VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

IN SILICO DRUG DESIGNING

BIOLOGICAL DATABASE

BIT- 2002

WINTER SEM 2019-20

J COMPONENT – REVIEW-III

TEAM MEMBERS:

SUNEHA GHOSH	18BCB0075
SIDDHARTHA MONDAL	18BCB0145

PROJECT UNDER THE GUIDANCE OF:

Dr. SUDANDIRA DOSS

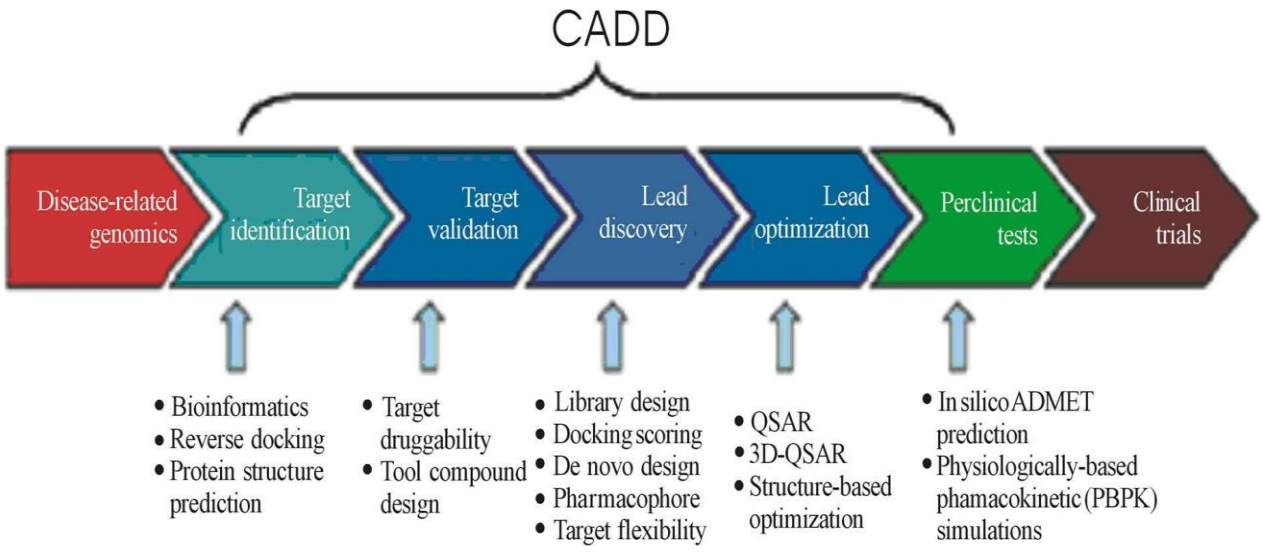
1. ABSTRACT

In the field of computational drug design, the identification and characterisation of the biological target of intrigue is a significant advance. In spite of the developing number of such settled protein structures each year, there are as yet many drug targets, especially membrane proteins, for which auxiliary assurance is very difficult. In these cases, experimental information on effectively decided bioactive molecules might be utilized effectively for computational ligand based drug design methods. However, for as far back as three decades, most medication revelation endeavours have been driven by the structure of the target biomolecule. Advances in structural biology techniques have given basic data of numerous molecules, offering ascend to the structure based drug design process as an integral asset for drug discovery in look into the scholarly community and pharmaceutical industry. Both fields in the in-silico drug design field are regularly utilized, every one relying upon foundation test data and important computational approaches.

The subject of in silico drug design is a rapidly growing place wherein many successes have befallen in current years. The explosion of bioinformatics, cheminformatics, genomics, proteomics, and
structural records have furnished hundreds of latest targets in addition to new ligands. Therefore, in silico drug design represents computational techniques and resources that are used to facilitate the possibilities for future drug lead discovery. This review stated a brief records of drug design and summarized the maximum important steps of in silico drug design approach for the discovery of latest molecular entities. The workflow of the entire virtual designing marketing campaign is discussed, from the selection of a target, the evaluation of a structure of that target,
ligand search, receptor theory to molecular docking, digital high-throughput screening, the pivotal questions to don't forget in choosing a way for drug lead discovery and evaluation of the drug leads.

