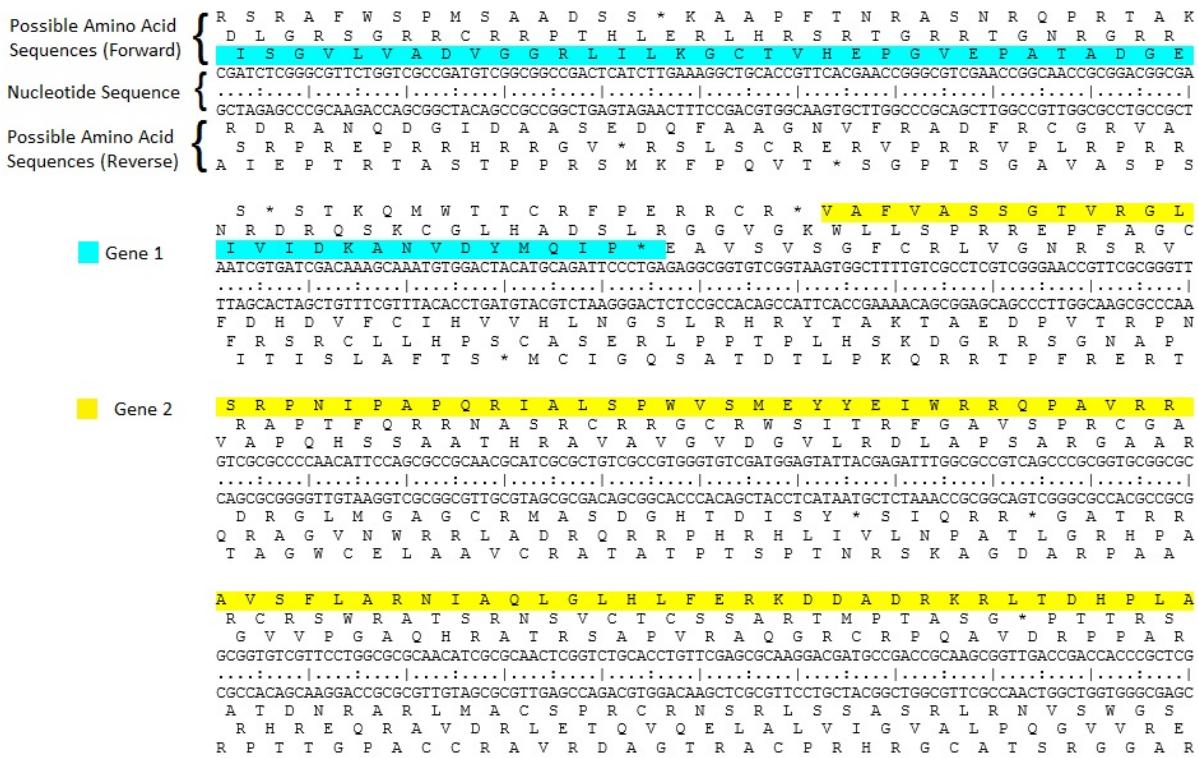


Fig. 9, Example of a six-frame translation. The nucleotide sequence is shown in the middle with forward translations above and reverse translations below. Two possible open reading frames with the sequences are highlighted

The ORF Finder (Open Reading Frame Finder) is a graphical analytic tool that detects all open reading frames of a user-specified minimum size in a user-supplied sequence or a sequence already in the database. This utility finds all open reading frames using conventional or other genetic codes. The deduced amino acid sequence can be saved in a variety of formats and compared against a database of sequences using the basic local alignment search tool (BLAST) service. The ORF Finder should aid in the preparation of comprehensive and correct sequence submissions. It also includes the Sequin sequence submission software (sequence analyser) [Pearson WR. et al 1997] (Fig. 9).

PDB Database

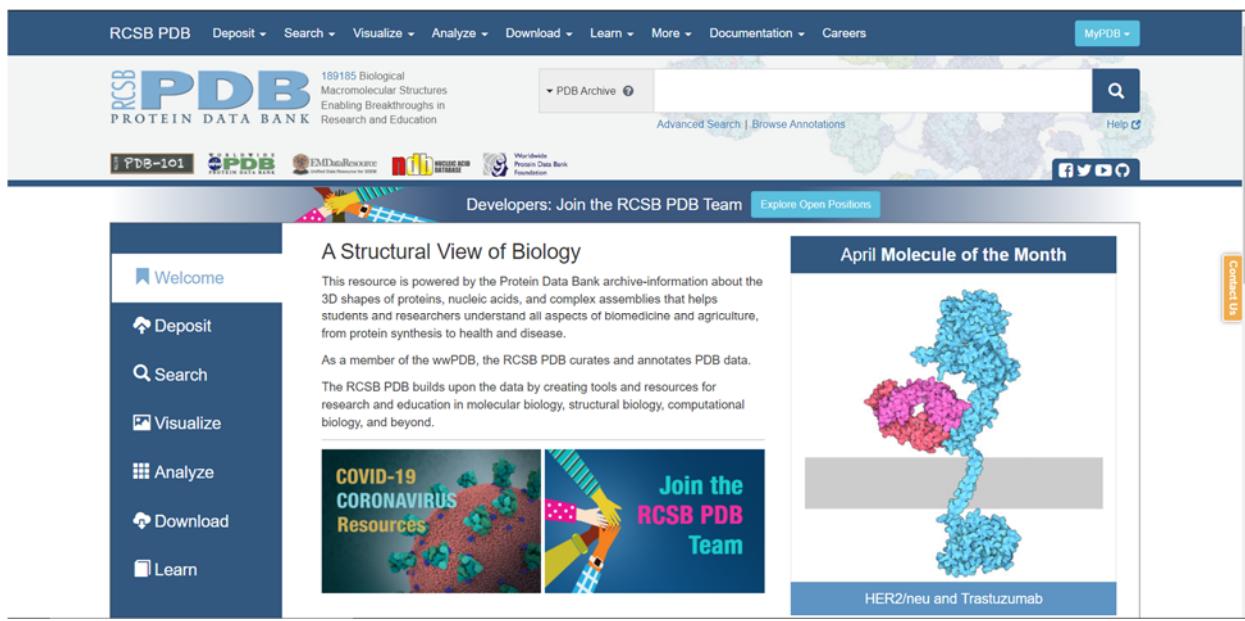


Fig. 10, Protein Data Bank homepage screenshot.

The Protein Data Bank (PDB) is a database that contains three-dimensional structural data for big biological entities including proteins and nucleic acids. The data, which is often collected by X-ray crystallography, NMR spectroscopy, or, increasingly, cryo-electron microscopy, and provided by biologists and biochemists from across the world, is freely available on the Internet via the websites of its member organizations (PDBe, PDBj, RCSB, and BMRB). The PDB is managed by the Worldwide Protein Data Bank, abbreviated as wwPDB (Fig. 10).

The PDB is essential in structural biology fields like as structural genomics. Scientists are now required to contribute their structural data to the PDB by most major scientific publications and several funding sources. Many additional databases make use of protein structures from the PDB. SCOP and CATH, for example, categorize protein structures, whereas PDBsum gives a graphical representation of PDB entries based on information from various sources such as Gene ontology.

The PDB file format was the first file format used by the PDB. The original format was limited to 80 characters per line due to the width of computer punch cards. The "macromolecular Crystallographic Information file" format, mmCIF, an expansion of the CIF format, was phased in around 1996. In 2014, mmCIF was adopted as the standard format for the PDB archive. The wwPDB declared in 2019 that crystallographic depositions will only be accepted in mmCIF format.

Experimental Method	Proteins	Nucleic Acids	Protein/Nucleic Acid complexes	Other	Total
X-ray diffraction	135170	2097	6945	4	144216
NMR	11337	1325	264	8	12934
Electron microscopy	3475	35	1136	0	4646
Hybrid	155	5	3	1	164
Other	286	4	6	13	309
<i>Total:</i>	150423	3466	8354	26	162269

Fig. 11, The PDB database, as well as its holdings list, is updated weekly (UTC+0 Wednesday). As of 1 April 2020, the PDB had the above-mentioned items in the above table.

Total 162269 of different data regarding proteins, nucleic acids, their complexes and others which have been studied via X-ray diffraction, NMR, Electron microscopy, Hybrid and Other techniques have been listed (Fig. 11).

Uniprot

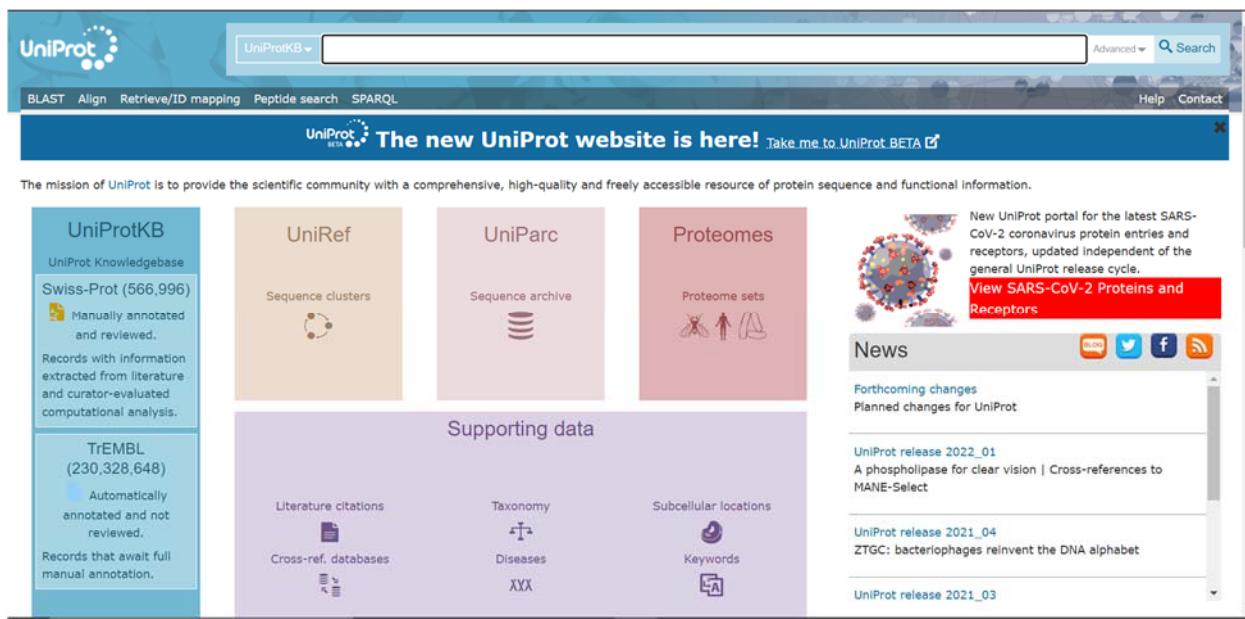


Fig. 12, Uniprot homepage screenshot.

UniProt is a publicly available database of protein sequence and functional information, with many entries coming from genome sequencing efforts. It offers a wealth of information drawn from the academic literature regarding the biological function of proteins. It is managed by the UniProt consortium, which is made up of numerous European bioinformatics organizations and a foundation based in Washington, DC (Fig. 12).

The Universal Protein Resource (UniProt) is a comprehensive database of protein sequence and annotation information. UniProt databases include the UniProt Knowledgebase (UniProtKB), UniProt Reference Clusters (UniRef), and UniProt Archive (UniParc). The UniProt collaboration and host institutions EMBL-EBI, SIB, and PIR are dedicated to preserving the UniProt databases in perpetuity [Uniprot Consortium 2009].

UniProt is a joint effort of the European Bioinformatics Institute (EMBL-EBI), the Swiss Institute of Bioinformatics (SIB), and the Protein Information Resource (PIR). More than 100 employees are active throughout the three institutes in various duties such as database curation, software development, and support.

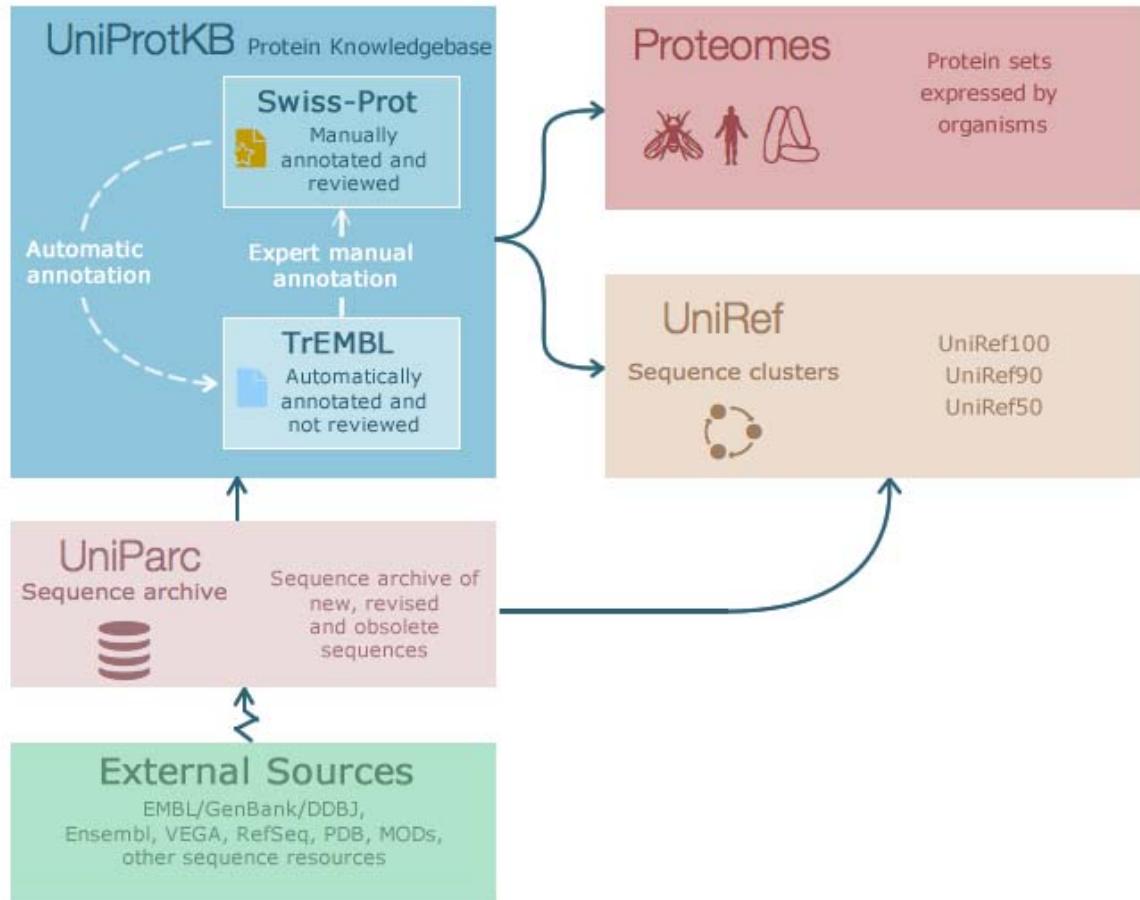


Fig. 13, UniProt databases include the UniProt Knowledgebase (UniProtKB), UniProt Reference Clusters (UniRef), and UniProt Archive (UniParc)

The flowchart in Fig. 13 depicts the interrelationships between several UniProt Databases and the data that they provide to users. Among these, UniProtKB is the most important and extensively used database.

ProtParam:

The screenshot shows the ProtParam page on the ExPasy website. At the top, there is a navigation bar with links for Home, About, SIB News, and Contact. Below the navigation bar is a search bar with a placeholder 'e.g. BLAST, UniProt, MSH6, Albumin...' and a magnifying glass icon. To the right of the search bar is a section titled 'What you can do with this resource' containing a list of protein analysis tools. Below the search bar is the ProtParam logo and a brief description of its function: 'Compute various physical and chemical parameters for a given protein sequence. The computed parameters include the molecular weight, theoretical pI (isoelectric point), amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY).'. A blue button labeled 'Browse the resource website' is located at the bottom left of the main content area.

Fig. 14, The above screenshots represent the ProtParam homepage.

ProtParam (References / Documentation) is a programme that calculates physical and chemical properties for a protein contained in Swiss-Prot or TrEMBL, as well as for a user-entered protein sequence. Upon clicking on ‘Browse the resource website’, protparam can be used by entering the protein sequence (Fig. 14).

ProtParam is hosted on the website ExPasy. It is a flexible and integrated portal that gives users access to more than 160 databases and supporting a wide range of life science and clinical research domains, including genomics, proteomics, and structural biology, as well as evolution and phylogeny, systems biology, and medical chemistry.

STRING Database

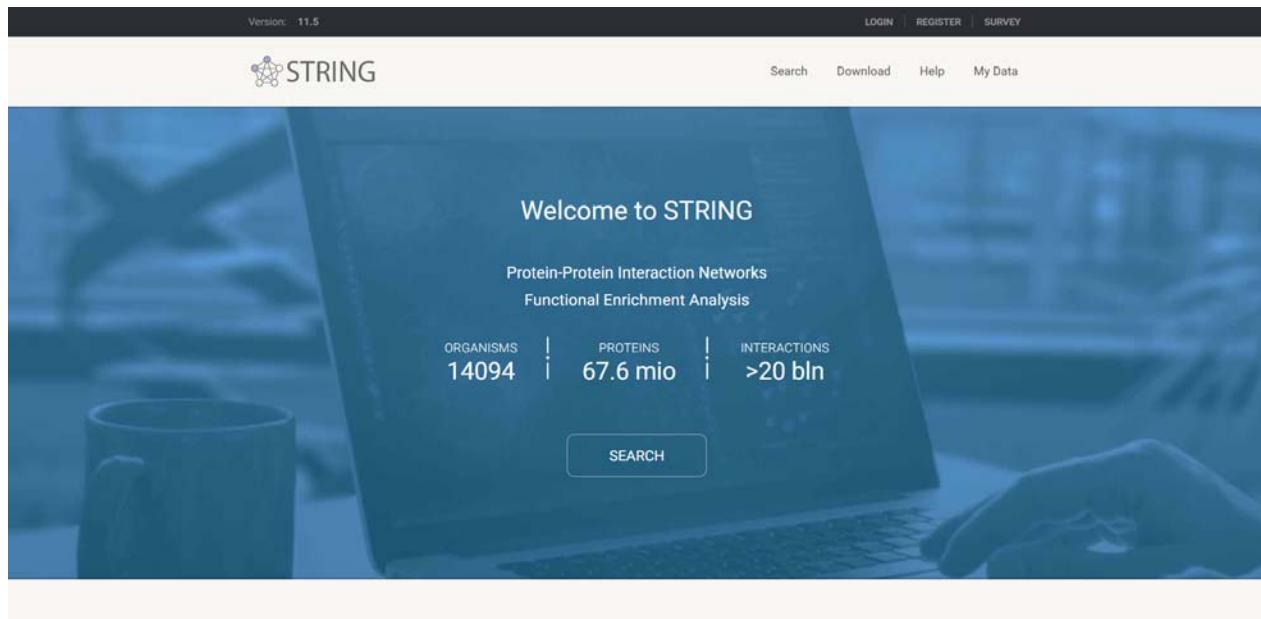


Fig. 15, The above screenshot is of the STRING Database homepage. Search option is used to proceed further and make any changes.

STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) is a web-based database and resource for known and predicted protein-protein interactions in molecular biology (Fig. 15).

The STRING database incorporates data from a variety of sources, including experimental data, computer prediction approaches, and publicly available text collections. It is open to the public and is updated on a regular basis. Using a variety of functional classification methods such as GO, Pfam, and KEGG, the resource additionally highlights functional enrichments in user-provided protein lists. The most recent version 11b contains data on around 24,5 million proteins from over 5000 species. A consortium of academic organizations, including CPR, EMBL, KU, SIB, TUD, and UZH, developed STRING.

For all protein interactions, the data is weighted and integrated, and a confidence score is obtained. The outcomes of various computational predictions can be examined from a variety of perspectives. All projected or imported interactions

are compared to a common functional partnership reference as defined by KEGG (Kyoto Encyclopedia of Genes and Genomes). [Mering C. et al 2003]

KEGG Pathway

KEGG PATHWAY Database
Wiring diagrams of molecular interactions, reactions and relations

KEGG2 PATHWAY BRITE MODULE KO GENES COMPOUND DISEASE DRUG

Select prefix Enter keywords Help

[New pathway maps | Update history]

Pathway Maps

KEGG PATHWAY is a collection of manually drawn pathway maps representing our knowledge of the molecular interaction, reaction and relation networks for:

1. Metabolism
Global/overview Carbohydrate Energy Lipid Nucleotide Amino acid Other amino Glycan
Cofactor/vitamin Terpenoid/PK Other secondary metabolite Xenobiotics Chemical structure
2. Genetic Information Processing
3. Environmental Information Processing
4. Cellular Processes
5. Organismal Systems
6. Human Diseases
7. Drug Development

KEGG PATHWAY is the reference database for pathway mapping in **KEGG Mapper**.

Pathway Identifiers

Each pathway map is identified by the combination of 2-4 letter prefix code and 5 digit number (see **KEGG Identifier**). The prefix has the following meaning:

map manually drawn reference pathway
ko reference pathway highlighting KOs
ec reference metabolic pathway highlighting EC numbers
rn reference metabolic pathway highlighting reactions
<org> organism-specific pathway generated by converting KOs to gene identifiers

and the numbers starting with the following:

011 global map (lines linked to KOs)
012 overview map (lines linked to KOs)

Fig. 16, The above figure is the from the KEGG Pathway webpage.

The Kyoto Encyclopedia of Genes and Genomes (KEGG) is a database that contains information about genomes, biological pathways, diseases, medications, and chemical compounds. Data analysis in genomics, metagenomics, metabolomics, and other omics studies, modelling and simulation in systems biology, and translational research in drug development are all examples of how KEGG is used in bioinformatics research and education (Fig. 16).