2. INTRODUCTION

When it involves rational drug design, the identification and characterisation of the organic target of interest is a primary step. For the past 3 decades, drug discovery has been driven with the aid of the shape of the goal biomolecule. Advances in spectroscopic techniques such as X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy supplied 3D data of many molecules, giving rise to the shape-based totally drug-design (SBDD) manner as a powerful tool for drug discovery in studies academia and pharmaceutical enterprise SBDD affords insights into the interplay of a particular tar-get protein with numerous small molecules, yielding information about specificity and affinity of the interplay. This drug-design approach aims to hit compound identity as well as medicinal chemistry optimisation. In case that the shape of the goal protein is unknown, ligand-based totally computer-aided drug-design techniques are used in order to design molecules in silico. Such design is based totally on records of active or inactive compounds, which already exist and are recognized that engage with a biomolecular target. The goal of this technique is to keep the crucial structural and physicochemical properties of the ligand for its interplay with the molecular target even as at the equal time discard the facts which isn't always related to those interactions. In this article, we strive to give an overview of the most typically used strategies and provide examples of widely used software program for in silico drug design.



IN-SILICO DRUG DESIGN

In silico is a time period that method "computer aided".
So in silico drug design way rational design with the aid of which tablets are designed/discovered by using the use of computational techniques. According to Kubinyi , most of the medicine within the past were located via twist of fate or trial and error method, or in different words, serendipity performed an important role in locating new capsules.
Current fashion in drug discovery is shifted from discovery to design, which needs expertise the biochemistry of the sickness, pathways, figuring out disorder causative proteins after which designing compounds that are capable of modulating the role of those proteins. This has become commonplace exercise in biopharmaceutical industries. Both experimental and computational strategies play tremendous roles inside the drug discovery and improvement and most of the times run complementing each other. The predominant aim of computer aided drug design (CADD) is to convey the nice chemical entities to experimental testing by way of reducing costs and late level attrition.

CADD involves:

a. Computer primarily based strategies to make greater efficient drug discovery and development system.

b. Building up chemical and biological facts databases about ligands and targets/proteins to identify and optimize novel tablets.

c. Devising in silico filters to calculate drug likeness or pharmacokinetic properties for the chemical compounds prior to screening to allow early detection of the compounds that are more likely to fail in clinical ranges and in addition to enhance detection of promising entities.

There are diverse computational techniques which are able to generating the desired effect at various ranges of the drug discovery technique . The two main disciplines of CADD which can control modern

day drug discovery process and which might be capable of accelerating drug discovery are bioinformatics and
cheminformatics.

In general:
a. Bioinformatic strategies hold lots of potential in goal identity (generally proteins/enzymes), goal validation, understanding the protein, evolution and phylogeny and protein modelling

b. Cheminformatic techniques maintain numerous potential in storage management and renovation of facts related to chemical substances and associated residences, and importantly inside the identity of novel
bioactive compounds, and further in lead optimization. Besides, cheminformatic techniques are extensively utilized in in silico ADME (Absorption, Distribution, Metabolism and Elimination) prediction and associated issues that help inside the discount of the late degree failure of compounds.

Need of computer aided Drug-Design:

Besides the significant charges and time related in bringing a brand-new drug to the market, a number of the major motives for the pharmaceutical industries to search for alternative or complementary strategies to experimental screening are:
a. In a survey study, 5 of the 40,000 compounds tested in animals attain human testing and simplest one out of those five reaching the clinical trials is finally approved.
b. On the other hand, the notable increment in chemical area and goal proteins/receptors will increase the call for for the HTS and could in turn call for brand spanking new lead identification strategies (rational approaches) to reduce charges and enhance efficacy.

c. Advances in computing technology on software program and hardware have enabled reliable computational strategies.

3. IDEA:

Using in-silico drug design process we are trying to come up with a drug that can be used to cure Crohn's disease.