The wiring diagram database, KEGG PATHWAY, is at the heart of the KEGG resource. It's a set of route maps that includes genes, proteins, RNAs, chemical compounds, glycans, and chemical reactions, as well as disease genes and therapeutic targets, all of which are recorded as separate entries in KEGG's other databases [Kanehisa M. et al 2000]. The following sections are used to categorise the pathway maps:

- Metabolism
- Genetic information processing (transcription, translation, replication and repair, etc.)
- Environmental information processing (membrane transport, signal transduction, etc.)
- Cellular processes (cell growth, cell death, cell membrane functions, etc.)
- Organismal systems (immune system, endocrine system, nervous system, etc.)
- Human diseases
- Drug development

DrugBank

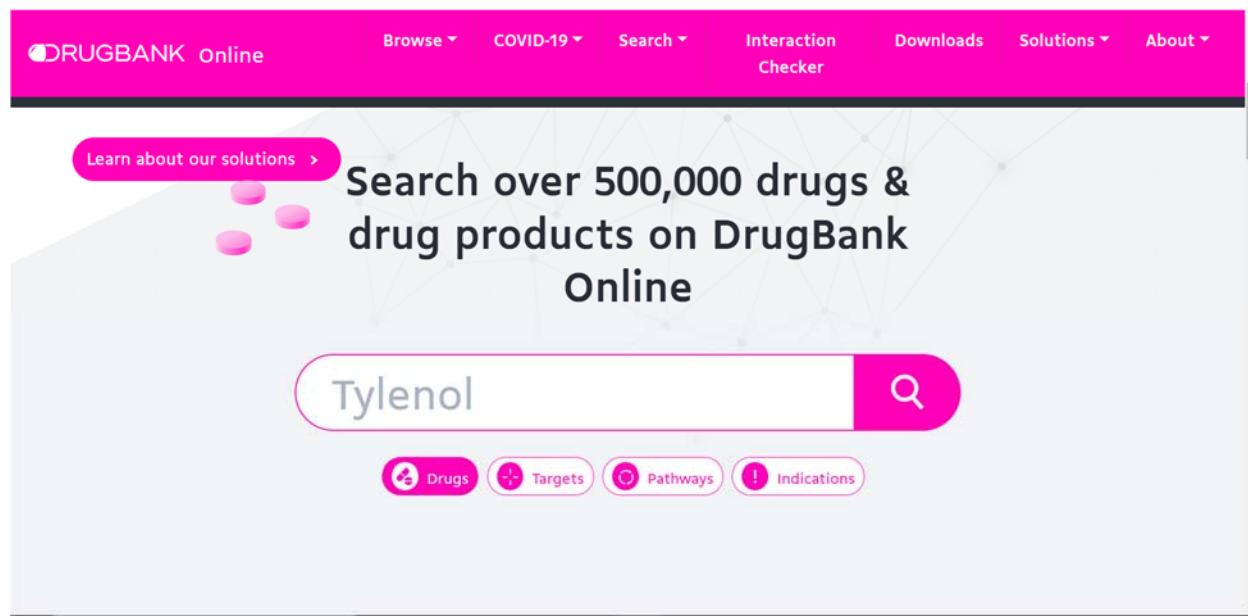


Fig. 17, DrugBank homepage screenshot.

The DrugBank database is a comprehensive, publicly available online database collecting information on medications and drug targets that was developed and is maintained by the University of Alberta and The Metabolomics Innovation Centre in Alberta, Canada (Fig. 17).

DrugBank is a bioinformatics and cheminformatics database that integrates detailed drug (chemical, pharmacological, and pharmaceutical) data with comprehensive drug target (sequence, structure, and route) data.

The DrugBank Online website is a free resource for the general population. However, use and re-distribution of content from DrugBank Online or the underlying DrugBank Data in whole or in part for any purpose necessitate a license. Academic users can apply for a free license for certain use cases, but all other users must purchase a license. [Wishart DS. et al 2006]

Out of the four options, users can choose in what relation is the drug being searched for:-

- Drugs
- Target
- Pathways
- Indications

PubChem

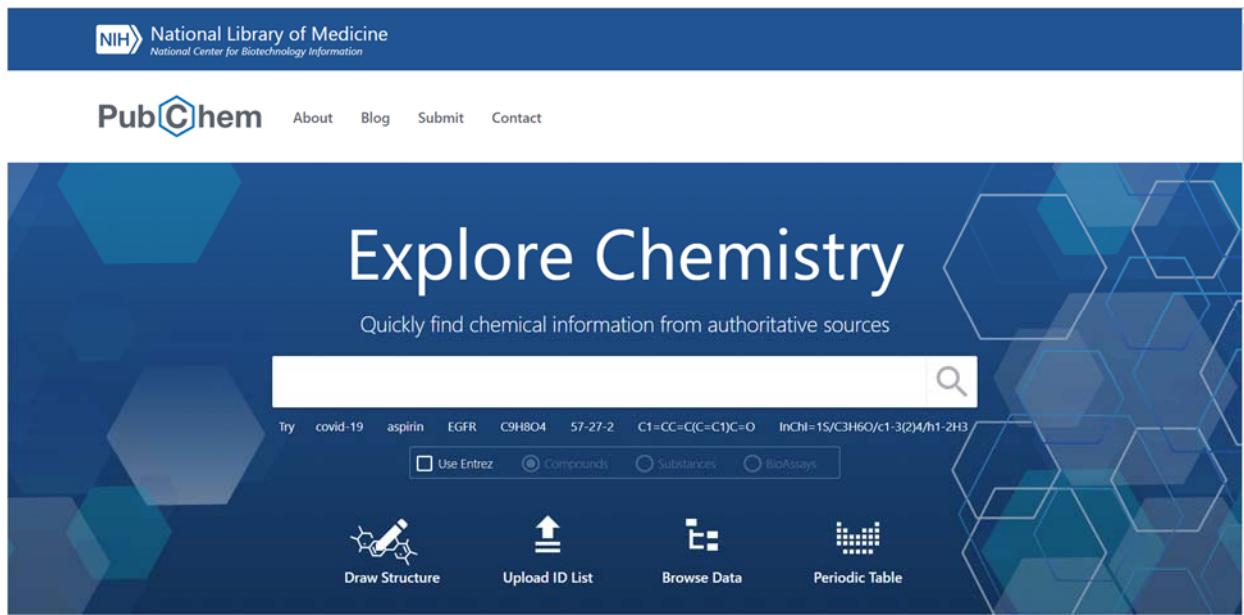


Fig. 18, The above screenshot represents the PubChem homepage.

PubChem is a database of chemical molecules and their activities against biological assays. The system is maintained by the National Center for Biotechnology Information (NCBI), a component of the National Library of Medicine, which is part of the United States National Institutes of Health (NIH). PubChem can be accessed for free through a web user interface. Millions of compound structures and descriptive datasets can be freely downloaded via FTP. PubChem contains multiple substance descriptions and small molecules with fewer than 100 atoms and 1,000 bonds. More than 80 database vendors contribute to the growing PubChem database (Fig. 18). [Sunghwan K. et al 2021]

PubChem consists of three dynamically growing primary databases. As of 5 November 2020 (number of BioAssays is unchanged):

- Compounds, 111 million entries (up from 94 million entries in 2017), contains pure and characterized chemical compounds.
- Substances, 293 million entries (up from 236 in 2017 and 163 million entries in Sept. 2014), contains also mixtures, extracts, complexes and uncharacterized substances.

SwissTargetPrediction

The screenshot shows the SwissTargetPrediction web interface. At the top, there is a navigation bar with links to SwissDrugDesign, SwissDock, SwissParam, SwissSidechain, SwissBioisostere, SwissTargetPrediction (which is the active page), SwissADME, SwissSimilarity, and About us. Below the navigation bar is the SIB logo and the text "Swiss Institute of Bioinformatics". The main title "SwissTargetPrediction" is centered above a menu bar with links to Home, FAQ, Help, Download, Contact, and Disclaimer. A large text box contains a brief description of the tool's purpose and its scientific foundation. Below this, there is a section titled "Select a species" with radio buttons for Homo sapiens, Mus musculus, and Rattus norvegicus, where Homo sapiens is selected. To the right is a Marvin JS molecule editor interface, which includes a toolbar with various chemical tools, a central workspace, and a periodic table sidebar. The workspace shows a molecular structure with atoms labeled H, C, N, O, S, F, P, Cl, Br, I, and *. Below the workspace, there is a note "(Provide a SMILES before submitting)". On the left, there is a text input field for "Paste a SMILES in this box, or draw a molecule", a dropdown for "Examples", and a "Clear" button. At the bottom, there is a "Predict targets" button.

Fig. 19, The above screenshot represents the Swiss Target Prediction.

SwissTargetPrediction is based on the assumption that bioactive compounds with comparable targets are more likely to share them. SwissTargetPrediction is a web service that uses a mix of 2D and 3D similarity metrics with known ligands to accurately predict the targets of bioactive compounds. Predictions can be made in five distinct animals, and for near paralogs and orthologs, mapping predictions by homology within and between species is possible (Fig. 19). [Gfeller D. et al 2014].

SWISS ADME

The screenshot shows the SwissADME homepage. At the top left is the SIB logo and "Swiss Institute of Bioinformatics" text. The top center features the "SwissADME" title. The top right has links for "Home", "FAQ", "Help", and "Terms of Use". Below the header, there's a text box containing descriptive text about the service, research papers, and development details. A note at the bottom states it's developed by the Molecular Modeling Group of the SIB | Swiss Institute of Bioinformatics. The main content area contains a Marvin JS molecular editor on the left and a large text input field on the right labeled "Enter a list of SMILES here:". At the bottom of the input field are buttons for "Fill with an example", "Clear", and "Run!". The ChemAxon logo is at the very bottom right.

Fig. 20, The above figure depicts SWISS ADME webpage. The two boxes provided correspond to the areas where we can either enter the SMILE of

the ligand and ‘Run!’ it or draw the entire structure of our need which will, in turn, generate a SMILE and then again ‘Run!’ it.

SwissADME is a free web tool for evaluating small molecule pharmacokinetics, drug-likeness, and medicinal chemistry friendliness (Fig. 20).

In SWISS ADME, ADME is an abbreviation for "absorption, distribution, metabolism, and excretion" in pharmacokinetics and pharmacology, and explains the disposition of a medicinal substance within an organism. The four criteria all have an impact on drug levels and kinetics of drug exposure in tissues, and hence on the compound's performance and pharmacological effectiveness as a medicine. [Diana A. et al 2017]

LogS is directly connected to a drug's water solubility and is defined as a common solubility unit equivalent to the 10-based logarithm of a molecule's solubility measured in mol/L.

Pharmacokinetics, sometimes known as PK, is a discipline of pharmacology concerned with determining the destiny of chemicals supplied to living organisms. BBB (Blood Brain Barrier) (A network of blood arteries and tissue made up of closely spaced cells that helps keep dangerous chemicals from reaching the brain) is one of the pharmacokinetic features. Some of the important factors to consider when taking a drug are blood-brain barrier (the blood-brain barrier allows some substances, such as water, oxygen, carbon dioxide, and general anesthetics, to pass into the brain) and GI absorption (GI absorption is altered secondary to delayed gastric emptying, decreased GI motility, and prolonged transit time through the GI tract).

The 'Lipinski' rule is one of the most significant principles in drug selection research, and it is one of the Druglikeness rules. Lipinski's rule of five, also known as Pfizer's rule of five or simply the rule of five (RO5), is a rule of thumb used to determine whether a chemical compound with a specific pharmacological or biological activity has chemical and physical properties that would make it a

likely orally active drug in humans. According to Lipinski's rule, an orally active medication should have no more than one violation of the following criteria [Lipinski CA. et al 2001]:

- No more than 5 hydrogen bond donors (the total number of nitrogen–hydrogen and oxygen–hydrogen bonds)
- No more than 10 hydrogen bond acceptors (all nitrogen or oxygen atoms)
- A molecular mass less than 500 daltons
- An octanol-water partition coefficient ($\log P$) that does not exceed 5

It's worth noting that all of the numbers are multiples of five, which is how the rule got its name. There are many exceptions to various rules of thumb, such as Baldwin's guidelines for ring closure.

Leadlikeness Lead-likeness is a tactical strategy for identifying chemical optimization starting points that have the highest possibility of producing "drug-like" candidates at the conclusion of drug development processes. [Alex P. et al 2008]

PreADME:



1. BACKGROUND

A significant bottleneck remains in the drug discovery procedure, in particular in the later stages of lead discovery, is analysis of the ADME and overt toxicity properties of drug candidates. Over 50% of the candidates failed due to ADME/Tox deficiencies during development. To avoid this failure at the development stage a set of in vitro ADME/Tox screens has been implemented in most pharmaceutical companies with the aim of discarding compounds in the discovery phase that are likely to fail further down the line. Even though the early stage in vitro ADME reduces the probability of the failure at the development stage, it is still time-consuming and resource-intensive. Therefore, we describe a new web-based application called PreADMET, which has been developed in response to a need for rapid prediction of drug-likeness and ADME/Tox data.

2. LINK

Site : <https://preadmet.qsarhub.com/>

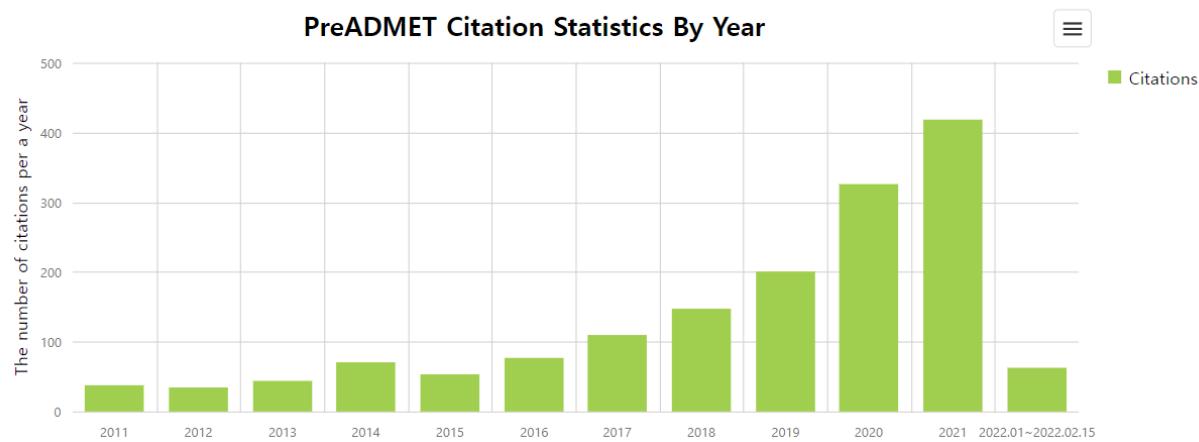


Fig. 21, The above screenshot represents the Pre ADME homepage. Different, options can be selected on the top right-hand corner of the homepage.

PreADMET is a web-based program that uses in silico methods to anticipate ADME data and construct drug-like libraries. PreAMDE offers somewhat same functionality of showing ADME Analysis of the compound (Fig. 21).

Molinspiration Tools

The screenshot displays the Molinspiration homepage with several sections:

- Molinspiration Products and Services:** Includes links for Calculation of Molecular Properties and Prediction of Bioactivity.
- Molinspiration Cheminformatics Software:** Describes the software's capabilities in molecule manipulation, SMILES conversion, and various molecular property calculations.
- Molinspiration now also on Touch Devices!** Shows mobile device interfaces and highlights the availability of touch devices.
- Free Web Tools for Cheminformatics Community:** Features a screenshot of a web tool for calculating molecular properties, mentioning over 80,000 processed molecules per month.
- More than 4500 Citations in Scientific Papers!** Notes the software's widespread use in academic research.
- Molinspiration Molecule Viewer:** Shows a grid of chemical structures and describes its features for visualization and search.

Fig. 22, The above screenshot is of the Molinspiration tool homepage. Different options as per the needs of the users can be accessed from the sidebar as per the user's needs.

Molinspiration provides a wide range of cheminformatics software tools. Cheminformatics (also known as chemoinformatics) is the application of physical chemistry theory to a variety of descriptive and prescriptive problems in the field of chemistry, including applications to biology and related molecular fields, using computer and information science techniques (so-called "in silico" techniques) (Fig. 22).

Pharmaceutical corporations and academic institutions, for example, use in silico approaches to aid and inform the drug discovery process, such as in the building of well-defined combinatorial libraries of synthetic compounds or to assist in structure-based drug design. Chemical and related businesses, as well as subjects like environmental science and pharmacology, where chemical processes are involved or investigated, can benefit from the methodologies. [Thomas E. et al 2006]

PyRx

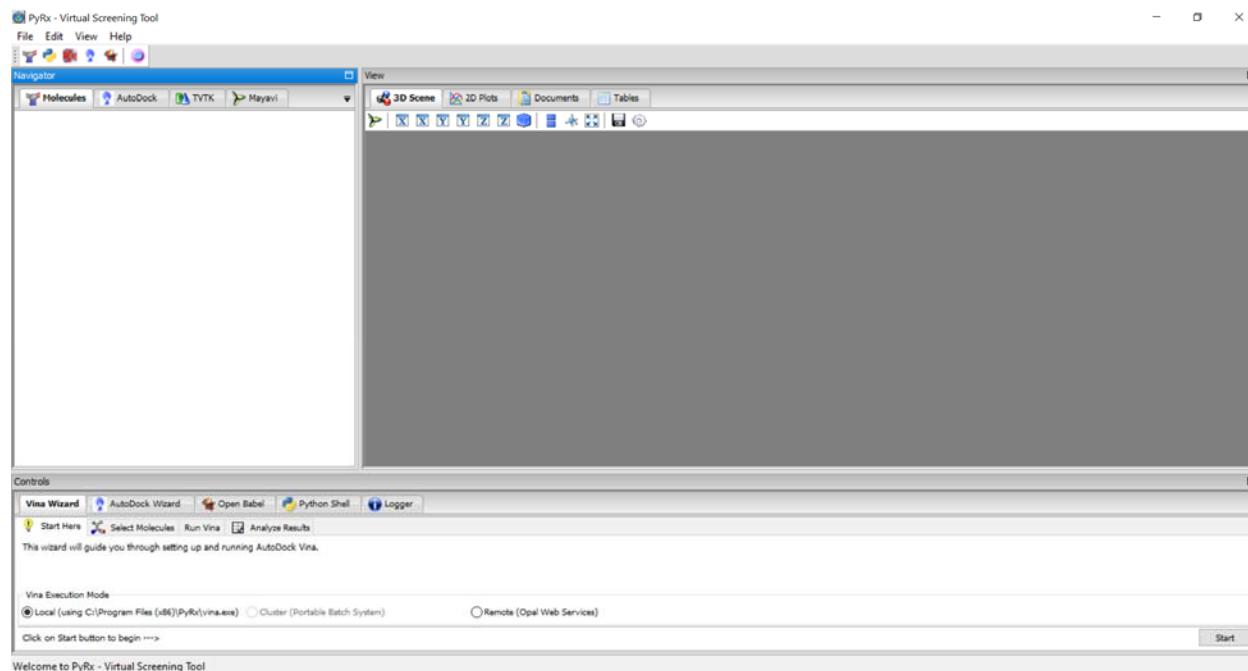


Fig. 23, The above screenshot represents the PyRx virtual screening tool which can be download and installed for PC usage.

PyRx is a Computational Therapeutic Discovery Virtual Screening software that can be used to screen libraries of compounds against prospective drug targets.

PyRx allows Medicinal Chemists to execute Virtual Screening from any platform, and it guides users through every step of the process, from data preparation to job submission and analysis. While there is no magic button in the drug discovery process, PyRx features a docking wizard with a simple user interface, making it a useful tool for Computer-Aided Drug Design. PyRx also offers a strong visualisation engine and chemical spreadsheet-like capability, both of which are critical for structure-based drug discovery.

PyRx provides a nice GUI for running virtual screening with AutoDock. PyRx includes a docking wizard and you can use it to run AutoDock Vina in the Cloud or HPC cluster (Fig. 23).

Biovia Discovery Studio

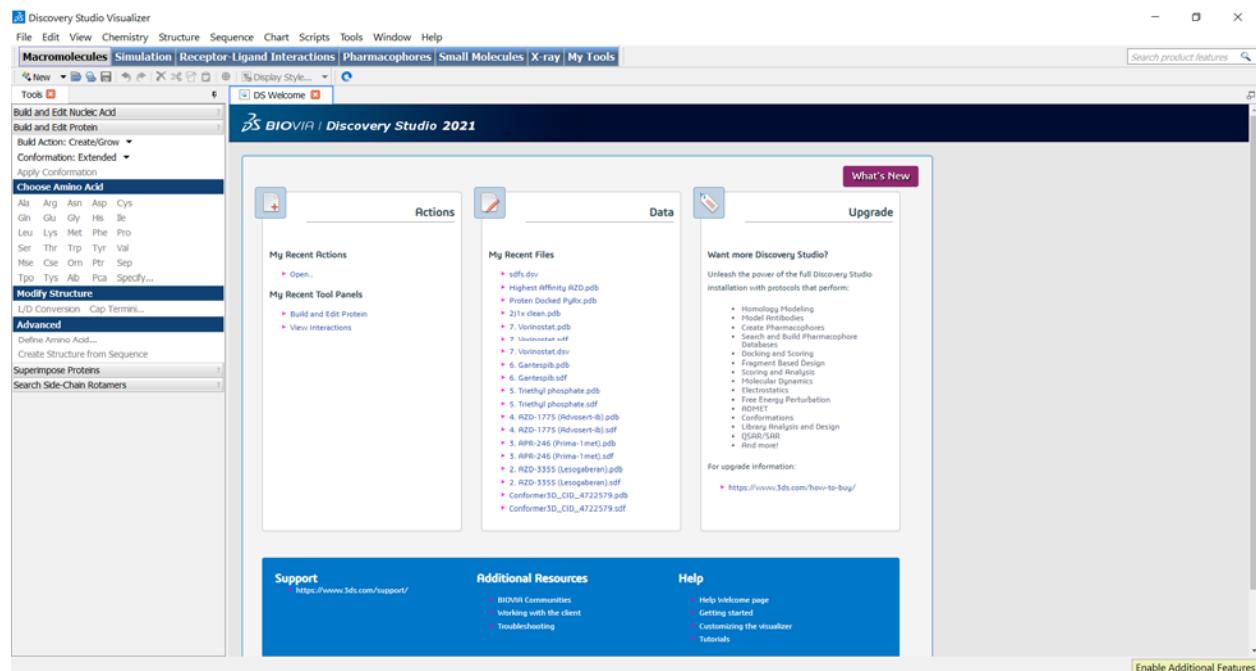


Fig. 24, The above figure represents screenshot of the Biovia Discovery Studio which is a downloadable software and can be installed on the PC.

Discovery Studio is a programme that simulates small molecule and macromolecule systems. Dassault Systemes BIOVIA is the company that created and distributed it.

Simulations, Ligand Design, Pharmacophore Moedling, Structure-based Design, Macromolecule Design and Validation, Macromolecule Engineering, QSAR, ADME, and Predictive Toxicity are some of the software applications offered by Discovery Studio. References (Fig. 24).

Results

NCBI

The screenshot shows the NCBI Gene search results for 'mutant p53'. The search term is entered in the dropdown search box. The results table lists three entries:

Name/Gene ID	Description	Location	Aliases	MIM
TP53	tumor protein p53 [Homo sapiens (human)] ID: 7157	Chromosome 17, NC_000017.11 (7668421..7687490, complement)	BCC7, BMFS5, LFS1, P53, TRP53	191170
Trp53	transformation related protein 53 [Mus musculus (house mouse)] ID: 22059	Chromosome 11, NC_000077.7 (69471174..69482699)	Tp53, bbl, bfy, p44, p53	
Tp53	tumor protein p53 [Rattus] ID: 22060	Chromosome 10,	Tp53, p53	

On the left sidebar, under 'Status', 'Current' is selected. Other filter categories like 'Gene sources', 'Categories', and 'Sequence content' are also listed.

Fig. 25, The above screen shot represents the search result of our target on NCBI.

Since our target is mutant p53 gene, we set the search through dropdown box to ‘Gene’ and searched it for ‘mutant p53’. The results of our search had been listed as seen in the ‘Search results’. We selected, the first result TP53 with NCBI Reference Sequence: NC_000017.11 (Fig. 25)

TP53 tumor protein p53 [*Homo sapiens* (human)]

[Download Datasets](#)

Gene ID: 7157, updated on 12-Jun-2022

Summary

Official Symbol TP53 provided by HGNC
Official Full Name tumor protein p53 provided by HGNC
Primary source HGNC:HGNC:11998
See related Ensembl:ENSG00000141510 MIM:191170; AllianceGenome:HGNC:11998
Gene type protein coding
RefSeq status REVIEWED
Organism *Homo sapiens*
Lineage Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as P53; BCC7; LFS1; BMFS5; TRP53
Summary This gene encodes a tumor suppressor protein containing transcriptional activation, DNA binding, and oligomerization domains. The encoded protein responds to diverse cellular stresses to regulate expression of target genes, thereby inducing cell cycle arrest, apoptosis, senescence, DNA repair, or changes in metabolism. Mutations in this gene are associated with a variety of human cancers, including hereditary cancers such as Li-Fraumeni syndrome. Alternative splicing of this gene and the use of alternate promoters result in multiple transcript variants and isoforms. Additional isoforms have also been shown to result from the use of alternate translation initiation codons from identical transcript variants (PMIDs: 12032546, 20937277). [provided by RefSeq, Dec 2016]
Expression Ubiquitous expression in spleen (RPKM 13.2), lymph node (RPKM 13.1) and 25 other tissues [See more](#)
Orthologs mouse all
NEW Try the new [Gene table](#)
Try the new [Transcript table](#)

Fig. 26, The result page of the TP53 tumour protein [*Homo sapiens* (humans)]

Upon opening the search result ‘TP53 tumour protein [*Homo sapiens* (humans)]’ which is on the serial no. 1 in the Fig. , the information for the gene became available which consisted of “Summary”, “Genomic context”, “Genomic regions, transcripts, and products”, “Expression”, “Bibliography” and so on relevant information to the gene. (Fig. 26).

FASTA nucleotide sequence for the NCBI Reference Sequence: NC_000017.11 is available on:

“https://www.ncbi.nlm.nih.gov/nuccore/NC_000017.11?report=fasta&from=7668421&to=7687490&strand=true”

BLAST

Description:

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	mutant p53 [Homo sapiens]	Homo sapiens	815	815	100%	0.0	100.00%	393	ACI25593.1
<input checked="" type="checkbox"/>	cellular tumor antigen p53 [Pan paniscus]	Pan paniscus	812	812	100%	0.0	99.75%	393	XP_003810114.2
<input checked="" type="checkbox"/>	tumor protein p53 [synthetic construct]	synthetic construct	812	812	100%	0.0	99.75%	394	AAX42852.1
<input checked="" type="checkbox"/>	tumor protein p53 (Li-Fraumeni syndrome). isoform CRA_c [Homo sapiens]	Homo sapiens	812	812	100%	0.0	99.75%	408	FAW90142.1
<input checked="" type="checkbox"/>	tumor protein p53 (Li-Fraumeni syndrome) [Homo sapiens]	Homo sapiens	811	811	100%	0.0	99.49%	393	AAV38428.1
<input checked="" type="checkbox"/>	p53 transformation suppressor [Homo sapiens]	Homo sapiens	810	810	100%	0.0	99.49%	393	CAA42634.1
<input checked="" type="checkbox"/>	p53 antigen [Homo sapiens]	Homo sapiens	810	810	100%	0.0	99.49%	393	AAA61211.1
<input checked="" type="checkbox"/>	phosphoprotein p53 [Homo sapiens]	Homo sapiens	810	810	100%	0.0	99.49%	393	AAA59987.1
<input checked="" type="checkbox"/>	mutant p53 protein [Homo sapiens]	Homo sapiens	810	810	100%	0.0	99.49%	393	AEY81367.1
<input checked="" type="checkbox"/>	p53 transformation suppressor [Homo sapiens]	Homo sapiens	810	810	100%	0.0	99.49%	393	CAA42628.1
<input checked="" type="checkbox"/>	p53 transformation suppressor [Homo sapiens]	Homo sapiens	810	810	100%	0.0	99.49%	393	CAA42629.1
<input checked="" type="checkbox"/>	tumor suppressor protein p53 [Homo sapiens]	Homo sapiens	810	810	100%	0.0	99.49%	393	AAD28535.1
<input checked="" type="checkbox"/>	tumor protein p53 [synthetic construct]	synthetic construct	810	810	100%	0.0	99.49%	393	AAX36369.1
<input checked="" type="checkbox"/>	Tumor protein p53 [Homo sapiens]	Homo sapiens	810	810	100%	0.0	99.49%	393	AAH03596.1
<input checked="" type="checkbox"/>	p53 transformation suppressor [Homo sapiens]	Homo sapiens	809	809	100%	0.0	99.49%	393	CAA42626.1
<input checked="" type="checkbox"/>	p53 transformation suppressor [Homo sapiens]	Homo sapiens	809	809	100%	0.0	99.49%	393	CAA42635.1

Fig. 27, The above screenshot represents the search result of the blast search for mutant p53 protein.

The ‘mutant p53 [Homo sapiens]’ in the description and has a E-value 0.0 and Per. Ident (100%) (Fig. 27). The results are defined as: Maximum Score is the highest alignment score (bit-score) between the query sequence and the database segments. It is sort-of inversely proportional to the e-value. A larger bit score is less likely to be obtained by chance than is a smaller bit score.

Alignment:

mutant p53 [Homo sapiens]

Sequence ID: [ACI25593.1](#) Length: 393 Number of Matches: 1

Range 1: 1 to 393 [GenPept](#) [Graphics](#)

[▼ Next Match](#) [▲ Pre](#)

Score 815 bits(2105)	Expect 0.0	Method Compositional matrix adjust.	Identities 393/393(100%)	Positives 393/393(100%)	Gaps 0/393(0%)
Query 1	MEEPQSDPSVEPPLSQETFSDLWKLPPENNVLSPQLPSQAMDDMLSPDDIEQWFTEDPGP			60	
Sbjct 1	MEEPQSDPSVEPPLSQETFSDLWKLPPENNVLSPQLPSQAMDDMLSPDDIEQWFTEDPGP			60	
Query 61	DEAPRMPEAAPRVAAPAPAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRLGFLHSGTAK			120	
Sbjct 61	DEAPRMPEAAPRVAAPAPAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRLGFLHSGTAK			120	
Query 121	SVTCTYSPALNKMFCQLAKTCPVQLWVDSTTPPGTRVRAMAIYKQSQHMTEVVRRCPHHE			180	
Sbjct 121	SVTCTYSPALNKMFCQLAKTCPVQLWVDSTTPPGTRVRAMAIYKQSQHMTEVVRRCPHHE			180	
Query 181	RCSDSDGLAPPQHLIRVEVNLRVEYLDDRNTFRHSVVVPYEPPEVGSDCTTIHYNYMCNS			240	
Sbjct 181	RCSDSDGLAPPQHLIRVEVNLRVEYLDDRNTFRHSVVVPYEPPEVGSDCTTIHYNYMCNS			240	
Query 241	SCMGGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPGRDRRTEEENLRKKGEPHHELP			300	
Sbjct 241	SCMGGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPGRDRRTEEENLRKKGEPHHELP			300	
Query 301	PGSTKRALPNNTSSSPQPKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPG			360	
Sbjct 301	PGSTKRALPNNTSSSPQPKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPG			360	
Query 361	GSRAHSSHLKSKKGQSTSRRHKKLMFKTEGPDS	393			
Sbjct 361	GSRAHSSHLKSKKGQSTSRRHKKLMFKTEGPDS	393			

Fig. 28, The above figure represents the amino acid sequence alignment.

As per the above alignment information, our query sequence that we entered in the ‘blastp’ have the above similarity (Fig. 28) with the query protein. Method used was ‘Compositional matrix adjust.’ Positives were 100% suggesting a good match.

Graphic:

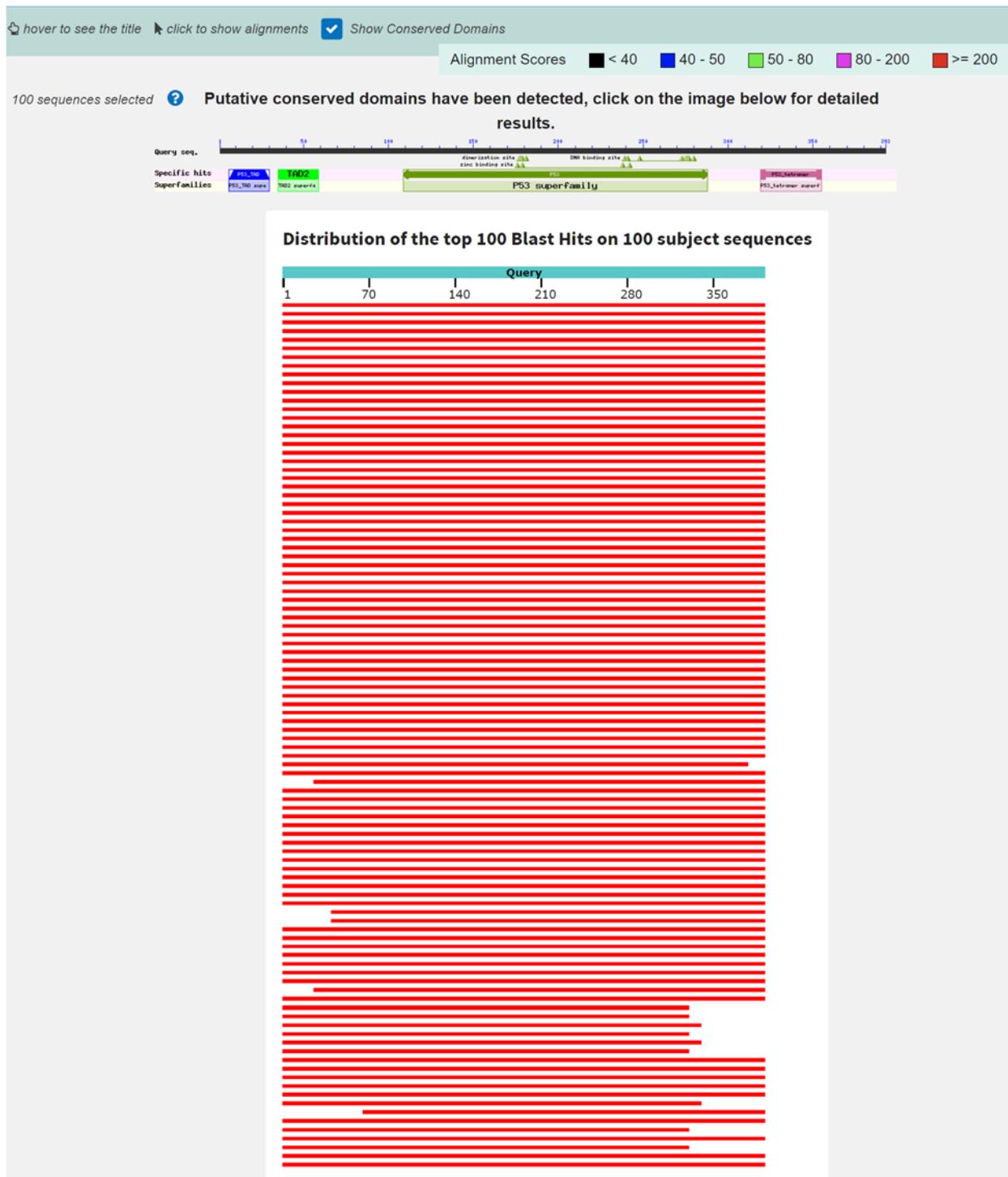


Fig. 29, The above screenshot represents the distribution of the top 100 Blast Hits on 100 subject sequences.

Each Red line corresponds to an organism that has the best hit in regards to our query (Fig. 29).

Conserved Domains:

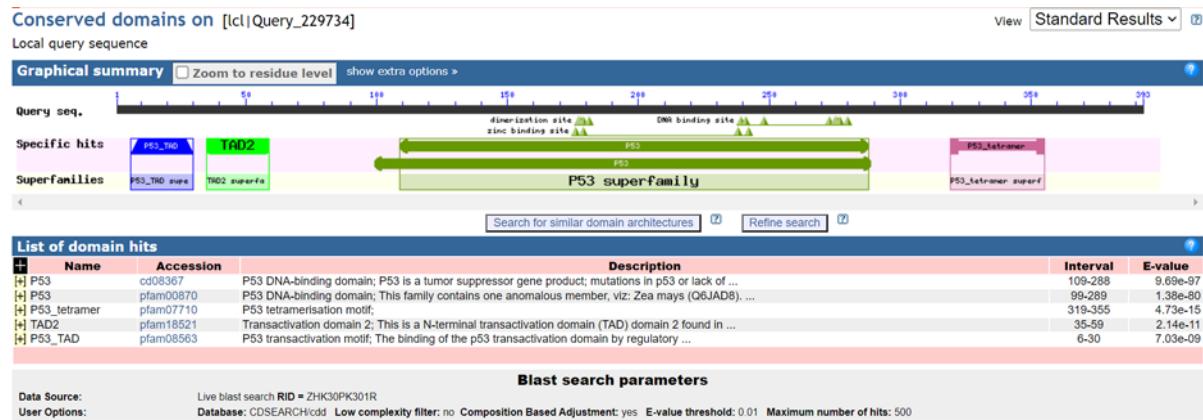


Fig. 30, The above screenshot represents the conserved domains of the the query sequence.

In total, there are about 5 domain hits, these are conserved domains and they can be observed in the graphical summary (Fig. 30).

CLUSTAL OMEGA

CLUSTAL O(1.2.4) multiple sequence alignment

sequence1	MEEPQSDPSVEPPLSQETFSDLWKL LPENNVLSPLPSQAMDDMLSPDDIEQWF TEDPGP	60
sequence2	MEEPQSDPSVEPPLSQETFSDLWKL LPENNVLSPLPSQAMDDMLSPDDIEQWF TEDPGP	60
sequence3	MEEPQSDPSVEPPLSQETFSDLWKL LPENNVLSPLPSQAMDDMLSPDDIEQWF TEDPGP	60

sequence1	DEAPRMPEAAPRVAPAPAAPT PAAPAPAPS WPLSSVPSQKTYQGSYGFRLGFLHSGTAK	120
sequence2	DEAPRMPEAAPPVAPAPAAPT PAAPAPAPS WPLSSVPSQKTYQGSYGFRLGFLHSGTAK	120
sequence3	DEAPRMPEAAPPVAPAPAAPT PAAPAPAPS WPLSSVPSQKTYQGSYGFRLGFLHSGTAK	120

sequence1	SVTCTYSPALNKMF CQLAKTCPVQLWVDST TPPPGTRVRAMAIYKQS QHMTEVVRRCPHHE	180
sequence2	SVTCTYSPALNKMF CQLAKTCPVQLWVDST TPPPGTRVRAMAIYKQS QHMTEVVRRCPHHE	180
sequence3	SVTCTYSPALNKMF CQLAKTCPVQLWVDST TPPPGTRVRAMAIYKQS QHMTEVVRRCPHHE	180

sequence1	RCSDSDGLAPPQHL IRVEVNLRVEYL LDDRNTFRHSVVVPY EPPEVGSDCTTIHY NYMCNS	240
sequence2	RCSDSDGLAPPQHL IRVEVNLRVEYL LDDRNTFRHSVVVPY EPPEVGSDCTTIHY NYMCNS	240
sequence3	RCSDSDGLAPPQHL IRVEVNLRVEYL LDDRNTFRHSVVVPY EPPEVGSDCTTIHY NYMCNS	240

sequence1	SCMGGMNRR PILTITLEDSSGN LLGRNSFEVRVC ACPGDRR TEENENLRKK GEPHHELP	300
sequence2	SCMGGMNRR PILTITLEDSSGN LLGRNSFEVRVC ACPGDRR TEENENLRKK GEPHHELP	300
sequence3	SCMGGMNRR PILTITLEDSSGN LLGRNSFEVRVC ACPGDRR TEENENLRKK GEPHHELP	300

sequence1	PGSTKRALPN NTSSSPQPKKPLD GEYFTLQIRGR ERFEMFREL NEALELKDAQAG KEPG	360
sequence2	PGSTKRALPN NTSSSPQPKKPLD GEYFTLQIRGR ERFEMFREL NEALELKDAQAG KEPG	360
sequence3	PGSTKRALPN NTSSSPQPKKPLD GEYFTLQIRGR ERFEMFREL NEALELKDAQAG KEPG	360

sequence1	GSRAHSSH LKSKKGQSTS RHKKLMFKTEGP DSD	393
sequence2	GSRAHSSH LKSKKGQSTS RHKKLMFKTEGP DSD	393
sequence3	GSRAHSSH LKSKKGQSTS RHKKLMFKTEGP DSD	393

Fig. 31, Three different sequences for mutant p53 protein were taken and the clustal omega was used in order to check the multiple sequence alignment. ‘Show Colors’ option was selected.

Multiple Sequence Alignment of the 3 sequences that are taken by us has been shown in the Fig. 31. The three sequence that were take were:

Sequence 1: mutant p53 [Homo sapiens]; GenBank: ACI25593.1

Sequence 2: cellular tumor antigen p53 isoform a [Homo sapiens]; NCBI Reference Sequence: NP_001119584.1

Sequence 3: cellular tumor antigen p53 isoform a [Homo sapiens]; NCBI Reference Sequence: NP_000537.3

GeneCard

The screenshot shows the GeneCard interface for the TP53 gene. At the top, there's a navigation bar with links for 'Follow Gene' and 'Phenotype Search'. Below the header, there's a row of logos for R&D, ORIGENE, SYNTHEGO, and InVivo Biosystems, each with associated services like Proteins, Antibodies, Assays, etc. The main content area has a section titled 'Aliases for TP53 Gene' which lists various aliases such as TP53, P53, LFS1, and BCC7. There are also sections for 'Jump to section' (with links to Paralogs, Pathways, Domains, Drugs, Expression, Genomics, Localization, and Orthologs), 'Protein Coding' (with a link to GCF17M007661), and 'Expression' (with a link to 50 publications).

Fig. 32, The above screenshot is the search result of the ‘mutant p53’ on the GeneCard website.

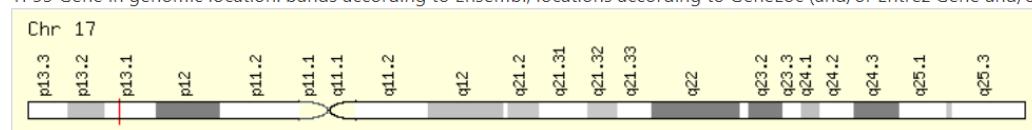
After entering the ‘mutant p53’ in the search query in the first option was selected which was tumour protein p53 from the list of results that appeared. Aliases for TP53 are shown in the Fig. 32, these correspond to the all the available related data to the tumour protein p53 or its similar. Different types of data as per the users needs can be accessed from the ‘Jump to section’.

Genomic View for TP53 Gene

Genes around TP53 on UCSC Golden Path with [GeneCards](#) custom track

Cytogenetic band: 17p13.1 by [HGNC](#) 17p13.1 by [Entrez Gene](#) 17p13.1 by [Ensembl](#)

TP53 Gene in genomic location: bands according to Ensembl, locations according to GeneLoc (and/or Entrez Gene and/or Ensembl if different)



 [GeneLoc](#) Genomic Neighborhood • Exon Structure • Gene Density

RefSeq DNA sequence for TP53 Gene

NC_000017.11

Fig. 33, The position of TP53 gene on the Chr 17.

The TP53 gene is situated on the chromosome number 17 in *Homo sapiens*. It is situated on the shorter arm of the chromosome 17. The thin red line on shorter arm of the chromosome number 17 (Fig. 33) denoted the precise location of the TP53 gene.