Crohn's disease

Crohn's disease is a type of inflammatory bowel disease. Still research are going on in this medical condition to understand it's causes, the people getting affected, why and how it is developing and how to keep it under control. Despite of advancement in medicines and treatment methods over three decades, still no suitable drug has been discovered therefore cure is not yet possible.
Crohn's disease commonly occurs in the small intestine and the colon. It can affect any part of the gastrointestinal (GI) tract. It can also deviate from GI and affect other parts of the body.

As of now the medication used to keep the disease from spreading in the body are immunomodulators, antibiotics, anti–inflammatory drugs, biological therapies, dietary changes, surgery.

4. METHODOLOGY:

Basic steps involved in *In silico* drug designing: -

Step 1. Disease related genomics

Step 2. Target identification

- ❖ Bioinformatics
- ❖ Reverse Docking
- ❖ Protein Structure prediction

Step 3. Target Validation

- ❖ Target druggability
- ❖ Tool compound design

Step 4. Lead discovery

- ❖ Library design
- ❖ Docking scoring
- ❖ De novo designs
- ❖ Pharmacophore
- ❖ Target flexibility

Step 5. Lead Optimization

- ❖ QSAR
- ❖ 3D-QSAR
- ❖ Structure based optimization

Step 6. Preclinical Tests

- ❖ In silico ADMET prediction
- ❖ Physiologically-based pharmacokinetic (PBPK) simulations

Step 7. Clinical trials

The disease we are concerned here is Crohn's disease and the objectives of our project shall be the following: -

1. Find all the non-synonymous 'SNPs' from 'NCBI'.

2. We will use 'SIFT' for the prediction of the 'SNPs'.

3. We then check the ‘sift’ results in ‘nssnp analyzer’.

4. Retrieve the ‘.pdb’ file from uniport and then analyse this file in ‘PYMOL’.

5. Then we will use another tool called ‘patchdock’ where we will upload the ‘.pdb’ file of the native protein which is mutated after analysis of ‘pymol’ results.

6. On reaching up-till this we need to study about the structures via literature survey that can be used as a ligand for the receptor.

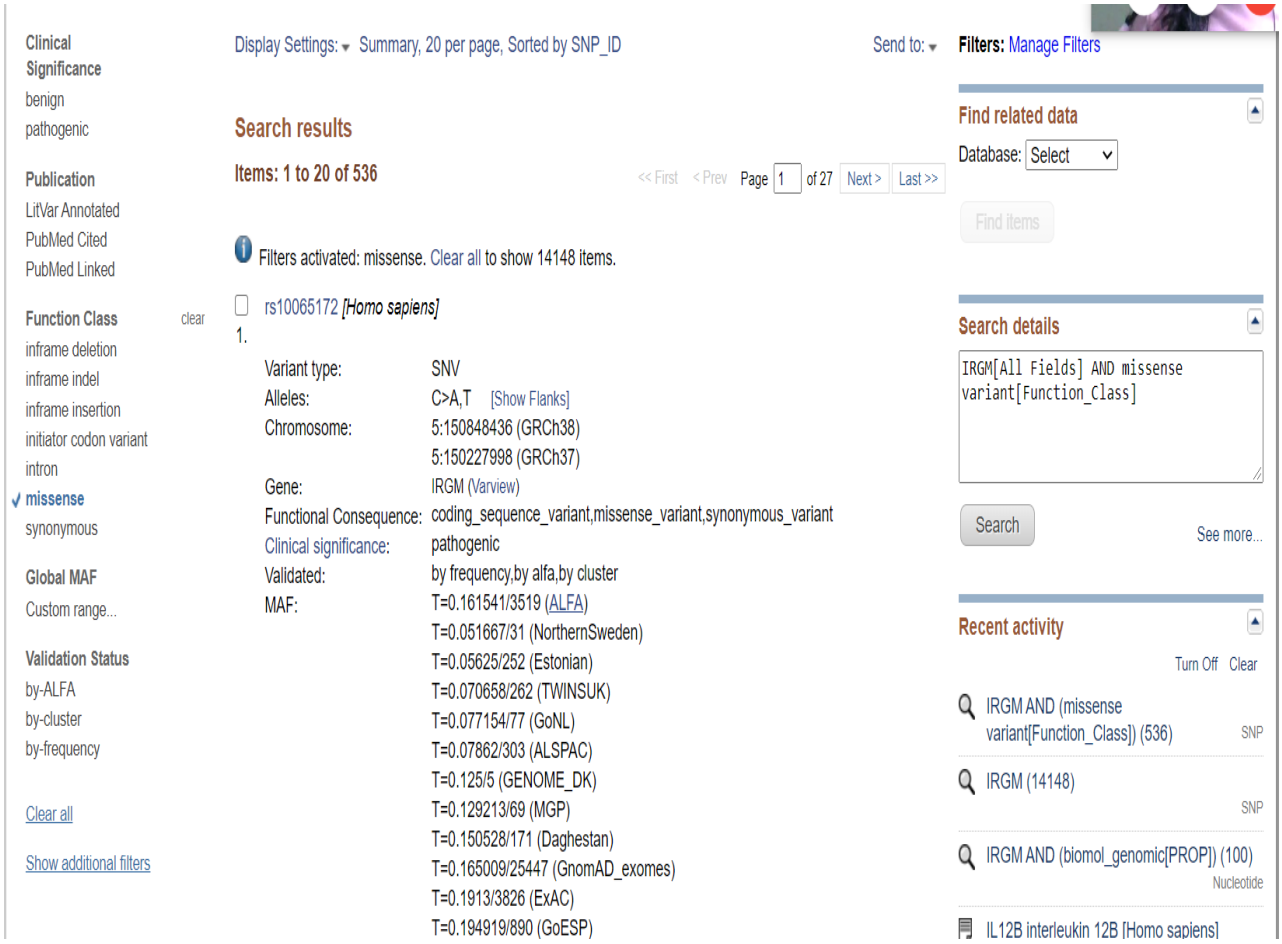
These steps in detail are as follows:

Initially we have to do some literature survey to get hold of a gene to work on. The same disease has many genes available, so we choose one and then copy the gene code and paste it on ncbi search bar. Here, we use the “IRGM”

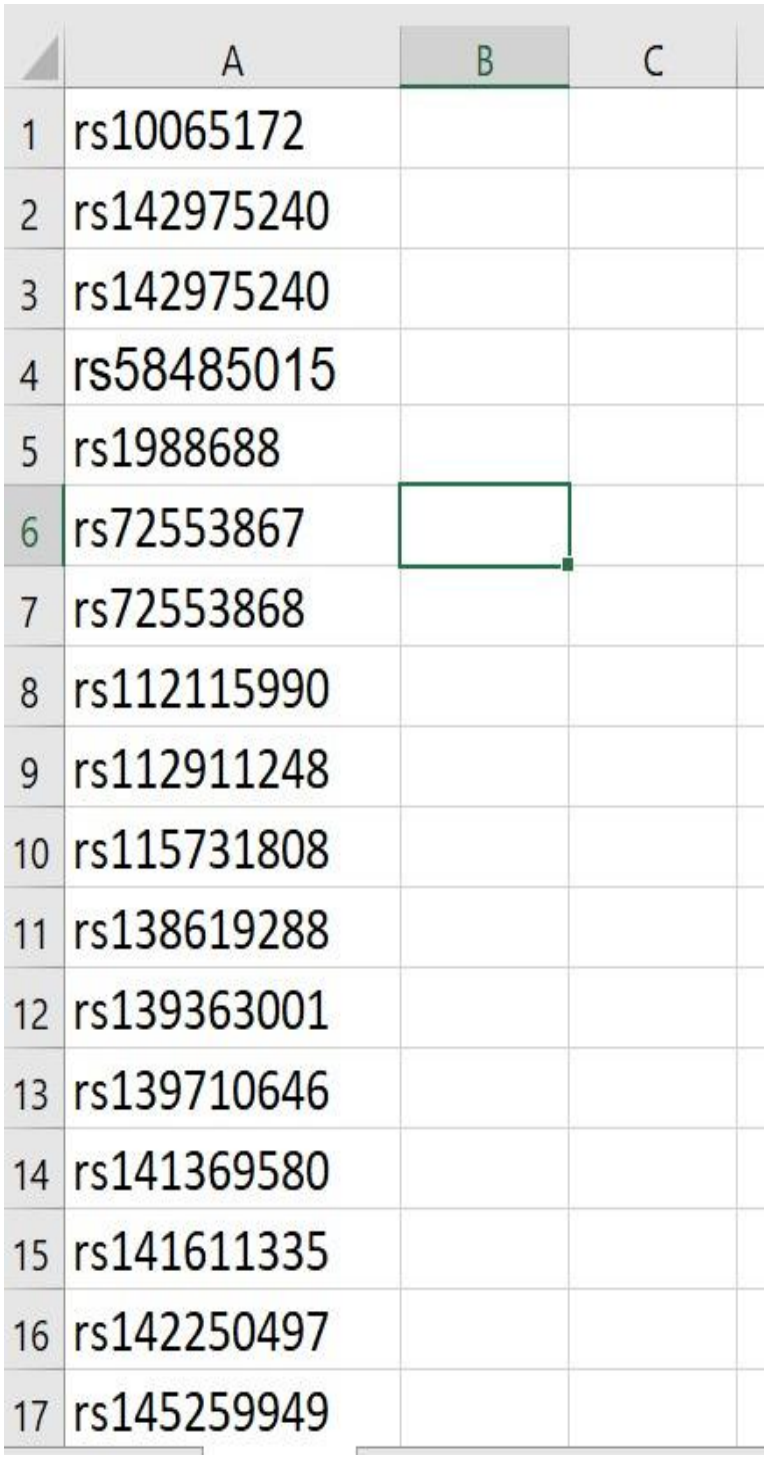
URL: https://www.ncbi.nlm.nih.gov/2

In the search bar, we select the ‘SNP’ option in the left panel and then search.

After the search results appear, we have to find the option ‘missense’ in the left panel of filters under the sub-option ‘Functional class’. The list changes.



Now from the list, we have to copy all the ‘rs’ id(s) of all the entries, into an excel sheet (ms-excel can be used). All data is put into one column.



	A	B	C
1	rs10065172		
2	rs142975240		
3	rs142975240		
4	rs58485015		
5	rs1988688		
6	rs72553867		
7	rs72553868		
8	rs112115990		
9	rs112911248		
10	rs115731808		
11	rs138619288		
12	rs139363001		
13	rs139710646		
14	rs141369580		
15	rs141611335		
16	rs142250497		
17	rs145259949		

After all the data entry work is done, we then go to the next website, which is 'sift'.