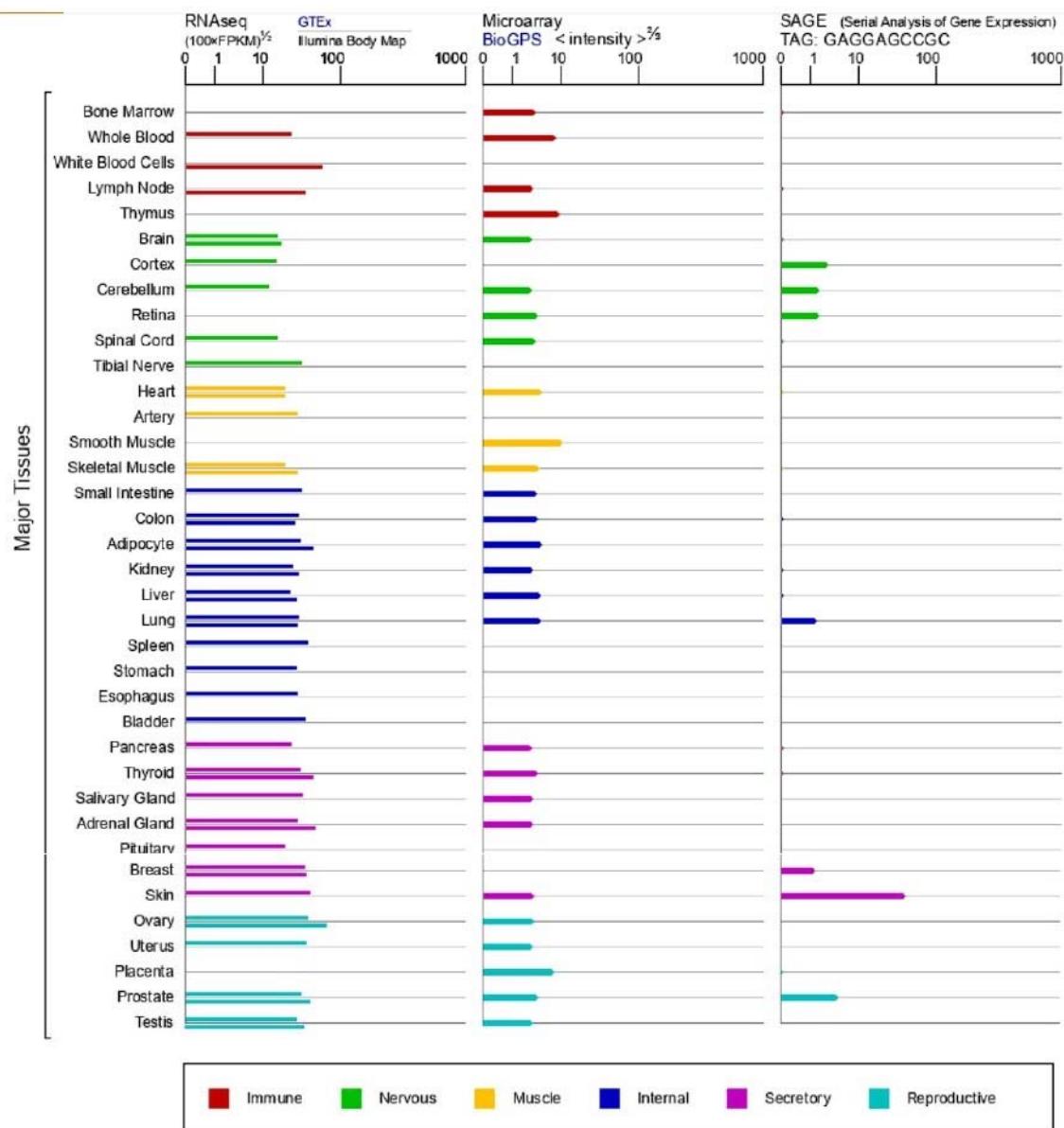


Fig. 34, Above if the screenshot of the ‘Function’ section of from the ‘Jump to section’ for TP53.

RNAseq, Microarray and SAGE which are all the sequencing analysis techniques have been given above with their data for the expression of TP53 in different tissues is shown (Fig. 33). As per the Fig. 34, most of the expression is observed in skin.

Open Reading Frame (ORF) Finder

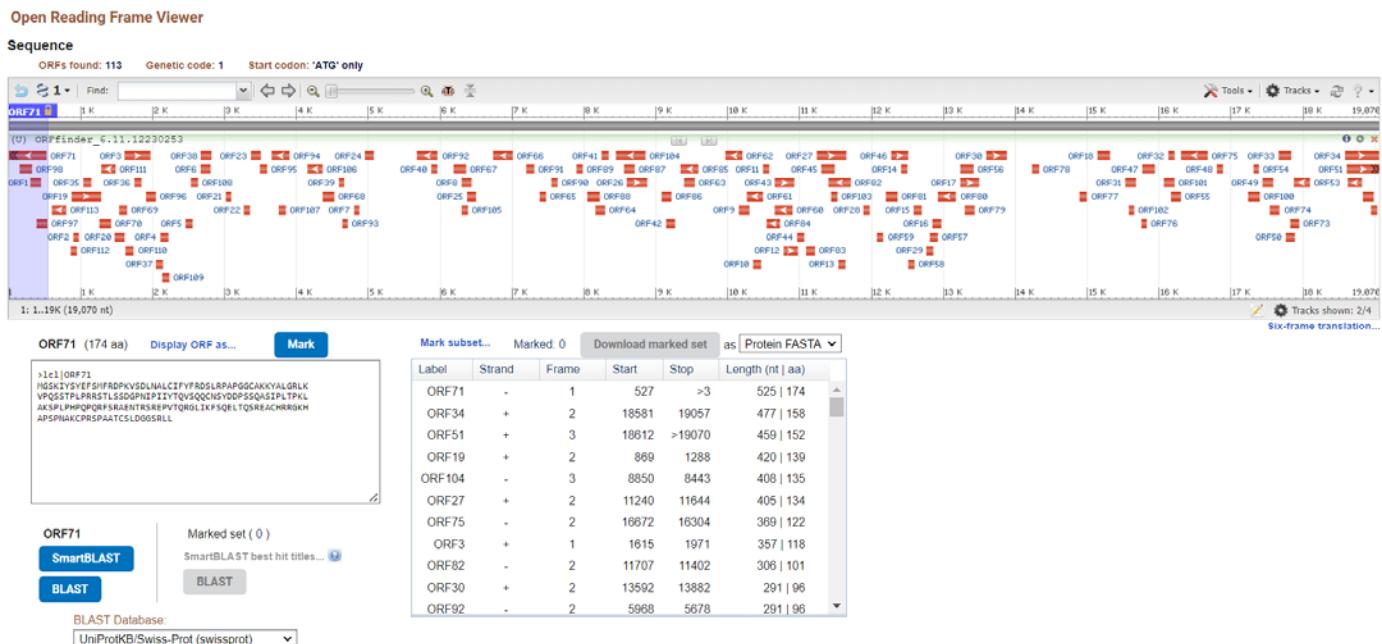


Fig. 35, The above the screenshot shows the ORF's in the mutant p53 gene nucleotide sequence.

Mutant p53 gene with accession number ‘ACI25593.1’ for Homo sapiens was used for finding the ORF’s (Fig. 35). The ‘Start Codon’ was set to ‘ATG Only’. Total of 113 open reading frames were found.

PDB Database

Property	Result
Protein	2J1X Human p53 core domain mutant M133L-V203A-Y220C-N239Y-N268D
UniProtKB accession	P04637

Structure	 <p>Global Symmetry: Asymmetric – C1 Global Stoichiometry: Monomer – A1</p>
PDB DOI	10.22110/pdb2J1X/pdb
Classification	NUCLEAR PROTEIN
Organism(s)	Homo sapiens
Expression System	Escherichia coli
Mutation(s) and their count (if present)	Yes and 5
Experimental Data Snapshot	

Method	X-RAY DIFFRACTION
Resolution	1.64 Å
R-Value Free	0.206
R-Value Work	0.185
R-Value Observed	0.185

Table 1, PDB Database search results for our selected protein 2J1X.

Relevant data for the study of 2J1X was extracted to see different protein of our target protein (Table 1).

Uniprot

UniProtKB - P04637 (P53_HUMAN)

Display Help video BLAST Align Format Add to basket History Community curation (1) Add a publication Feedback

Entry Protein Cellular tumor antigen p53
Gene TP53
Organism Homo sapiens (Human)
Status Reviewed + Annotation score: ★★★★★ - Experimental evidence at protein level

Function¹

None

- Function
- Names & taxonomy
- Subcellular location
- Pathology & Biotech
- PTM / Processing
- Expression
- Interaction
- Structure

Acts as a tumor suppressor in many tumor types; induces growth arrest or apoptosis depending on the physiological circumstances and cell type (PubMed:11025664, PubMed:12524540, PubMed:12810724, PubMed:15186775, PubMed:15340061, PubMed:17317671, PubMed:17349958, PubMed:19556538, PubMed:20673990, PubMed:20959462, PubMed:22726440, PubMed:24051492, PubMed:9840937, PubMed:24652652).
Involved in cell cycle regulation as a trans-activator that acts to negatively regulate cell division by controlling a set of genes required for this process (PubMed:11025664, PubMed:12524540, PubMed:12810724, PubMed:15186775, PubMed:15340061, PubMed:17317671, PubMed:17349958, PubMed:19556538, PubMed:20673990, PubMed:20959462, PubMed:22726440, PubMed:24051492, PubMed:9840937, PubMed:24652652).
One of the activated genes is an inhibitor of cyclin-dependent kinases. Apoptosis induction seems to be mediated either by stimulation of BAX and FAS antigen expression, or by repression of Bcl-2 expression. Its pro-apoptotic activity is activated via its interaction with PPP1R13B/ASPP1 or TP53BP2/ASPP2 (PubMed:12524540).
However, this activity is inhibited when the interaction with PPP1R13B/ASPP1 or TP53BP2/ASPP2 is disrupted by DDX101 or DDX102.

Fig. 36, The above screenshot represents the result of the Uniprot search for the mutant p53.

Here it's observable that the result for 'Cellular tumour antigen p53' is the protein which is our target. UniProtKB – P04637 is the uniprot id and the side bar can be accessed to see various properties of our target as per our need (Fig. 36).

Keywords ⁱ	
Molecular function	Activator, DNA-binding, Repressor
Biological process	Apoptosis, Biological rhythms, Cell cycle, Host-virus interaction, Necrosis, Transcription, Transcription regulation

Fig. 37, Keyword section of Uniprot result.

The keyword clearly suggest that our search result point out to the specific mutant p53 of our need as words like Apoptosis suggest cancerous properties of the gene (Fig. 37).

Family & Domainsⁱ

Region				Actions	Graphical view	Length
Feature key	Position(s)	Description				
Region ⁱ	1 – 320	Interaction with CCAR2	1 Publication	Add BLAST		320
Region ⁱ	1 – 83	Interaction with HRMT1L2	1 Publication	Add BLAST		83
Region ⁱ	1 – 44	Transcription activation (acidic)		Add BLAST		44
Region ⁱ	50 – 96	Disordered	Sequence analysis	Add BLAST		47
Region ⁱ	66 – 110	Interaction with WWOX		Add BLAST		45
Region ⁱ	100 – 370	Interaction with HIPK1	By similarity	Add BLAST		271
Region ⁱ	100 – 300	Required for interaction with ZNF385A	1 Publication	Add BLAST		201
Region ⁱ	113 – 236	Required for interaction with FBXO42	1 Publication	Add BLAST		124
Region ⁱ	116 – 292	Interaction with AXIN1	By similarity	Add BLAST		177
Region ⁱ	241 – 248	Interaction with the 53BP2 SH3 domain		Add BLAST		8
Region ⁱ	256 – 294	Interaction with E4F1	1 Publication	Add BLAST		39
Region ⁱ	273 – 280	Interaction with DNA		Add BLAST		8
Region ⁱ	282 – 325	Disordered	Sequence analysis	Add BLAST		44
Region ⁱ	300 – 393	Interaction with CARM1	1 Publication	Add BLAST		94
Region ⁱ	319 – 360	Interaction with HIPK2		Add BLAST		42
Region ⁱ	325 – 356	Oligomerization		Add BLAST		32
Region ⁱ	351 – 393	Disordered	Sequence analysis	Add BLAST		43
Region ⁱ	359 – 363	Interaction with USP7		Add BLAST		5
Region ⁱ	368 – 387	Basic (repression of DNA-binding)		Add BLAST		20

Fig. 38, The ‘Family and Domain’ information for the uniprot has been shown in the above screenshot.

Different ‘Families and Domains’ can be observed in the above screenshot for the mutant p53 and the relations which other proteins with CCAR2 and HIPK1 showing the highest similarity (Fig. 38).

ProtParam

Property	Value
Number of amino acids	393
Molecular weight	43754.33
Theoretical pI	6.47
Amino acid composition	Amino acid with highest quantity: 11.2% Amino acid with lowest quantity: Pyl (O) and Sec(O)
Total number of negatively charged residues (Asp + Glu)	50
Total number of positively charged residues (Arg + Lys)	47
Atomic composition	Carbon C 1902 Hydrogen H 2991 Nitrogen N 551 Oxygen O 592 Sulfur S 22
Formula	C ₁₉₀₂ H ₂₉₉₁ N ₅₅₁ O ₅₉₂ S ₂₂
Total number of atoms	6058
Extinction coefficients (Extinction coefficients are in units of M-1 cm-1, at 280 nm measured in water.)	Ext. coefficient 36035 Abs 0.1% (=1 g/l) 0.824, assuming all pairs of Cys residues form cystines

Table 2, ProtParam search results for Tumour p53 cancer's amino acid sequence for Homo sapiens (Humans)

ProtParam was used to get essential information about tumour protein p53 and different properties can be seen in the Table 2.

String Database

STRING

Search Download Help My Data

There are several matches for 'p53'. Please select one from the list below and press Continue to proceed.

<- BACK CONTINUE ->

1672 matches		showing page 1 of 84 • first • previous • next • last
organism	protein	
1) <input type="checkbox"/> Drosophila melanogaster	p53 - P53 protein long form variant 1; P53 is a transcriptional factor required for adaptive responses to genotoxic stress, including cell death, compensatory proliferation and DNA repair	
2) <input checked="" type="checkbox"/> Homo sapiens	TP53 - Cellular tumor antigen p53; Acts as a tumor suppressor in many tumor types; induces growth arrest or apoptosis depending on the physiological circumstances and cell type. Involved in cell cycle regulation as a trans-activator that acts to negatively regulate cell division by controlling a set of genes required for this process. One of the activated genes is an inhibitor of cyclin-dependent kinases. Apoptosis induction seems to be mediated either by stimulation of BAX and FAS antigen expression, or by repression of Bcl-2 expression. In cooperation with mitochondrial PPIF is involved in [...] [a.k.a. P53 , ENST00000514944 , AGA62702.1 , p53]	
3) <input type="checkbox"/> Mus musculus	Trp53 - Transformation related protein 53; Cellular tumor antigen p53; Acts as a tumor suppressor in many tumor types; induces growth arrest or apoptosis depending on the physiological circumstances and cell type. Involved in cell cycle regulation as a trans-activator that acts to negatively regulate cell division by controlling a set of genes required for this process. One of the activated genes is an inhibitor of cyclin-dependent kinases. Apoptosis induction seems to be mediated either by stimulation of BAX and FAS antigen expression, or by repression of Bcl-2 expression. In cooperation with [...] [a.k.a. Tp53 , P53 , Trp53-004 , p53]	
4) <input type="checkbox"/> Drosophila melanogaster	betaTub60D - Tubulin beta-3 chain; Tubulin is the major constituent of microtubules. It binds two moles of GTP, one at an exchangeable site on the beta chain and one at a non-exchangeable site on the alpha chain; Belongs to the tubulin family [a.k.a. FBgn0003888 , CG3401 , TubB60C , p53]	
5) <input type="checkbox"/> Drosophila melanogaster	hth - Homeobox protein homothorax; All isoforms are required for patterning of the embryonic cuticle. Acts with exd to delimit the eye field and prevent inappropriate eye development. Isoforms that carry the homeodomain are required for proper localization of chordotonal organs within the peripheral nervous system and antennal identity; required to activate antennal-specific genes, such as sal and to repress the leg-like expression of dac. Necessary for the nuclear localization of the essential HOX cofactor, extradenticle (exd). Both necessary and sufficient for inner photoreceptors to adopt [...] [a.k.a. FBgn0001235 , CG17117 , dtl , P53]	
6) <input type="checkbox"/> Homo sapiens	RFWD2 - E3 ubiquitin-protein ligase RFWD2; E3 ubiquitin-protein ligase that mediates ubiquitination and subsequent proteasomal degradation of target proteins. E3 ubiquitin ligases accept ubiquitin from an E2 ubiquitin-conjugating enzyme in the form of a thioester and then directly transfers the ubiquitin to targeted	

Fig. 39, The string database search result for p53 protein is being visible. As we are working on Homo Sapien in terms of cancer by mutant p53 gene. We have selected the 'Homo Sapiens' in the 'organisms' section.

From the list of all the available results (Fig. 39), our choice of selection which is "2) Homo sapiens" was chosen as our study revolves with this chosen organism.

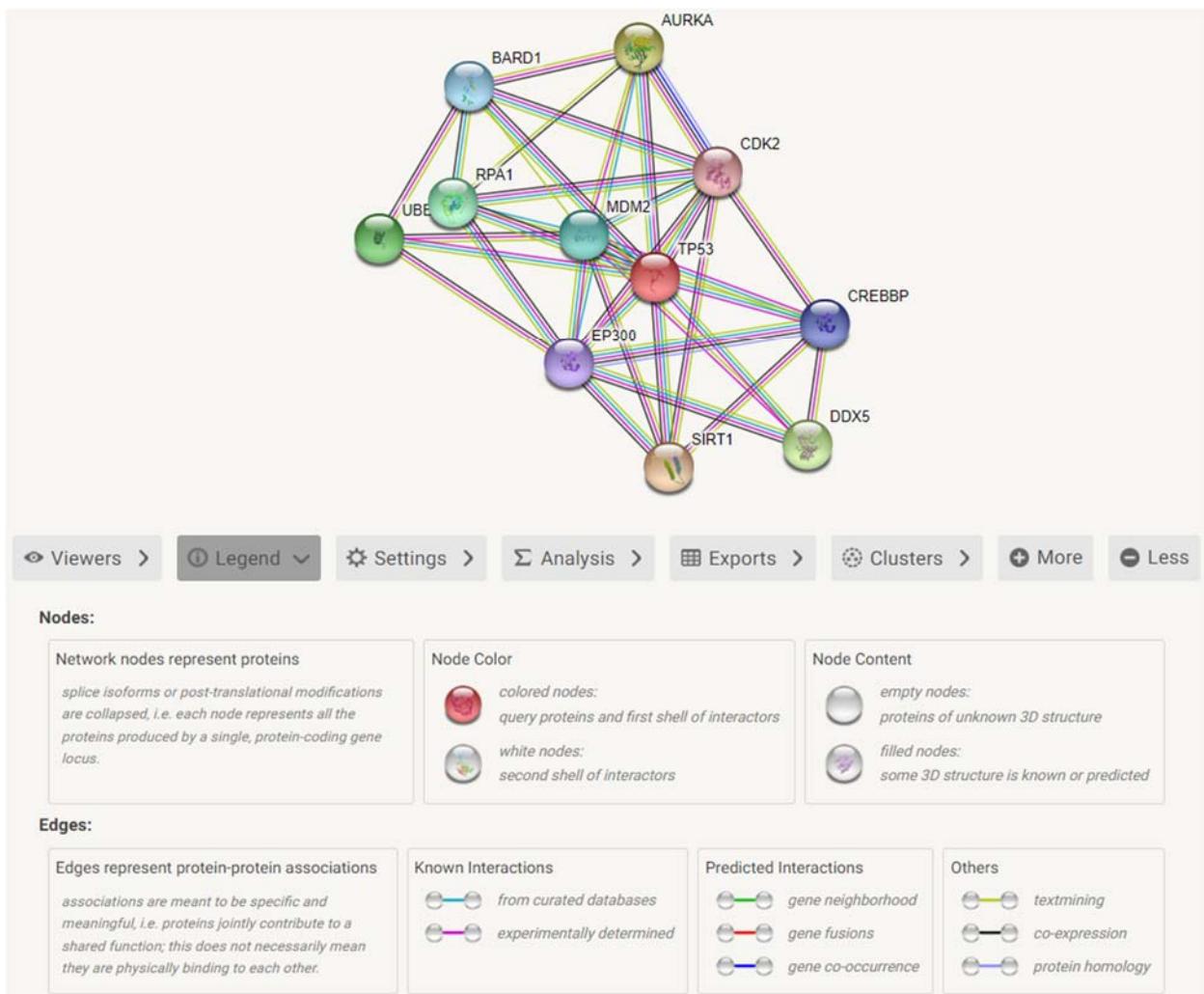


Fig. 40, The above figure represents the result of the search for ‘p53’ on the string database. Description of various component in the figure is also available in the the figure itself.

Nodes represented in the string diagram of p53 are proteins. All these nodes represent are proteins that are coded by a single, protein coding gene. In the ‘Node Color’ section, coloured nodes represent the 1st shell interactors which are the proteins directly associated with your input protein(s) and the query protein (the protein of our search). But, white nodes represent the second shell of interactors. The proteins in the second shell of interactors are those that are associated with the proteins in the first shell or with your input protein (s) (Fig. 40). It is possible for a 2nd shell protein to be directly related to your input

protein(s), but this is generally a weaker relationship, and so it will not be among the set number of 1st shell interactors. The color of the bubble indicates which shell the protein belongs to, while 2nd shell proteins are usually grey.

Edges like the figure describes (Fig. 40) represents the protein - protein interaction. As stated in the the figure in the “Edges” section, meaning proteins jointly contribute to a shared function, this does not necessarily mean they are physically binding to each other”. The known interaction, predicted interaction and the others are determined by their different color formats in the “Edge” section in the figure.

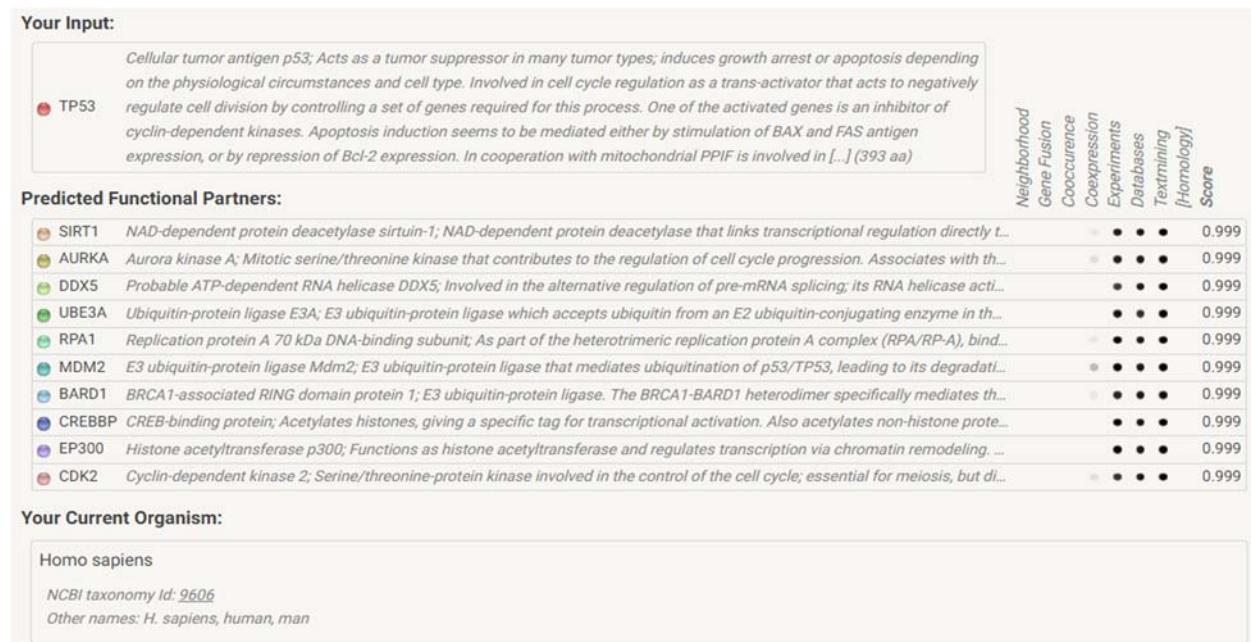


Fig. 41, This is one of the important part of the string database searches. Depending on our input (“Your Input:”), the string database has provided us with “Predicted Functional Partners” with scores.

Functional links between proteins can often be inferred from genomic associations between the genes that encode them: groups of genes that are required for the same function tend to show similar species coverage, are often located in close proximity on the genome (in prokaryotes), and tend to be involved in gene-fusion

events. The database STRING is a precomputed global resource for the exploration and analysis of these associations. Since the three types of evidence differ conceptually, and the number of predicted interactions is very large, it is essential to be able to assess and compare the significance of individual predictions. Thus, STRING contains a unique scoring framework based on benchmarks of the different types of associations against a common reference set, integrated in a single confidence score per prediction. (Fig. 41).

As visible under ‘Coexpression’ column (This identifies which genes have a tendency to show a coordinated expression pattern across a group of samples) of the ‘Predicted Functional Partners’, is it observable that MDM2 which is encoded by the mouse double minute 2 (MDM2) gene is the primary negative regulatory factor of the p53 protein and can ligate the p53 protein via its E3 ubiquitin ligase, and the ubiquitinated p53 can be transferred to the cytoplasm and degraded by proteasomes has the most Coexpression with the p53. (Fig. 41)

KEGG Pathway

Pathway Text Search					
Number of entries in a page 20 ▾ Hide thumbnail					
Page : 1 Go of 4 Items : 1 - 20 of 66 Top Previous Next Bottom					
map04115		p53 signaling pathway	p53 activation is induced by a number of stress signals, including DNA damage, oxidative stress and25 (CDKN1A) K10137 (ZMAT3) K04726 (BID) K10133 (TP53I3), K10134 (E124) K04390 (TNFRSF6), K10129 (GTSE...)	... cycle Apoptosis TSC2 Wip1 ΔNp73 Cyclin G Cop-1 p53R2 P48 TSAP6 PTEN Maspin KAI GD-Af TSP1 IGF-BP3...	

Fig. 42. It represents the search result for the p53 in the KEGG Pathway search. Here we have only shown only ‘map04115’ as it is the out preferred result for studying the p53 tumor pathway. However, the result had 66 search results for the same.

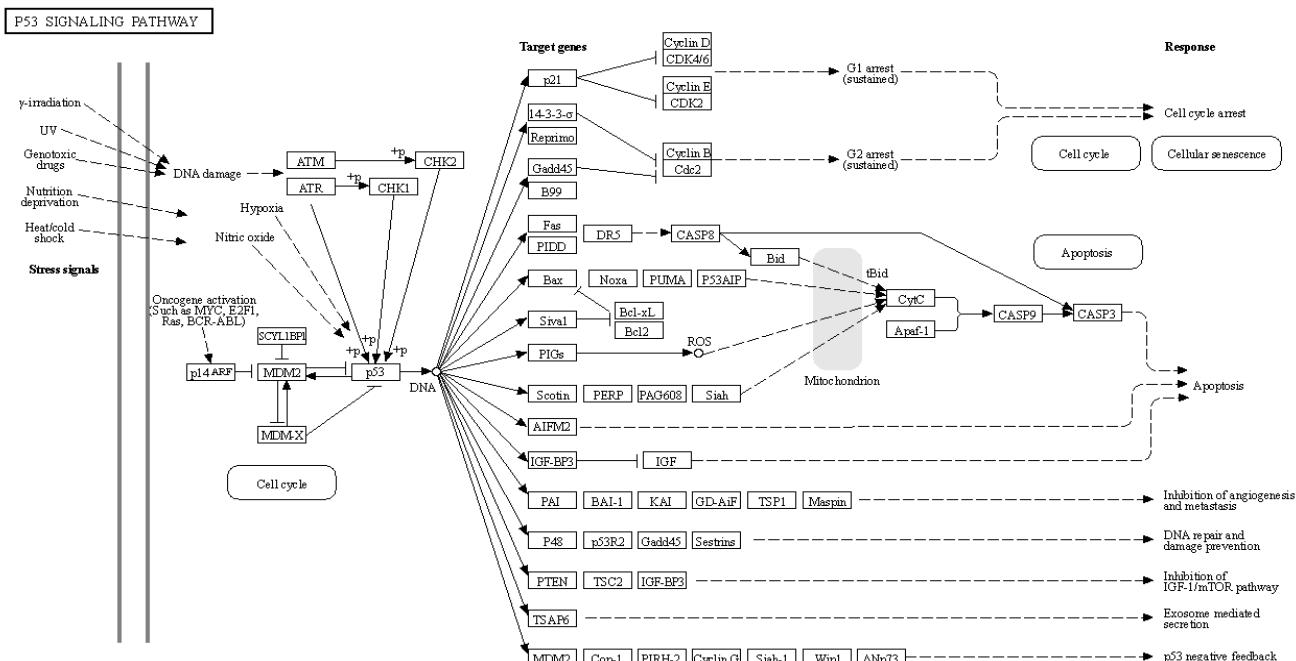


Fig. 43. The above figure refers to the ‘map04115’ which is the p53 signaling pathway.

As per Fig. 43, A variety of stress signals, including DNA damage, oxidative stress, and activated oncogenes, cause p53 activation. The p53 protein functions as a transcriptional activator of genes controlled by p53. This has three primary consequences: cell cycle arrest, cellular senescence, or apoptosis. Other p53-regulated gene activities interact with neighboring cells, repair damaged DNA, or establish positive and negative feedback loops that boost or decrease the p53 protein's actions and integrate these stress responses with other signal transduction pathways. Gamma-Radiation, UV and Genotoxic Drugs are some of the reasons that cause the DNA Damage and cause p53 (indirectly) to undergo mutation and show the cancerous properties. Also, Nutrition deprivation and sudden temperature changes also put stress and cause similar effects.

DRUGBANK

Table for the selected drugs against cellular tumour p53:-

Name	Target Name	DrugBank Id	Stage
1-(9-ethyl-9H-carbazol-3-yl)-N-methylmethanamine (PhiKan 083)	P53	DB08363	Experimental
AZD 3355	P53	DB05404	Investigational

(Lesogaberan)			
APR-246 (Prima-1met)	P53	DB11684	Investigational
AZD 1775 (Adavosertib)	Cancers and P53	DB11740	Investigational
Triethyl phosphate	P53	DB03347	Experimental
Ganetespib	BREAST CANCER, Small Cell Lung Cancer, Acute Myeloid Leukaemia, and Myelodysplas- tic Syndrome and P53	DB12047	Investigational
Vorinostat	Cutaneous T cell lymphoma	DB02546	Approved, Investigational

	(CTCL) and P53	
--	-------------------	--

Table 3, The above table contains drugs that had been selected from drug bank and various research papers for the purpose of our studies.

These were various categories including experimental, investigational, approved plus experimental and approved plus investigational. These have been selected for analysis of their effects and properties in relation to the mutant p53 cancer gene (Table 3).

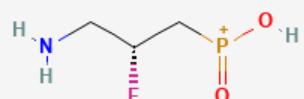
PUBCHEM

Table for structures of drugs:-

Drugs	2D - Structure	3D - Structure
1-(9-ethyl-9H-carbazol-3-yl)-N-methylmethanamine (PhiKan 083)		

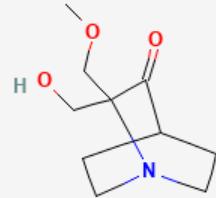
AZD-3355

(Lesogaberan)



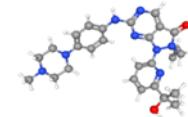
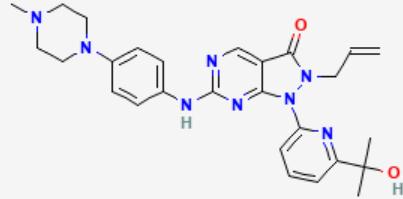
APR-246

(Prima-1met)



AZD-1775

(Adavosertib)



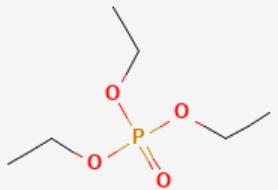
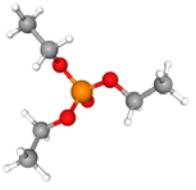
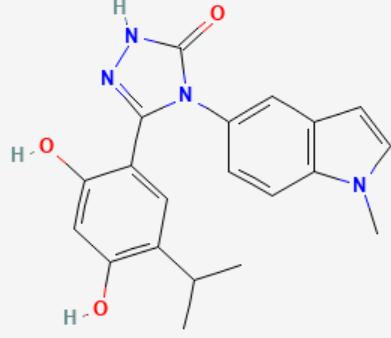
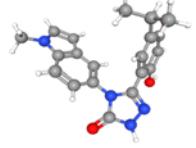
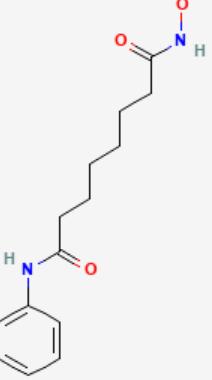
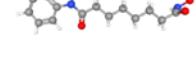
Triethyl phosphate		
Ganetespib		
Vorinostat		

Table 4, The above table depicts the 2D Strtucutes and 3D Confirmations of the different drugs selected for the studies.

Chemical and Physical Properties of Drugs:-

Drugs	1-(9-ethyl-9H-carbazol-1-yl)-N-methylmethanamine (PhiKan 083)	AZD-3355 (Lesogaberan)	APR-246 (Prima-1met)	AZD-1775 (Adavosert-ib)	Triethyl phosphate	Ganetes pib	Vorinostat
Molecular Formula	C16H18 N2	C3H8F NO2P+	C10H17 NO3	C27H32 N8O2	C6H15O4P	C20H20 N4O3	C14H20 N2O3
Molecular Weight	238.33	140.07	199.25	500.6	182.15	364.4	264.32
Canonical Smile	CCN1C2=C(C=C2)C(C=C3=CC=C31)CNC(C=C2)CO	C(C(C[P+](=O)(=O)FN)C(=O)C2CCN1CC2)CO	COCCC(=O)C2CCN1C(C=C2)=CC=C1N2C3=NC(=N=C=C3C(=O)N2C=C=C3N)C4=CC	CC(C)(C1=NC(C=C2)=CC=C1)N2C3=NC(=N=C=C3C(=O)N2C=C=C3N)C4=CC	CCOP(=O)(OCC)C1=C(C=C2)C(C(=C1)NC(=O)CCC1)C2=N	CC(C)C1=C(C=C2)C(C(=C1)NC(=O)CCC1)C2=N	C1=CC=C(C=C1)NC(=O)CCC1NC(=O)CCCC(=O)NO

				=C(C=C 4)N5CC N(CC5) C)O		C)O)O	
H Bond- Donors	1	2	1	2	0	3	3
H Bond- Accepto rs	1	4	4	9	4	4	3
Rotatab le Bonds	3	3	3	7	6	3	8
Topolog icalPola r Surface Area	17 Å²	63.3 Å²	49.8 Å²	101 Å²	44.8 Å²	90.1 Å²	78.4 Å²
Formal Charge	0	1	0	0	0	0	0
Comple xtiy	278	89.4	236	795	113	610	276
Covalen tly- Bonded	1	1	1	1	1	1	1

Unit						
Count						

Table 5, Different chemical and physical properties of drugs that are being used for the purpose of our studies have listed in the above table. These are some of the essential properties of the drugs that have been selected.

Pubchem is one of the best websites to study different types of compounds. For our studies we extracted information regarding the organic compounds or drugs such as SMILES, Molecular Formula, Structures, etc.,(Table 5)

SwissTargetPrediction

Target results of all the drugs in our studies have been listed in the tables below. Further information for the target with highest proportions have been listed in the tables itself.

1. 1-(9-ethyl-9H-carbazol-3-yl)-N-methylmethanamine or (PhiKan 083)

Property/Topic	Result														
Target Proportions	<p>A pie chart illustrating the distribution of target proportions. The largest segment is Family A G protein-coupled receptor at 73.3%. The remaining segments are Kinase (6.7%), Cytochrome P450 (6.7%), Membrane receptor (6.7%), and Isomerase (6.7%).</p> <table><thead><tr><th>Target Class</th><th>Proportion (%)</th></tr></thead><tbody><tr><td>Family A G protein-coupled receptor</td><td>73.3%</td></tr><tr><td>Isomerase</td><td>6.7%</td></tr><tr><td>Cytochrome P450</td><td>6.7%</td></tr><tr><td>Membrane receptor</td><td>6.7%</td></tr><tr><td>Kinase</td><td>6.7%</td></tr></tbody></table>	Target Class	Proportion (%)	Family A G protein-coupled receptor	73.3%	Isomerase	6.7%	Cytochrome P450	6.7%	Membrane receptor	6.7%	Kinase	6.7%		
Target Class	Proportion (%)														
Family A G protein-coupled receptor	73.3%														
Isomerase	6.7%														
Cytochrome P450	6.7%														
Membrane receptor	6.7%														
Kinase	6.7%														
Target	<table><thead><tr><th>Target</th><th>Common name</th><th>Uniprot ID</th><th>ChEMBL ID</th><th>Target Class</th><th>Probability*</th><th>Known actives (3D/2D)</th></tr></thead><tbody><tr><td>Serotonin 2a (5-HT2a) receptor</td><td>HTR2A</td><td>P28223</td><td>CHEMBL224</td><td>Family A G protein-coupled receptor</td><td><div style="width: 100px; height: 10px; background-color: green;"></div></td><td>247 / 40 ↘</td></tr></tbody></table>	Target	Common name	Uniprot ID	ChEMBL ID	Target Class	Probability*	Known actives (3D/2D)	Serotonin 2a (5-HT2a) receptor	HTR2A	P28223	CHEMBL224	Family A G protein-coupled receptor	<div style="width: 100px; height: 10px; background-color: green;"></div>	247 / 40 ↘
Target	Common name	Uniprot ID	ChEMBL ID	Target Class	Probability*	Known actives (3D/2D)									
Serotonin 2a (5-HT2a) receptor	HTR2A	P28223	CHEMBL224	Family A G protein-coupled receptor	<div style="width: 100px; height: 10px; background-color: green;"></div>	247 / 40 ↘									
UniProtKB	P28223 (HT2A_HUMAN)														
Gene	HTR2A														

Protein	Human Serotonin GPCR
Function	<ul style="list-style-type: none"> Beta-arrestin family members inhibit signaling via G proteins and mediate activation of alternative signaling pathways (PubMed:28129538).
GO: Biological Function	<ul style="list-style-type: none"> Cell Death

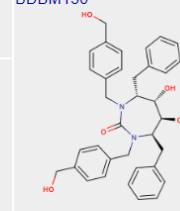
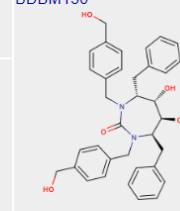
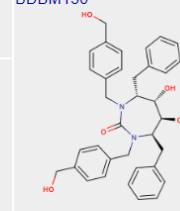
2. AZD-3355 (Lesogaberan)

Topic	Result														
Target Proportions	<table border="1"> <tr> <td>Family C G protein-coupled receptor</td> <td>Enzyme</td> <td>Protease</td> </tr> <tr> <td>Family A G protein-coupled receptor</td> <td>Transferase</td> <td></td> </tr> </table>	Family C G protein-coupled receptor	Enzyme	Protease	Family A G protein-coupled receptor	Transferase									
Family C G protein-coupled receptor	Enzyme	Protease													
Family A G protein-coupled receptor	Transferase														
Target	<table border="1"> <thead> <tr> <th>Target</th> <th>Common name</th> <th>Uniprot ID</th> <th>ChEMBL ID</th> <th>Target Class</th> <th>Probability*</th> <th>Known activity (3D)</th> </tr> </thead> <tbody> <tr> <td>GABA-B receptor</td> <td>GABBR2 GABBR1</td> <td>Q75899 Q9UBS5</td> <td>CHEMBL2111463</td> <td>Family C G protein-coupled receptor</td> <td><div style="width: 100%; background-color: green; height: 10px;"></div></td> <td>0 / 1</td> </tr> </tbody> </table>	Target	Common name	Uniprot ID	ChEMBL ID	Target Class	Probability*	Known activity (3D)	GABA-B receptor	GABBR2 GABBR1	Q75899 Q9UBS5	CHEMBL2111463	Family C G protein-coupled receptor	<div style="width: 100%; background-color: green; height: 10px;"></div>	0 / 1
Target	Common name	Uniprot ID	ChEMBL ID	Target Class	Probability*	Known activity (3D)									
GABA-B receptor	GABBR2 GABBR1	Q75899 Q9UBS5	CHEMBL2111463	Family C G protein-coupled receptor	<div style="width: 100%; background-color: green; height: 10px;"></div>	0 / 1									

UniProtKB	O75899(GABR2_HUMAN)
Gene	GABBR2
Protein	Gamma-aminobutyric acidtype B receptor subunit 2
Function	<ul style="list-style-type: none"> Component of a heterodimeric G-protein coupled receptor for GABA, formed by GABBR1 and GABBR2 (PubMed:9872316, PubMed:9872744, PubMed:15617512, PubMed:18165688, PubMed:22660477, PubMed:24305054).
GO: Biological Function	Nothing specific to TP53

3. APR-246 (Prima-1met)

NOTE: Searched using Binding DB

Property/Topic	Result																	
Target Proportions	Not mentioned in Binding DB																	
Target	<table border="1"> <thead> <tr> <th>Target/Host (Institution)</th> <th>Ligand</th> <th>Target/Host Links</th> <th>Ligand Links</th> <th>Trg + Lig Links</th> </tr> </thead> <tbody> <tr> <td> Human immunodeficiency virus type 1 protease (98/99 = 99%)[†] (Human immunodeficiency virus type 1) </td> <td>  BDBM150 ((4R,5S,6S,7R)-4,7-dibenzyl-5,6-dihydroxy-1,3-bis([...]) Show SMILES Show InChI </td> <td> PDB MMDB UniProtKB/TrEMBL DrugBank GoogleScholar </td> <td> CHEBI MMDB PC cid PC sid PDB UniChem Patents Similar </td> <td> MMDB PDB Article PubMed </td> </tr> <tr> <td>Uppsala University Curated by ChEMBL</td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table>	Target/Host (Institution)	Ligand	Target/Host Links	Ligand Links	Trg + Lig Links	Human immunodeficiency virus type 1 protease (98/99 = 99%) [†] (Human immunodeficiency virus type 1)	 BDBM150 ((4R,5S,6S,7R)-4,7-dibenzyl-5,6-dihydroxy-1,3-bis([...]) Show SMILES Show InChI	PDB MMDB UniProtKB/TrEMBL DrugBank GoogleScholar	CHEBI MMDB PC cid PC sid PDB UniChem Patents Similar	MMDB PDB Article PubMed	Uppsala University Curated by ChEMBL						
Target/Host (Institution)	Ligand	Target/Host Links	Ligand Links	Trg + Lig Links														
Human immunodeficiency virus type 1 protease (98/99 = 99%) [†] (Human immunodeficiency virus type 1)	 BDBM150 ((4R,5S,6S,7R)-4,7-dibenzyl-5,6-dihydroxy-1,3-bis([...]) Show SMILES Show InChI	PDB MMDB UniProtKB/TrEMBL DrugBank GoogleScholar	CHEBI MMDB PC cid PC sid PDB UniChem Patents Similar	MMDB PDB Article PubMed														
Uppsala University Curated by ChEMBL																		