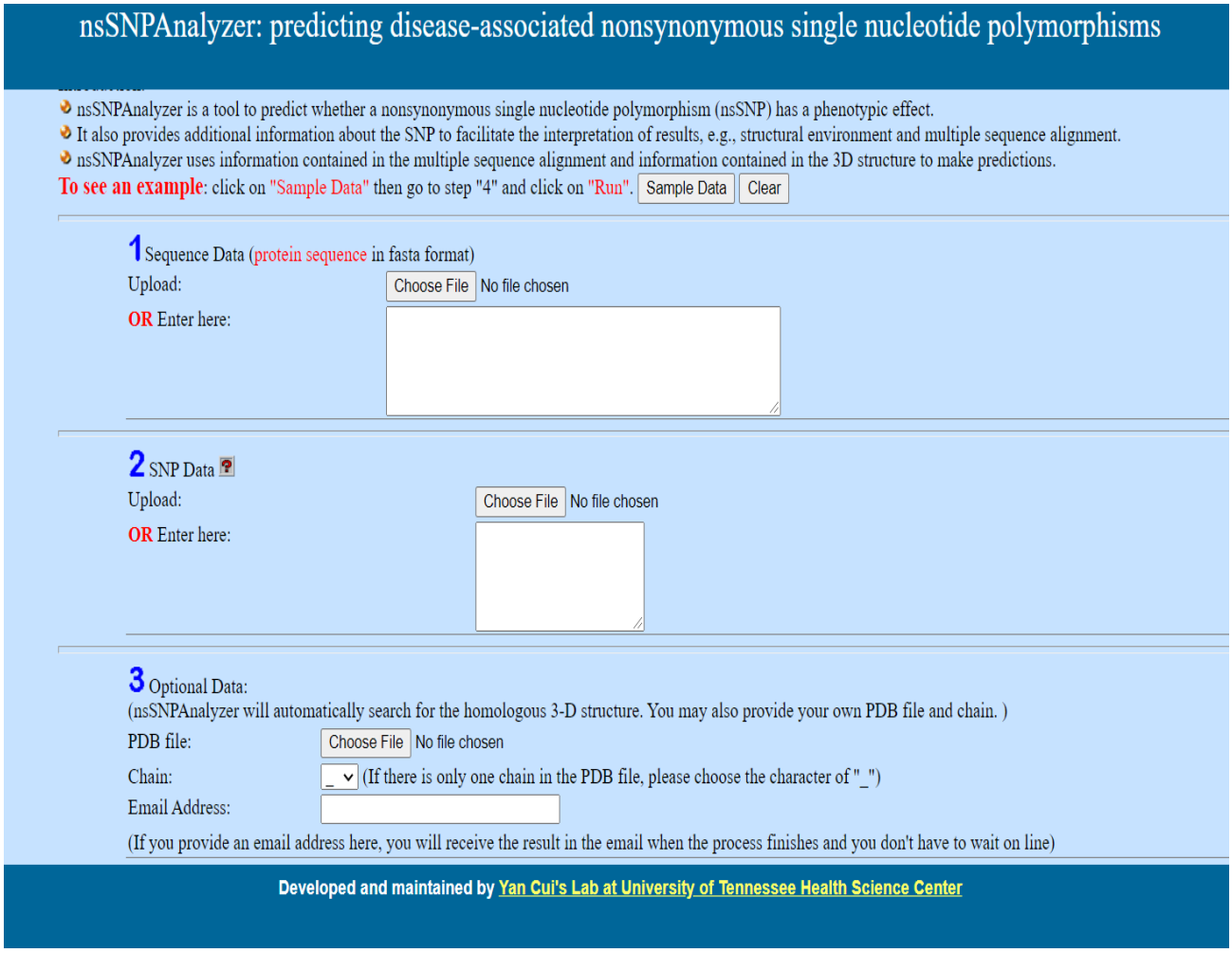
URL: https://sift.bii.a-star.edu.sg/

From the home page we then select the option sift for genomes and then another page opens up.

Here, we add the variant file, which actually is our excel file, and then submit our document, and then save the results.

Now, we need to check the deleterious SNPs and that is done using this page. Therefore, we go to 'nssnp' page.