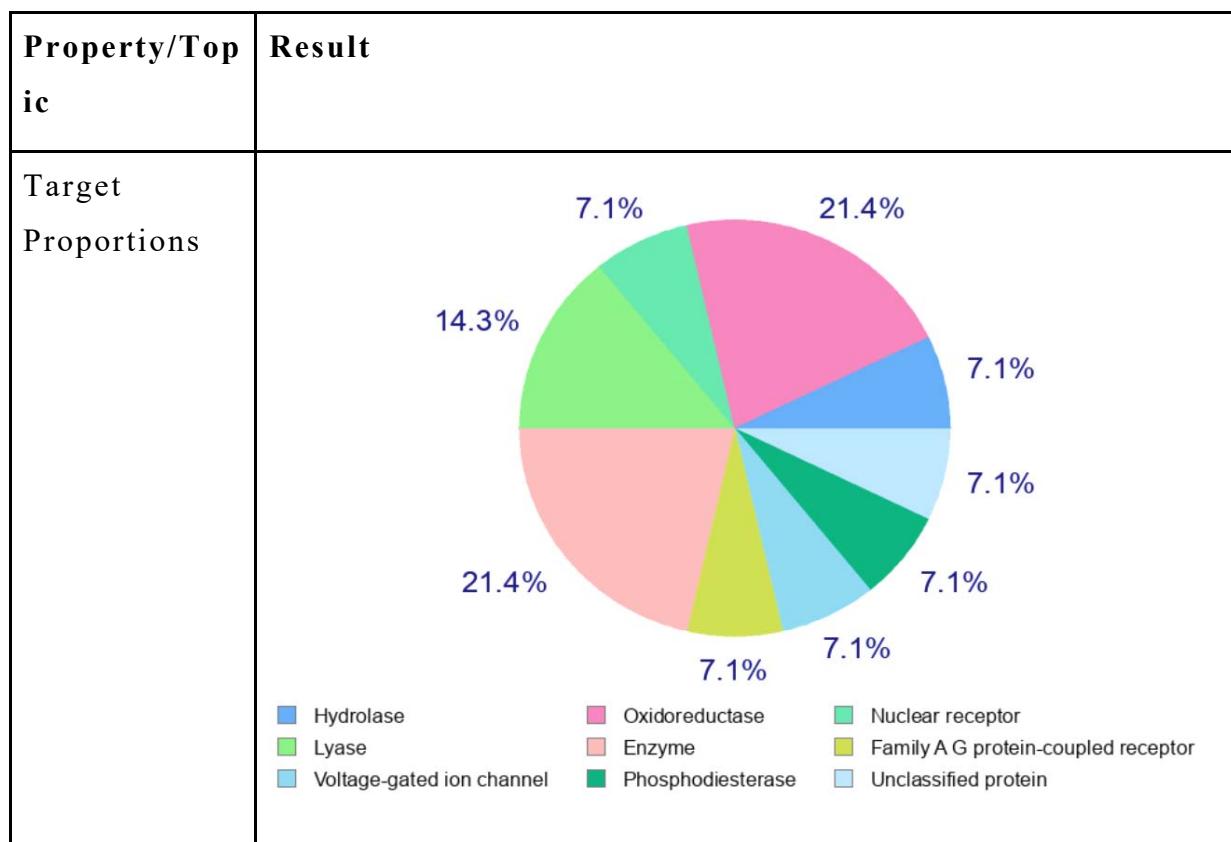
UniProtKB	Q72874 (Q72874_9HIV1)
Organism	Human immunodeficiency virus 1
Gene	pol
Protein	Submitted name: Pol polyprotein
Function	Only Molecular function stated
GO: Biological Function	Not Stated on UniProt
GO Molecular function -	<ul style="list-style-type: none"> • aspartic-type endopeptidase activity

4. AZD-1775 (Adavosert-ib)

Property/Topic	Result														
Target Proportions	<table border="1"> <tr> <td>Kinase</td> <td>Toll-like and IL-1 receptors</td> <td>Protease</td> </tr> <tr> <td>Other cytosolic protein</td> <td>Family A G protein-coupled receptor</td> <td></td> </tr> </table>	Kinase	Toll-like and IL-1 receptors	Protease	Other cytosolic protein	Family A G protein-coupled receptor									
Kinase	Toll-like and IL-1 receptors	Protease													
Other cytosolic protein	Family A G protein-coupled receptor														
Target	<table border="1"> <thead> <tr> <th>Target</th> <th>Common name</th> <th>Uniprot ID</th> <th>ChEMBL ID</th> <th>Target Class</th> <th>Probability*</th> <th>Known actives (3D/2D)</th> </tr> </thead> <tbody> <tr> <td>c-Jun N-terminal kinase 1</td> <td>MAPK8</td> <td>P45983</td> <td>CHEMBL2276</td> <td>Kinase</td> <td><div style="width: 100%;"><div style="width: 10%;"> </div></div></td> <td>56 / 0 ↘</td> </tr> </tbody> </table>	Target	Common name	Uniprot ID	ChEMBL ID	Target Class	Probability*	Known actives (3D/2D)	c-Jun N-terminal kinase 1	MAPK8	P45983	CHEMBL2276	Kinase	<div style="width: 100%;"><div style="width: 10%;"> </div></div>	56 / 0 ↘
Target	Common name	Uniprot ID	ChEMBL ID	Target Class	Probability*	Known actives (3D/2D)									
c-Jun N-terminal kinase 1	MAPK8	P45983	CHEMBL2276	Kinase	<div style="width: 100%;"><div style="width: 10%;"> </div></div>	56 / 0 ↘									
UniProtKB	P45983 (MK08_HUMAN)														
Organism	Homo sapiens (Humans)														
Gene	MAPK8														
Protein	Mitogen-activated protein kinase 8														
Function	<ul style="list-style-type: none"> Loss of this interaction abrogates the acetylation required for replication initiation. Promotes stressed cell apoptosis by phosphorylating key regulatory factors 														

	including p53/TP53 and Yes-associates protein YAP1 (PubMed:21364637).
GO: Biological Function	<ul style="list-style-type: none"> negative regulation of apoptotic process positive regulation of apoptotic process positive regulation of cell killing positive regulation of protein insertion into mitochondrial membrane involved in apoptotic signaling pathway

5. Triethyl phosphate



Target	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left; padding: 2px;">Target</th><th style="text-align: center; padding: 2px;">Common name</th><th style="text-align: center; padding: 2px;">Uniprot ID</th><th style="text-align: center; padding: 2px;">ChEMBL ID</th><th style="text-align: center; padding: 2px;">Target Class</th><th style="text-align: center; padding: 2px;">Probability*</th><th style="text-align: center; padding: 2px;">Known actives (3D/2D)</th></tr> </thead> <tbody> <tr> <td style="padding: 2px;">Acetylcholinesterase</td><td style="text-align: center; padding: 2px;">ACHE</td><td style="text-align: center; padding: 2px;">P22303</td><td style="text-align: center; padding: 2px;">CHEMBL220</td><td style="text-align: center; padding: 2px;">Hydrolase</td><td style="text-align: center; padding: 2px;"></td><td style="text-align: center; padding: 2px;">3 / 1 </td></tr> </tbody> </table>							Target	Common name	Uniprot ID	ChEMBL ID	Target Class	Probability*	Known actives (3D/2D)	Acetylcholinesterase	ACHE	P22303	CHEMBL220	Hydrolase		3 / 1
Target	Common name	Uniprot ID	ChEMBL ID	Target Class	Probability*	Known actives (3D/2D)															
Acetylcholinesterase	ACHE	P22303	CHEMBL220	Hydrolase		3 / 1															
UniProtKB	P22303 (ACES_HUMAN)																				
Organism	Homo sapiens (Human)																				
Gene	ACHE																				
Protein	Acetylcholinesterase																				
Function	<ul style="list-style-type: none"> • Hydrolyzes rapidly the acetylcholine neurotransmitter released into the synaptic cleft allowing to terminate the signal transduction at the neuromuscular junction. Role in neuronal apoptosis. 																				
GO: Biological Function	<ul style="list-style-type: none"> • synapse assembly 																				

6. Ganetespib

Property/Topic	Result														
Target Proportions	<p>A pie chart illustrating the distribution of target types. The segments are labeled with their respective percentages: Other cytosolic protein (33.3%), Enzyme (26.7%), Family A G protein-coupled receptor (20.0%), Kinase (13.3%), and Oxidoreductase (6.7%).</p> <table border="1"> <thead> <tr> <th>Target Class</th> <th>Percentage</th> </tr> </thead> <tbody> <tr> <td>Other cytosolic protein</td> <td>33.3%</td> </tr> <tr> <td>Enzyme</td> <td>26.7%</td> </tr> <tr> <td>Family A G protein-coupled receptor</td> <td>20.0%</td> </tr> <tr> <td>Kinase</td> <td>13.3%</td> </tr> <tr> <td>Oxidoreductase</td> <td>6.7%</td> </tr> </tbody> </table>	Target Class	Percentage	Other cytosolic protein	33.3%	Enzyme	26.7%	Family A G protein-coupled receptor	20.0%	Kinase	13.3%	Oxidoreductase	6.7%		
Target Class	Percentage														
Other cytosolic protein	33.3%														
Enzyme	26.7%														
Family A G protein-coupled receptor	20.0%														
Kinase	13.3%														
Oxidoreductase	6.7%														
Target	<table border="1"> <thead> <tr> <th>Target</th> <th>Common name</th> <th>Uniprot ID</th> <th>ChEMBL ID</th> <th>Target Class</th> <th>Probability*</th> <th>Known actives (3D/2D)</th> </tr> </thead> <tbody> <tr> <td>Heat shock protein 75 kDa, mitochondrial</td> <td>TRAP1</td> <td>Q12931</td> <td>CHEMBL1075132</td> <td>Other cytosolic protein</td> <td><div style="width: 100%;"><div style="width: 100%; background-color: green;"></div></div></td> <td>3 / 1 </td> </tr> </tbody> </table>	Target	Common name	Uniprot ID	ChEMBL ID	Target Class	Probability*	Known actives (3D/2D)	Heat shock protein 75 kDa, mitochondrial	TRAP1	Q12931	CHEMBL1075132	Other cytosolic protein	<div style="width: 100%;"><div style="width: 100%; background-color: green;"></div></div>	3 / 1
Target	Common name	Uniprot ID	ChEMBL ID	Target Class	Probability*	Known actives (3D/2D)									
Heat shock protein 75 kDa, mitochondrial	TRAP1	Q12931	CHEMBL1075132	Other cytosolic protein	<div style="width: 100%;"><div style="width: 100%; background-color: green;"></div></div>	3 / 1									
UniProtKB	Q12931 (TRAP1_HUMAN)														
Organism	Homo sapiens (Human)														
Gene	TRAP1														
Protein	Heat shock protein 75kDa, mitochondrial														
Function	<ul style="list-style-type: none"> Chaperone that expresses an ATPase activity. Involved in maintaining mitochondrial function and polarization, downstream of PINK1 and mitochondrial complex I. Is a 														

	negative regulator of mitochondrial respiration able to modulate the balance between oxidative phosphorylation and aerobic glycolysis. The impact of TRAP1 on mitochondrial respiration is probably mediated by modulation of mitochondrial SRC and inhibition of SDHA.
GO: Biological Function	<ul style="list-style-type: none"> negative regulation of intrinsic apoptotic signaling pathway in response to hydrogen peroxide

7. Vorinostat

Property/Topic	Result														
Target Proportions	<table border="1"> <tr> <td>Eraser</td> <td>Protease</td> <td>Ligand-gated ion channel</td> </tr> <tr> <td>86.7%</td> <td>6.7%</td> <td>6.7%</td> </tr> </table>	Eraser	Protease	Ligand-gated ion channel	86.7%	6.7%	6.7%								
Eraser	Protease	Ligand-gated ion channel													
86.7%	6.7%	6.7%													
Target	<table border="1"> <thead> <tr> <th>Target</th> <th>Common name</th> <th>Uniprot ID</th> <th>ChEMBL ID</th> <th>Target Class</th> <th>Probability*</th> <th>Known actives (3D/2D)</th> </tr> </thead> <tbody> <tr> <td>Histone deacetylase 3</td> <td>HDAC3</td> <td>O15379</td> <td>CHEMBL1829</td> <td>Eraser</td> <td><div style="width: 100%;"><div style="width: 100%; background-color: green;"></div></div> 270 / 85 🔍</td> <td></td> </tr> </tbody> </table>	Target	Common name	Uniprot ID	ChEMBL ID	Target Class	Probability*	Known actives (3D/2D)	Histone deacetylase 3	HDAC3	O15379	CHEMBL1829	Eraser	<div style="width: 100%;"><div style="width: 100%; background-color: green;"></div></div> 270 / 85 🔍	
Target	Common name	Uniprot ID	ChEMBL ID	Target Class	Probability*	Known actives (3D/2D)									
Histone deacetylase 3	HDAC3	O15379	CHEMBL1829	Eraser	<div style="width: 100%;"><div style="width: 100%; background-color: green;"></div></div> 270 / 85 🔍										

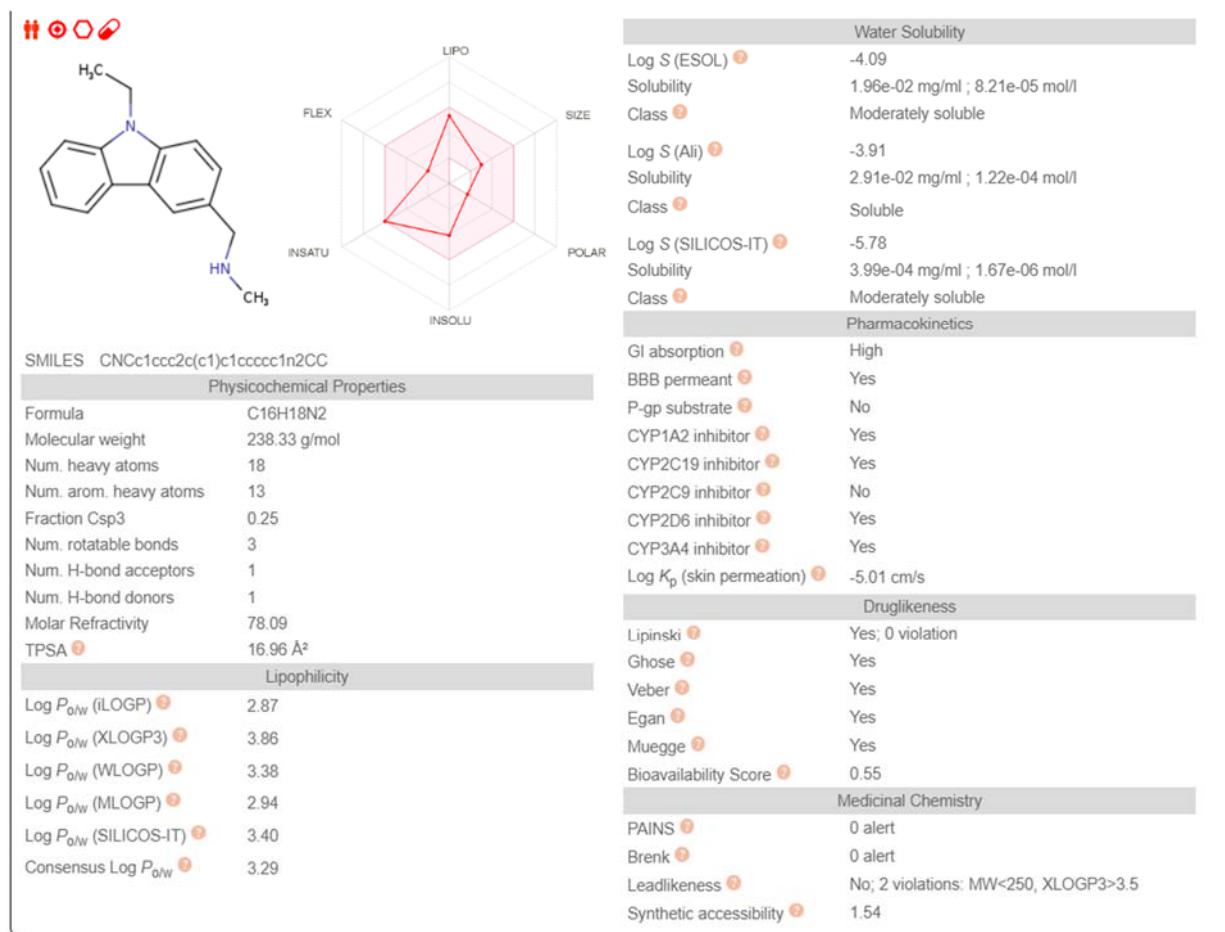
UniProtKB	O15379 (HDAC3_HUMAN)
Organism	Homo sapiens (Human)
Gene	HDAC3
Protein	Histone deacetylase 3
Function	<ul style="list-style-type: none"> • Histone deacetylation gives a tag for epigenetic repression and plays an important role in transcriptional regulation, cell cycle progression and developmental events (PubMed:23911289).
GO: Biological Function	<ul style="list-style-type: none"> • negative regulation of the apoptotic process

Swiss ADME

The important terms and factors that we have taken into consideration in Swiss ADME and their details are as follows:-

The figures in the tables with cobweb-like formations will represent the 'Bioavailability Radar' in the figures section (right-hand side image). The Bioavailability Radar provides an initial assessment of a molecule's drug-likeness.

1. 1-(9-ethyl-9H-carbazol-3-yl)-N-methylmethanamine or (PhiKan 083)



The LogS of the drug is soluble for all the categories (ESOL, Ali and SILICOS-IT). In the pharmacokinetic properties, BBB is ‘Yes’ and GI absorption is ‘High’ which is good for the drug. Also, the drug obeys Lipinski’s rule. But, there is a violation of Leadlikeness with 2 violations.

2. AZD-3355 (Lesogaberan)



The LogS of the drug is soluble for all the categories (ESOL, Ali and SILICOS-IT). In the pharmacokinetic properties, BBB is ‘No’ and GI absorption is ‘High’ which is good for the drug. Also, the drug obeys Lipinski’s rule. But, there is a violation of Leadlikeness with 1 violations.

3. APR-246 (Prima-1met)



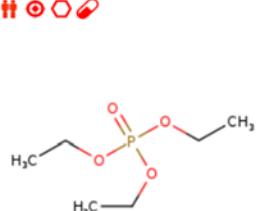
The LogS of the drug is soluble for all the categories (ESOL, Ali and SILICOS-IT). In the pharmacokinetic properties, BBB is ‘No’ and GI absorption is ‘High’ which is good for the drug. Also, the drug obeys Lipinski’s rule. But, there is a violation of Leadlikeness with 1 violation.

4. AZD-1775 (Adavosert-ib)

		Water Solubility
		Log S (ESOL) ⓘ -4.85
		Solubility 7.00e-03 mg/ml ; 1.40e-05 mol/l
		Class ⓘ Moderately soluble
		Log S (Ali) ⓘ -4.96
		Solubility 5.50e-03 mg/ml ; 1.10e-05 mol/l
		Class ⓘ Moderately soluble
		Log S (SILICOS-IT) ⓘ -6.22
		Solubility 2.98e-04 mg/ml ; 5.96e-07 mol/l
		Class ⓘ Poorly soluble
Physicochemical Properties		Pharmacokinetics
SMILES	C=CCn1c(=O)c2c(n1c1cccc(n1C(O)(C)Nc(nc2)Nc1ccc(cc1)N1CCN(CC1)C	GI absorption ⓘ High
Formula	C27H32N8O2	BBB permeant ⓘ No
Molecular weight	500.60 g/mol	P-gp substrate ⓘ Yes
Num. heavy atoms	37	CYP1A2 inhibitor ⓘ No
Num. arom. heavy atoms	21	CYP2C19 inhibitor ⓘ Yes
Fraction Csp3	0.33	CYP2C9 inhibitor ⓘ Yes
Num. rotatable bonds	7	CYP2D6 inhibitor ⓘ Yes
Num. H-bond acceptors	6	CYP3A4 inhibitor ⓘ Yes
Num. H-bond donors	2	Log K_p (skin permeation) ⓘ -7.15 cm/s
Molar Refractivity	152.75	
TPSA ⓘ	104.34 Å ²	
Lipophilicity		Druglikeness
Log $P_{o/w}$ (iLOGP) ⓘ	4.15	Lipinski ⓘ Yes; 1 violation: MW>500
Log $P_{o/w}$ (XLOGP3) ⓘ	3.10	Ghose ⓘ No; 2 violations: MW>480, MR>130
Log $P_{o/w}$ (WLOGP) ⓘ	2.02	Veber ⓘ Yes
Log $P_{o/w}$ (MLOGP) ⓘ	2.77	Egan ⓘ Yes
Log $P_{o/w}$ (SILICOS-IT) ⓘ	1.77	Muegge ⓘ Yes
Consensus Log $P_{o/w}$ ⓘ	2.76	Bioavailability Score ⓘ 0.55
		Medicinal Chemistry
		PAINS ⓘ 1 alert: anil_di_alk_A ⓘ
		Brenk ⓘ 1 alert: isolated_alkene ⓘ
		Leadlikeness ⓘ No; 1 violation: MW>350
		Synthetic accessibility ⓘ 4.23

The LogS of the drug is soluble for all the categories (ESOL, Ali and SILICOS-IT) but SILICOS-IT is poorly soluble. In the pharmacokinetic properties, BBB is ‘No’ and GI absorption is ‘High’ which is good for the drug. Also, the drug obeys Lipinski’s rule with 1 violation in terms of molecular weight. There is ‘No’ for leadlikeness with 1 violation.

5. Triethyl phosphate

		Water Solubility
		Log S (ESOL) ⓘ -1.08
		Solubility 1.52e+01 mg/ml ; 8.37e-02 mol/l
		Class ⓘ Very soluble
		Log S (Ali) ⓘ -1.53
		Solubility 5.41e+00 mg/ml ; 2.97e-02 mol/l
		Class ⓘ Very soluble
		Log S (SILICOS-IT) ⓘ -1.50
		Solubility 5.83e+00 mg/ml ; 3.20e-02 mol/l
		Class ⓘ Soluble
Pharmacokinetics		
		GI absorption ⓘ High
		BBB permeant ⓘ Yes
		P-gp substrate ⓘ No
		CYP1A2 inhibitor ⓘ No
		CYP2C19 inhibitor ⓘ No
		CYP2C9 inhibitor ⓘ No
		CYP2D6 inhibitor ⓘ No
		CYP3A4 inhibitor ⓘ No
		Log K_p (skin permeation) ⓘ -6.84 cm/s
Lipophilicity		Druglikeness
Log $P_{o/w}$ (iLOGP) ⓘ 2.69		Lipinski ⓘ Yes; 0 violation
Log $P_{o/w}$ (XLOGP3) ⓘ 0.80		Ghose ⓘ Yes
Log $P_{o/w}$ (WLOGP) ⓘ 2.20		Veber ⓘ Yes
Log $P_{o/w}$ (MLOGP) ⓘ 0.35		Egan ⓘ Yes
Log $P_{o/w}$ (SILICOS-IT) ⓘ 0.78		Muegge ⓘ No; 1 violation: MW<200
Consensus Log $P_{o/w}$ ⓘ 1.37		Bioavailability Score ⓘ 0.55
		Medicinal Chemistry
		PAINS ⓘ 0 alert
		Brenk ⓘ 1 alert: phosphor ⓘ
		Leadlikeness ⓘ No; 1 violation: MW<250
		Synthetic accessibility ⓘ 3.84

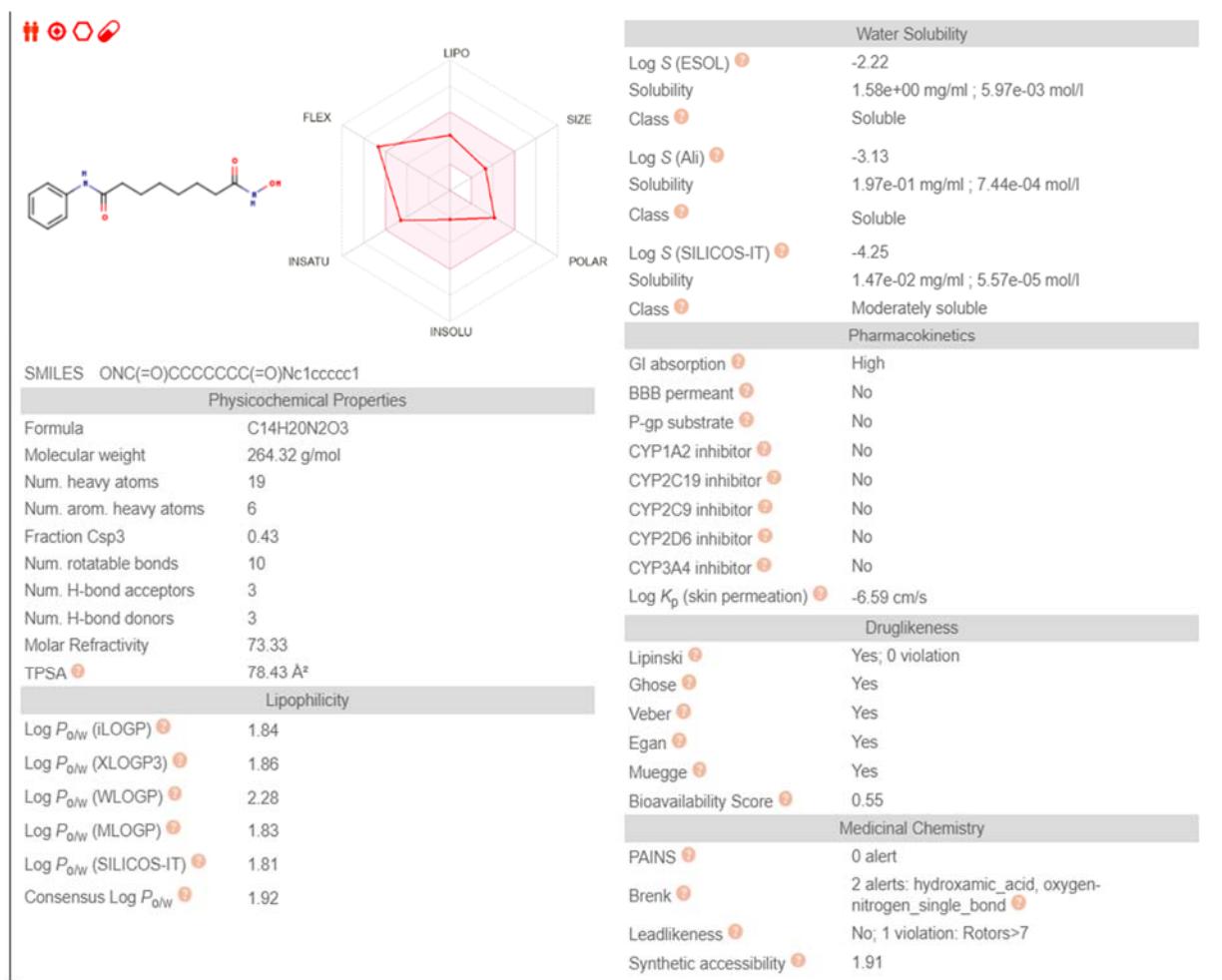
The LogS of the drug is soluble for all the categories (ESOL, Ali and SILICOS-IT). In the pharmacokinetic properties, BBB is ‘Yes’ and GI absorption is ‘High’ which is good for the drug. Also, the drug obeys Lipinski’s rule. There is ‘No’ for leadlikeness with 1 violation.

6. Ganetespib

		Water Solubility
		Log S (ESOL) ⓘ -4.40
		Solubility Class ⓘ Moderately soluble
		Log S (Ali) ⓘ -4.78
		Solubility Class ⓘ Moderately soluble
		Log S (SILICOS-IT) ⓘ -5.05
		Solubility Class ⓘ Moderately soluble
Physicochemical Properties		Pharmacokinetics
SMILES Oc1cc(O)c(cc1c1n[nH]c(=O)n1ccc2c(c1)ccn2C)C(C)C		GI absorption ⓘ High
Formula C20H20N4O3		BBB permeant ⓘ No
Molecular weight 364.40 g/mol		P-gp substrate ⓘ No
Num. heavy atoms 27		CYP1A2 inhibitor ⓘ No
Num. arom. heavy atoms 20		CYP2C19 inhibitor ⓘ No
Fraction Csp3 0.20		CYP2C9 inhibitor ⓘ Yes
Num. rotatable bonds 3		CYP2D6 inhibitor ⓘ No
Num. H-bond acceptors 4		CYP3A4 inhibitor ⓘ No
Num. H-bond donors 3		Log K_p (skin permeation) ⓘ -6.33 cm/s
Molar Refractivity 105.00		Druglikeness
TPSA ⓘ 96.07 Å ²		Lipinski ⓘ Yes; 0 violation
	Lipophilicity	Ghose ⓘ Yes
Log $P_{o/w}$ (iLOGP) ⓘ 2.50		Veber ⓘ Yes
Log $P_{o/w}$ (XLOGP3) ⓘ 3.09		Egan ⓘ Yes
Log $P_{o/w}$ (WLOGP) ⓘ 3.25		Muegge ⓘ Yes
Log $P_{o/w}$ (MLOGP) ⓘ 2.66		Bioavailability Score ⓘ 0.55
Log $P_{o/w}$ (SILICOS-IT) ⓘ 2.69		Medicinal Chemistry
Consensus Log $P_{o/w}$ ⓘ 2.84		PAINS ⓘ 0 alert
		Brenk ⓘ 0 alert
		Leadlikeness ⓘ No; 1 violation: MW>350
		Synthetic accessibility ⓘ 3.10

The LogS of the drug is soluble for all the categories (ESOL, Ali and SILICOS-IT). In the pharmacokinetic properties, BBB is ‘No’ and GI absorption is ‘High’ which is good for the drug. Also, the drug obeys Lipinski’s rule. There is ‘No’ for leadlikeness with 1 violation.

7. Vorinostat



The LogS of the drug is soluble for all the categories (ESOL, Ali and SILICOS-IT). In the pharmacokinetic properties, BBB is ‘No’ and GI absorption is ‘High’ which is good for the drug. Also, the drug obeys Lipinski’s rule. There is ‘No’ for leadlikeness with 1 violation.

PreADME (Toxicity)

Amongst PreADME usage results, the toxicity was the out main area of concern for result generation, hence the results for toxicity of out chosen drugs in different test configurations are listed in the tables below by using PreADME.

1. 1-(9-ethyl-9H-carbazol-3-yl)-N-methylmethanamine or (PhiKan 083)

ID	Value
algae_at	0.0446127
Ames_test	mutagen
Carcino_Mouse	positive
Carcino_Rat	positive
daphnia_at	0.0641795
hERG_inhibition	medium_risk
medaka_at	0.00724942
minnow_at	0.016777
TA100_10RLI	negative
TA100_NA	positive

2. AZD-3355 (Lesogaberan)

ID	Value
algae_at	1.31967
Ames_test	mutagen
Carcino_Mouse	positive
Carcino_Rat	positive
daphnia_at	136.94
hERG_inhibition	low_risk
medaka_at	13482.5
minnow_at	2233.66
TA100_10RLI	negative
TA100_NA	positive

3. APR-246 (Prima-1met)

ID	Value
algae_at	1.18375
Ames_test	mutagen
Carcino_Mouse	positive
Carcino_Rat	positive

daphnia_at	6.02297
hERG_inhibition	low_risk
medaka_at	33.7906
minnow_at	18.5551
TA100_10RLI	negative
TA100_NA	negative

4. AZD-1775 (Adavosert-ib)

ID	Value
algae_at	0.0052455
Ames_test	non-mutagen
Carcino_Mouse	negative
Carcino_Rat	negative
daphnia_at	0.00591148
hERG_inhibition	medium_risk
medaka_at	0.00010446
minnow_at	0.000263351
TA100_10RLI	negative
TA100_NA	negative

5. Triethyl phosphate

ID	Value
algae_at	0.0822102
Ames_test	mutagen
Carcino_Mouse	negative
Carcino_Rat	positive
daphnia_at	1.69532
hERG_inhibition	low_risk
medaka_at	2.65577
minnow_at	0.566376
TA100_10RLI	negative
TA100_NA	negative

6. Ganetespib

ID	Value
algae_at	0.0109912
Ames_test	mutagen
Carcino_Mouse	negative
Carcino_Rat	negative

daphnia_at	0.0160637
hERG_inhibition	medium_risk
medaka_at	0.000605989
minnow_at	0.00181303
TA100_10RLI	negative
TA100_NA	negative

7. Vorinostat

ID	Value
algae_at	0.0502196
Ames_test	mutagen
Carcino_Mouse	negative
Carcino_Rat	negative
daphnia_at	0.324618
hERG_inhibition	low_risk
medaka_at	0.148241
minnow_at	0.175024
TA100_10RLI	positive
TA100_NA	negative

Molinspiration Tool

Predict Bioactivity of the following sample with Molinspiration

Molinspiration tools was used to predict the bioactivity of the all the organic compounds (drugs) of our choosing. Different score against each property can be deduced from the Molinspiration tool.

Molinspiration bioactivity scores:-

1. 1-(9-ethyl-9H-carbazol-3-yl)-N-methylmethanamine or (PhiKan 083)

Property	Score
GPCR ligand	-0.03
Ion channel modulator	-0.16
Kinase inhibitor	0.12
Nuclear receptor ligand	-0.39
Protease inhibitor	-0.41
Enzyme inhibitor	0.07

2. AZD-3355 (Lesogaberan)

Property	Score
GPCR ligand	-1.90
Ion channel modulator	-1.94
Kinase inhibitor	-2.40

Nuclear receptor ligand	-2.70
Protease inhibitor	-2.06
Enzyme inhibitor	-1.61

3. APR-246 (Prima-1met)

Property	Score
GPCR ligand	-0.03
Ion channel modulator	-0.27
Kinase inhibitor	-0.76
Nuclear receptor ligand	-1.00
Protease inhibitor	-0.31
Enzyme inhibitor	-0.17

4. AZD-1775 (Adavosert-ib)

Property	Score
GPCR ligand	0.23
Ion channel modulator	0.01
Kinase inhibitor	0.34

Nuclear receptor ligand	-0.49
Protease inhibitor	-0.21
Enzyme inhibitor	0.24

5. Triethyl phosphate

Property	Score
GPCR ligand	-0.34
Ion channel modulator	0.28
Kinase inhibitor	-0.43
Nuclear receptor ligand	-0.71
Protease inhibitor	-0.33
Enzyme inhibitor	0.38

6. Ganetespib

Property	Score
GPCR ligand	0.09
Ion channel modulator	-0.12
Kinase inhibitor	0.10

Nuclear receptor ligand	-0.02
Protease inhibitor	-0.32
Enzyme inhibitor	-0.03

7. Vorinostat

Property	Score
GPCR ligand	-0.07
Ion channel modulator	-0.30
Kinase inhibitor	0.03
Nuclear receptor ligand	-0.39
Protease inhibitor	0.42
Enzyme inhibitor	0.40

PyRx

PyRx was one of the main tools used in our studies. It was extensively used in order to find Binding Affinity which is the strength of the binding interaction between a single biomolecule (e.g. protein or DNA) to its ligand/binding partner (e.g. drug or inhibitor). More negative the value of binding affinity is, the better the binding affinity gets which means that ligand will bind more effectively with the target with less energy expenditure and better confirmation and stability. (Table 6).

Ligand	Binding Affinity
1._PhiKan083_uff_E=436.14	-5.7
2._AZD-3355_(Lesogaberan)_uff_E=-140.23	-3.3
3._APR-246_(Prima-1met)_uff_E=404.49	-4.6
4._AZD-1775_(Advosert-ib)_uff_E=869.78	-7.7
5._Triethyl_phosphate_uff_E=409.38	-3.8
6._Gantespib_uff_E=808.15	-7.4
7._Vorinostat_uff_E=163.71	-5.5

Table 6, Docking Results of 2J1X after preparation with are chosen ligands (drugs).

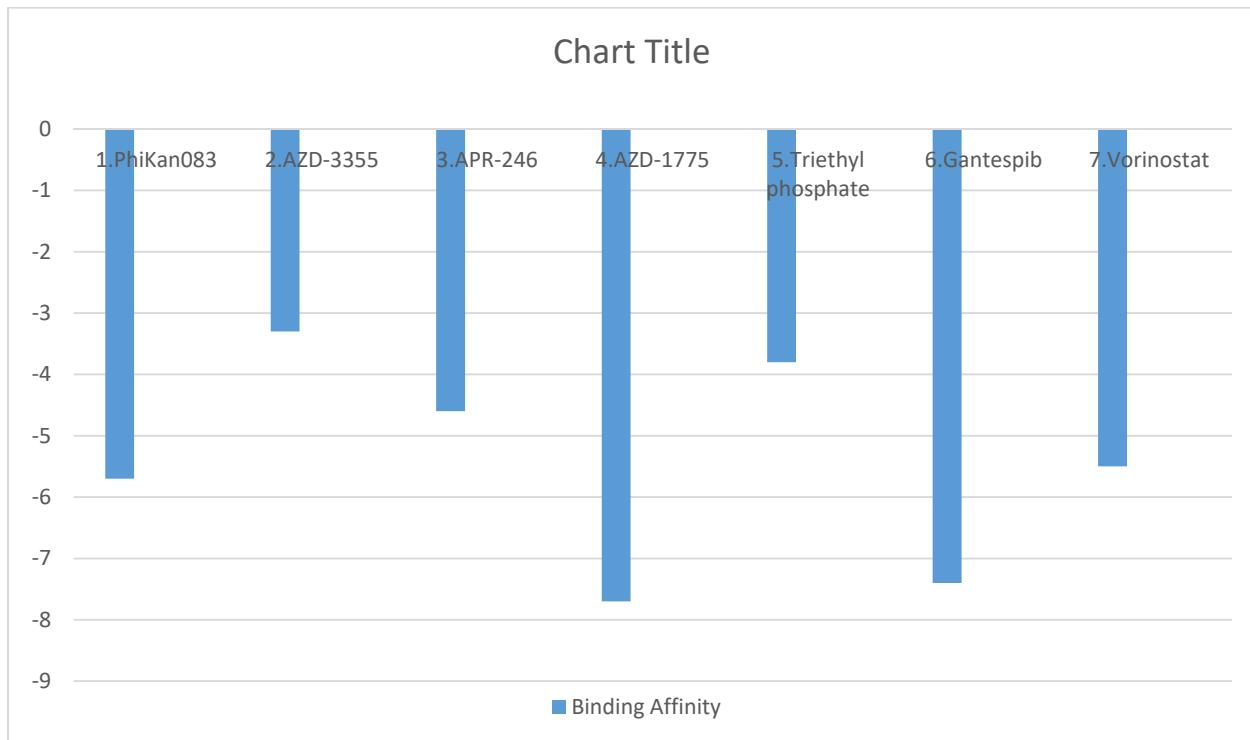


Fig. 44, The AZD-1775 have the highest binding affinity among all the other ligands (drugs) as well as Ganetespib is close by with negligible difference.

Since, the most negative value of the binding affinity is for the AZD – 1775 (Table 6) and the comparison can be clearly visible in the bar chart (Fig. 44) where besides AZD-1775 with highest binding affinity is also almost competed by Gantespib falling minutely behind it in binding affinity.



Fig. 45, The protein above is the prepared 2J1X protein structure in Biovia Discovery Studio.

The cellular tumour protein was cleaned using the biovia discovery studio before it could be used for docking of the ligands in the PyRx. Cleaned structure had a chain removed and water molecules eliminated from it. (Fig. 45).

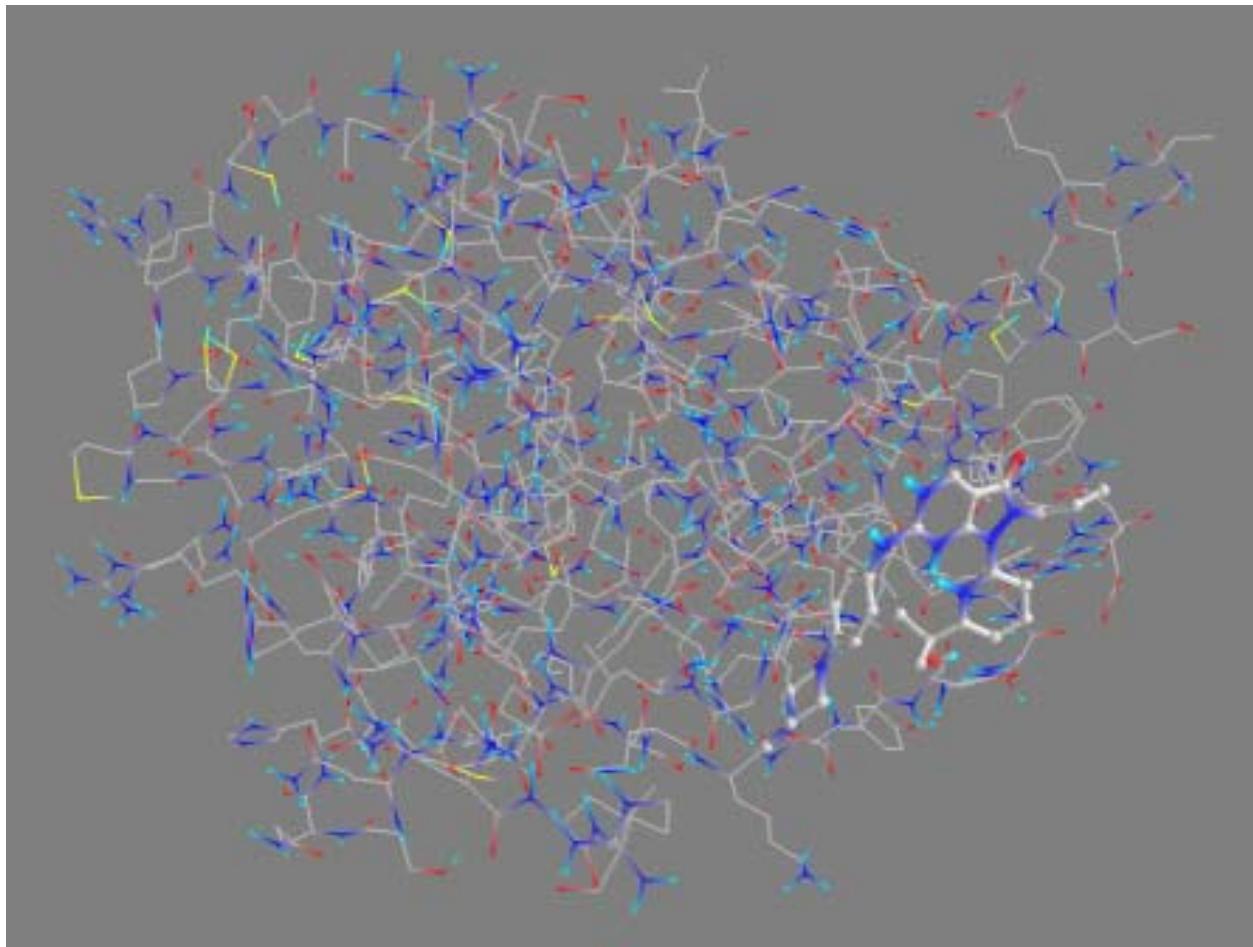


Fig. 46, In PyRx, this is the 3D structure of docked 2J1X protein.

Here the thicker structure at the right hand corner of the figure represents the ligand (drug) AZD-1775 which is in the docked position to the 2J1X protein and is showing the best binding affinity (Fig. 46)

Biovia Discovery Studio

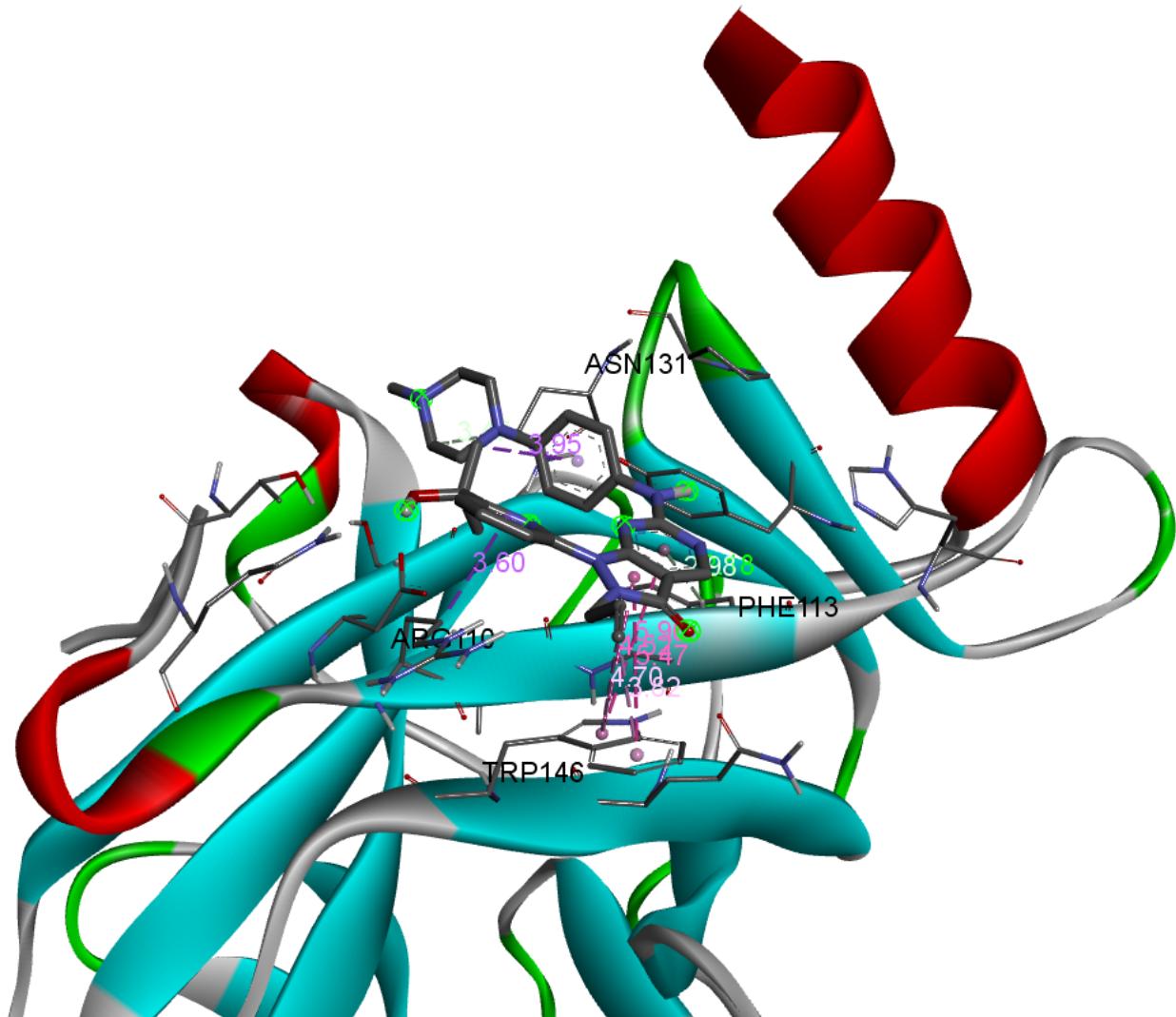


Fig. 47, Cleaned cellular tumour protein p53 in a docked confirmation with ligand AZD-1775.