url: http://snpanalyzer.uthsc.edu/

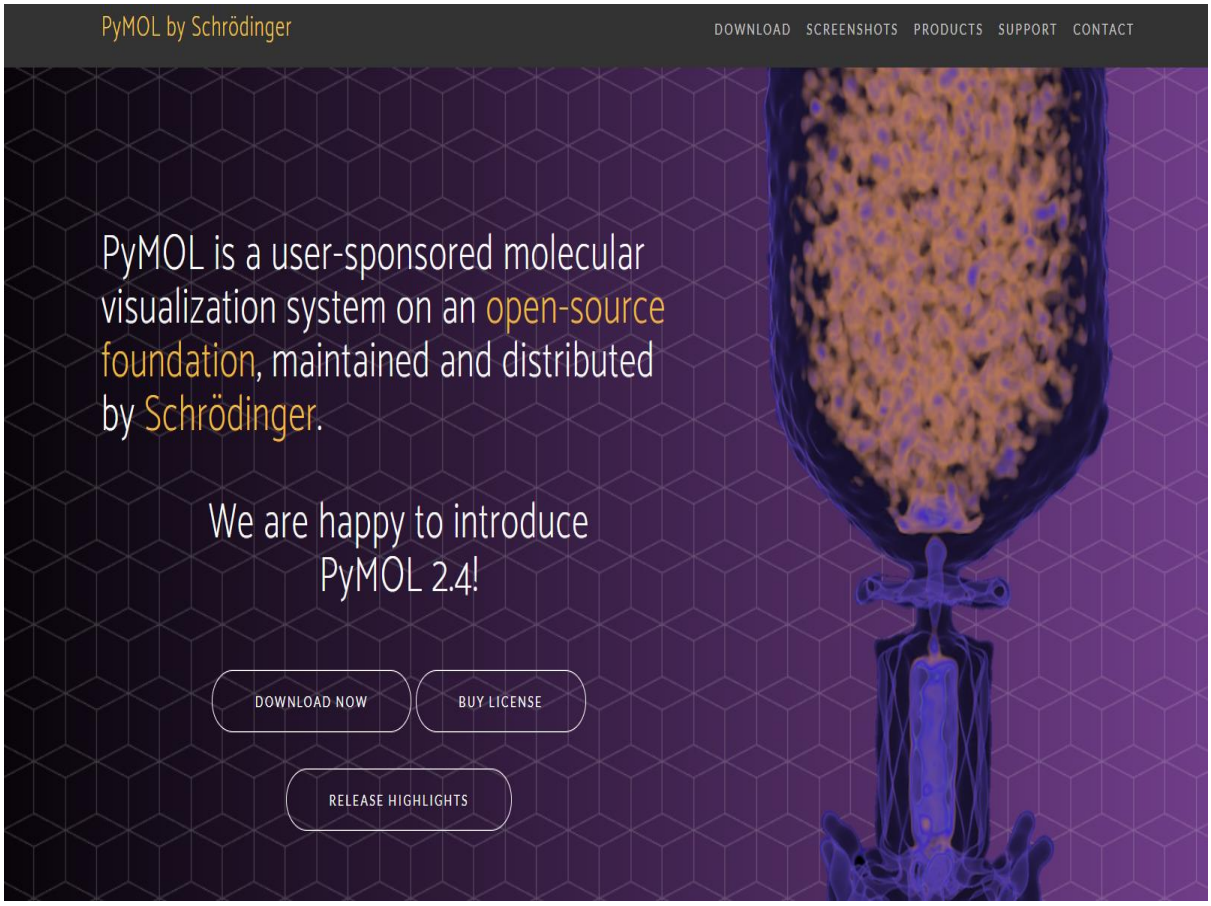


Now, using 'chimera' we mutate the wild type protein structure. Our aim is to use 'mutagenesis' to mutate the native(wild-type) protein with deleterious substitutions.

Download link: https://www.cgl.ucsf.edu/chimera/cgi-bin/secure/chimera-get.py?file=win64/chimera-1.14-win64.exe

This can also be done in 'pymol'.