Thick dark grey stick like structure (ligand) can be seen in the binded state at the p53 protein which is in beta sheet and alpha helix structure. Binding sites ARC110, TRP146, ASN131 and PHE113 are visible (Fig. 47). These sites are where the ligand had bind the p53 protein in the best possible confirmation as per Biovia's simulation.

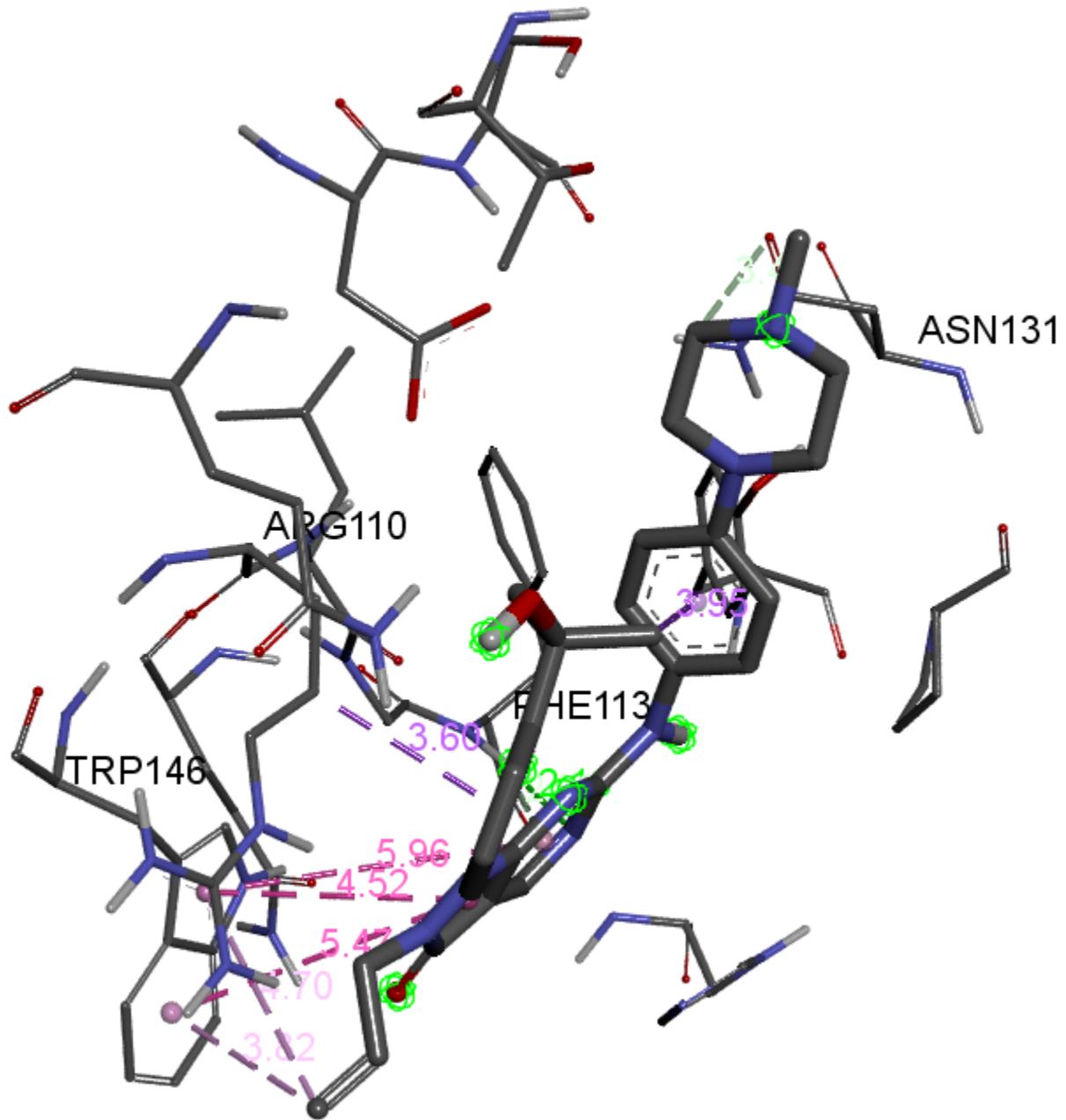
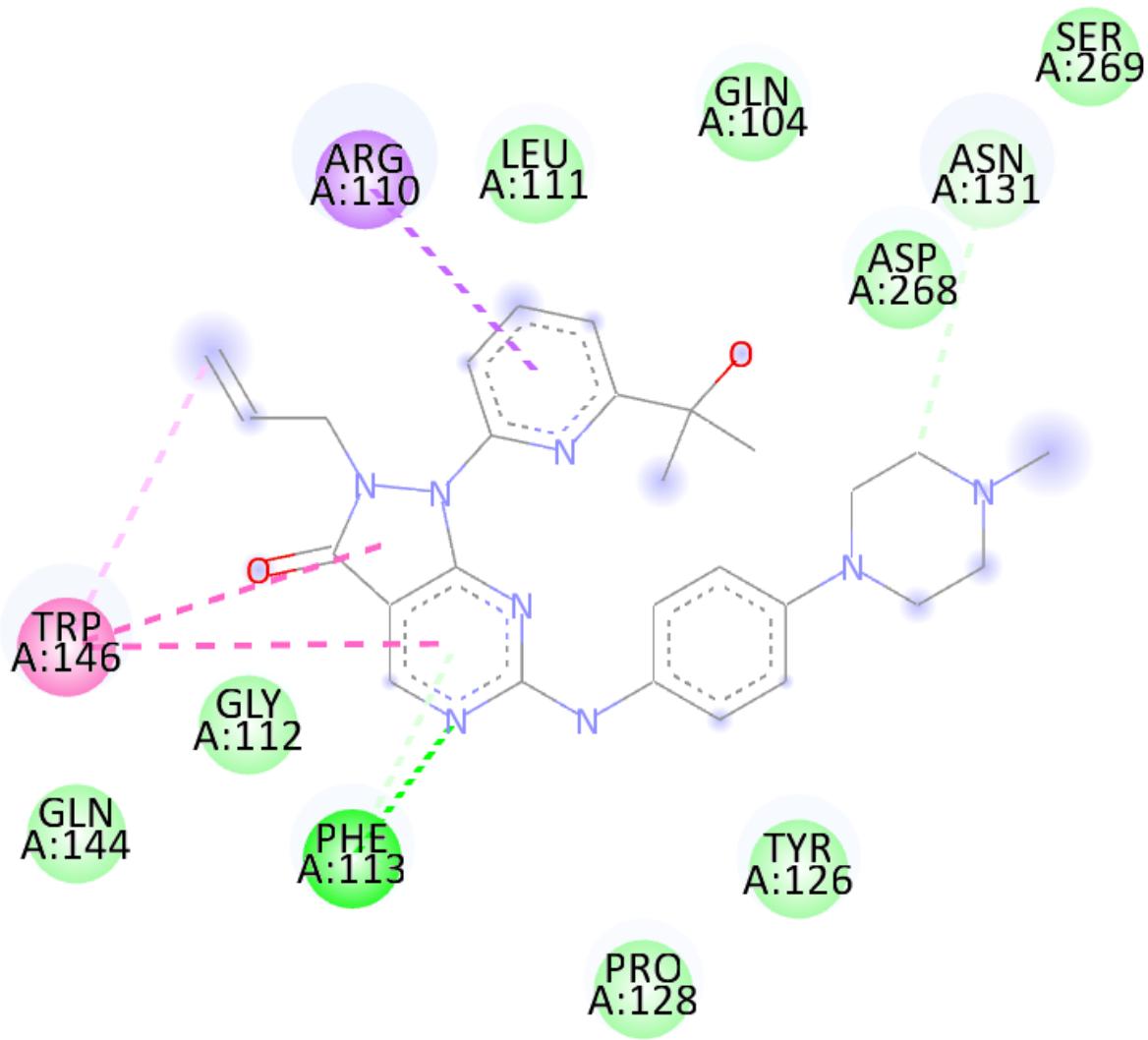


Fig. 48, Docked 2J1X protein's site. Here interactions sites with ligand are again visible like the previous figure.

Here the p53 protein is also in a stick structure but in a thinner texture compared to thicker p53 protein. The pink lines signifies the distance between the ligand (AZD-1775) and the target (p53) (Fig. 48).



Interactions

- | | |
|--|---------------------------|
| [Green Box] van der Waals | [Purple Box] Pi-Sigma |
| [Green Box] Conventional Hydrogen Bond | [Pink Box] Pi-Pi Stacked |
| [Green Box] Carbon Hydrogen Bond | [Light Blue Box] Pi-Alkyl |
| [Green Box] Pi-Donor Hydrogen Bond | |

Fig. 49, This is a 2D – Structure of the same interaction as visible in the previous figures.

Heres, the 2D increases the viewing capacity of the interacting atoms and amino acids, bonding interactions (such as van der waals, hydrogen bonds, etc.,). Also, Pi-sigma, Pi-Pi Stacked and Pi-Alkyl which are electronic confirmations of bonds (Fig. 49).

Discussion

Different ligands were used to test their suitability to suppress mutant p53 gene's effect which would otherwise cause cancer. In our studies we took drugs from different research papers as well from the DrugBank which were: 1-(9-ethyl-9H-carbazol-3-yl)-N-methylmethanamine (PhiKan 083), AZD 3355 (Lesogaberan), APR-246 (Prima-1met), AZD 1775 (Adavosertib), Triethyl phosphate, Ganetespib and Vorinostat where are all a virtual collection done by us of commercially available screenable compounds.

Upon doing ADME analysis using, it was found that most of the ligands gave 'Yes' for Blood Brain Barrier' and 'GI absorption'. Also, Lipsinski's rules was obeyed by all the ligands with 1 or 2 violations max by some ligands. String database analysis showed that MDM2 was showed the most coexpression with our target TP53.

SWISS Target Prediction was used in order to get find the different targets for all our ligands. The targets with most proportions were chosen and upon opening the targets data with highest proportion, most of the ligands showed G.O. Biological functions such as apoptosis, cell cycle arrest, cell death and relation to apoptotic pathways which clearly suggested their relevance to our study of mutant p53 which is carcinogenic.

PyRx was used for the purpose of docking all the ligands to the target protein. AZD-3355 docked to the amino acid TRP, ARG, PHE and ASN on positions 146, 110, 113 and 131 respectively of the target protein 2J1X. It was deduced upon docking all the ligands that AZD-1775 (DrugBank Accession Number: DB11740) had the best binding with the ligand with Binding Affinity of -7.7 and Ganetespib with a negligible decrease of -0.3. Least effective drug against our target was AZD-3355 with only -3.3 binding affinity.

Conclusion

The development of small-molecule compounds and short peptides to restore the conformation and transcriptional activity of wild-type p53 to mutp53 is one of the most widely explored strategies. Due of the relative undruggability of mutp53 with varied thermostability or conformational configurations, this technique is difficult to implement.

Site of mutation in the oncogenic mutant p53 is a druggable target. Our studies showed the final result for use of small-molecule compounds or organic compounds against mutant p53 that AZD-1775 is the best compound in our table (Table 3) of selected drugs for the use in the suppression of cellular tumor protein p53 as it has the highest binding affinity to the target.

The selectivity, efficacy, and safety of these promising techniques to target mutp53-expressing human malignancies should be improved in the future. It's critical to learn more about how mutant p53 causes oncogenic function, figure out how mutant p53 is stabilised or degraded in tumours, and find mutant p53-specific downstream signalling pathways or binding partners.

Significance

Research is an activity that leads to the discovery of new facts and information, as well as supporting us in verifying current knowledge and prompting us to question things that are difficult to comprehend based on available evidence. To be a good manager, you must understand how to make the best judgments possible by understanding the numerous phases involved in solving problems.

A gene that makes a protein that is found inside the nucleus of cells and plays a key role in controlling cell division and cell death. Mutations (changes) in the p53 gene may cause cancer cells to grow and spread in the body.

It is important to research about the p53 mutation and stop it as its oncogenic nature is harmful to Homo sapiens and other mammals as well. Study of relevant drugs/organic compounds as a ligand for the purpose of stopping this p53 mutation and restore normalcy is necessary or the proper functioning of the body.

References

- "Cancer – Signs and symptoms". NHS Choices. (8 June 2014)
<https://www.nhs.uk/conditions/cancer/symptoms/> [accessed: 01/01/2022]
- "Cancer". World Health Organization. (12 September 2018).
https://www.who.int/health-topics/cancer#tab=tab_1 [accessed: 07/01/2022]
- "Defining Cancer". National Cancer Institute. (17 September 2007).
<https://www.cancer.gov/about-cancer/understanding/what-is-cancer> [accessed: 07/01/2022]
- "PubChem Source Information". The PubChem Project. USA: National Center for Biotechnology Information.
<https://pubchem.ncbi.nlm.nih.gov/sources> [accessed: 27/04/2022]
- "Research Institute Posts Gene Data on Internet". The New York Times. (1997).
<https://archive.nytimes.com/www.nytimes.com/library/cyber/week/062697gene.html> [accessed: 01/02/2022]
- "The Human Genome Project". The New York Times. (2005)
<https://archive.nytimes.com/www.nytimes.com/library/national/science/genome-index.html> [accessed: 01/02/2022]
- 'How are the scores computed?', String (2022)
<http://version10.string-db.org/help/faq/> [accessed: 30/03/2022]
- "File Formats and the PDB".
[www.pdb.org.](http://www.pdb.org) [accessed: 25/02/22]
- "Genetics versus genomics"
<https://web.archive.org/web/20040630170807/http://www.who.int/genomics/geneticsVSGenomics/en/> [accessed: 01/02/2022]
- "The Genetics of Cancer" (2018)
<https://www.cancer.net/navigating-cancer-care/cancer-basics/genetics/genetics-cancer> [accessed: 07/01/2022]
- "What is Cancer?" (2019)
<https://www.cancer.net/navigating-cancer-care/cancer-basics/what-cancer> [accessed: 01/01/2022]

- Anshul Tiwari, Sakshi Singh, Computational approaches in drug designing in Bioinformatics. Chapter 3. (2022).
- Atema, A. & Chene, P. The gain of function of the p53 mutant Asp281Gly is dependent on its ability to form tetramers. *Cancer Lett.* 185, 103–109 (2002).
- Bajorath J. Computer-aided drug discovery [version 1; peer review: 3 approved]. F1000Research 2015, 4(F1000 Faculty Rev):630
- Balint E E, Vousden KH., Activation and activities of the p53 tumor suppressor protein. *Br J Cancer* 85:1813-23 (2001)
- Basic Local Alignment Search Tool
<https://blast.ncbi.nlm.nih.gov/Blast.cgi> [accessed 17/1/2022].
- Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (July 2003). "Multiple sequence alignment with the Clustal series of programs". *Nucleic Acids Research*. 31 (13): 3497–500.
- Cohen S, Murphy ML, Prather AA (January 2019). "Ten Surprising Facts About Stressful Life Events and Disease Risk". *Annual Review of Psychology*. 70: 577–597.
- Dassault Systemes (2022)
<https://discover.3ds.com/discovery-studio-visualizer-download>[accessed: 07/06/22]
- DrugBank Homepage (2022)
<https://go.drugbank.com/> [accessed 14/04/2022]
- EMBL-EBI. "Current Release Statistics < Uniprot < EMBL-EBI". www.ebi.ac.uk. Archived from the original on 25 April 2019.
[accessed: 01/03/22]
- Expasy, ProtParam tool (2022)
<https://web.expasy.org/protparam> [accessed: 17/03/22].
- Gfeller, D.; Grosdidier, A.; Wirth, M.; Daina, A.; Michelin, O.; Zoete, V. (2014). SwissTargetPrediction: a web server for target prediction of bioactive small molecules. *Nucleic Acids Research*, 42(W1), W32–W38.

- Hainaut P, Hollstein M. p53 and human cancer: the first ten thousand mutations. *Adv Cancer Res.* 2000;77:81-137 [PubMed] [Google Scholar] [Ref list]]
- Hartwell, L.; Kastan, M. (1994). Cell cycle control and cancer. , 266(5192), 1821–1828.
- Kanehisa M, Goto S (2000). "KEGG: Kyoto Encyclopedia of Genes and Genomes". *Nucleic Acids Res.* 28 (1): 27–30.
- Kim, Sunghwan; Chen, Jie; Cheng, Tiejun; Gindulyte, Asta; He, Jia; He, Siqian; Li, Qingliang; Shoemaker, Benjamin A; Thiessen, Paul A; Yu, Bo; Zaslavsky, Leonid; Zhang, Jian; Bolton, Evan E (8 January 2021). "PubChem in 2021: new data content and improved web interfaces". *Nucleic Acids Research.* 49 (D1): D1388–D1395.
- Koonin EV (March 2001). "Computational genomics". *Current Biology.* 11 (5): R155–8.
- Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (March 2001). "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings". *Advanced Drug Delivery Reviews.* 46 (1–3): 3–26.
- Mering, C. v. (2003). STRING: a database of predicted functional associations between proteins. *Nucleic Acids Research*, 31(1), 258–261.
- Multiple Sequence Alignment (2022)
<https://www.ebi.ac.uk/Tools/msa/clustalo/> [Accessed 21/01/2022]
- NCBI, Welcome to NCBI (2022)
<https://www.ncbi.nlm.nih.gov/> [Accessed 01/01/2022].
- Nicole Zache; Jeremy M.R. Lambert; Nina Rökaeus; Jinfeng Shen; Pierre Hainaut; Jan Bergman; Klas G. Wiman; Vladimir J.N. Bykov (2008). *Mutant p53 targeting by the low molecular weight compound STIMA-1.* , 2(1), 70–80.
- Oehmen, C.; Nieplocha, J. (2006). "ScalaBLAST: A Scalable Implementation of BLAST for High-Performance Data-Intensive Bioinformatics Analysis". *IEEE Transactions on Parallel and Distributed Systems.* 17 (8): 740.

- Open Reading Frame Finder (2022)
<https://www.ncbi.nlm.nih.gov/orffinder/> [Accessed 15/02/2022]
- PDB Database, “A Structural View of Biology” (2022)
<https://www.rcsb.org/> [Accessed 25/02/2022]
- Pearson WR, Wood T, Zhang Z, Miller W (November 1997). "Comparison of DNA sequences with protein sequences". *Genomics*. 46 (1): 24–36.
- Polinsky, Alex (2008). The Practice of Medicinal Chemistry Volume 112. Lead-Likeness and Drug-Likeness. , 244–254.
- Rui P S Patrício, Paula A Videira, Florbela Pereira, A computer-aided drug design approach to discover tumour suppressor p53 protein activators for colorectal cancer therapy, PMID: 34861473
- Singh J, Chuaqui CE, Boriack-Sjodin PA, Lee WC, Pontz T, Corbley MJ, Cheung HK, Arduini RM, Mead JN, Newman MN, Papadatos JL, Bowes S, Josiah S, Ling LE (Dec 2003). "Successful shape-based virtual screening: the discovery of a potent inhibitor of the type I TGFbeta receptor kinase (TbetaRI)". *Bioorganic & Medicinal Chemistry Letters*. 13 (24): 4355–9.
- Thomas Engel (2006). "Basic Overview of Chemoinformatics". *J. Chem. Inf. Model.* 46 (6): 2267–77.
- TP53 gene tumor protein p53 (2022)
<https://medlineplus.gov/genetics/gene/tp53/> [accessed: 25/02/22]
- Uniprot Home, (2022)
<https://www.uniprot.org/> [accessed: 01/03/2022]
- Uniprot, C. (2009). "The Universal Protein Resource (UniProt) in 2010". *Nucleic Acids Research*. 38 (Database issue): D142–D148.
- Vousden, K., Lane, D. p53 in health and disease. *Nat Rev Mol Cell Biol* 8, 275–283 (2007).
- Welcome to the PreADMET, Home (2022)
<https://preadmet.webservice.bmdrc.org/>, [accessed: 31/05/2022].
- Welcome to the PyRx Website, Home, Introduction
<https://pyrx.sourceforge.io/>, [accessed: 03/06/2022]

- Wishart, DS; Knox C; Guo AC; et al. (Jan 2006). "DrugBank: a comprehensive resource for in silico drug discovery and exploration". Nucleic Acids Research. 34 (Database issue): D668-72.