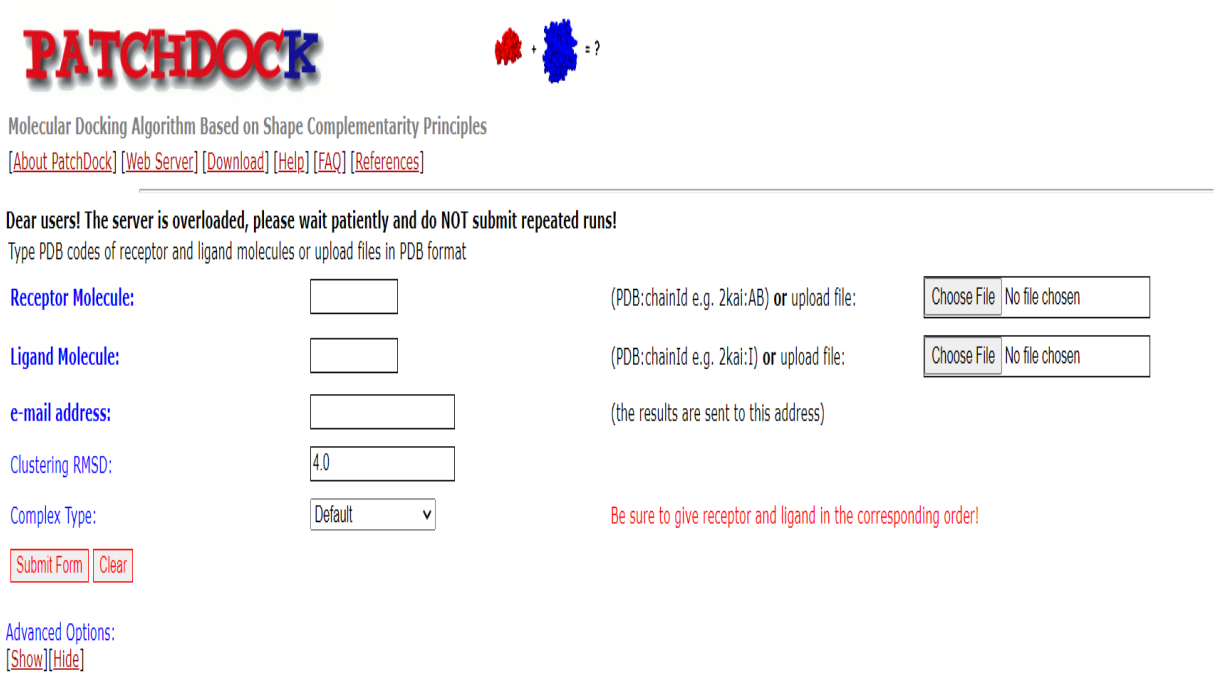
url: https://pymol.org/2/



Then, for each mutated protein we shall have one .pdb file and this need to be trimmed for further use.

Then we go into the ‘patchdock’ page.

url: https://bioinfo3d.cs.tau.ac.il/PatchDock/php.php



Here, we add the ligand and receptor and then enter a valid email id where the results of the docking shall be received. So, the receptors are each mutated .pdb file that we had saved earlier after analysis in chimera/pymol. On the other hand, the ligands shall be found for these via brute force of literature survey. After this the patchdock docks and checks the compatibility, whose results re mailed to us on our given email id.

5. LITERATURE SURVEY:

TITLE	AUTHOR NAME	DATE OF PUBLICATION	SUMMARIZATION
From *in silico* target prediction to multi-target drug design: Current databases, methods and applications	Author links open overlay panel AlexiosKoutsoukas[a] BenjaminSimms[b] JohannesKirchmair[a] Peter J.Bond[a] Alan V.Whitmore[cd] StevenZimmer[c] Malcolm P.Young[ce] Jeremy L.Jenkins[b] MeirGlick[b] Robert C.Glen[a] AndreasBender[a]	18 May 2011	Given the tremendous growth of bioactivity databases, the use of computational tools to predict protein targets of small molecules has been gaining importance in recent years. In this review, we firstly survey databases that can be used for ligand-based target prediction and which have grown tremendously in size in the past. We furthermore outline methods for target prediction that exist, both based on the knowledge of bioactivities from the ligand side and methods that can be applied in situations when a protein structure is known. Applications of successful *in silico* target identification attempts are discussed in detail, which were based partly or in whole on computational target predictions in the first instance. Finally, we will conclude with the prospective application of databases to not only predict, retrospectively, the protein targets of a small molecule, but also how to design ligands with desired polypharmacology in a prospective manner.

In Silico Studies in Drug Research Against Neurodegenerative Diseases	Farahnaz Rezaei Makhouri[1] and Jahan B. Ghasemi[2,*]	2018 Jul	In the present review, the authors provide a basic background about neurodegenerative diseases and in silico techniques in the drug research. Furthermore, they review the various in silico studies reported against various targets in neurodegenerative diseases, including homology modelling, molecular docking, virtual high-throughput screening, quantitative structure activity relationship (QSAR), hologram quantitative structure activity relationship (HQSAR), 3D pharmacophore mapping, proteochemometrics modelling (PCM), fingerprints, fragment-based drug discovery, Monte Carlo simulation, molecular dynamic (MD) simulation, quantum-mechanical methods for drug design, support vector machines, and machine learning approaches.
In Silico Approach for Predicting Toxicity of Peptides and Proteins	Sudheer Gupta, Pallavi Kapoor, Kumardeep Chaudhary, Ankur Gautam, Rahul Kumar, and Gajendra P. S.Raghava	2013 Sep 13	We obtained toxic peptides having 35 or fewer residues from various databases for developing prediction models. Non-toxic or random peptides were obtained from SwissProt and TrEMBL. It was observed that certain residues like Cys, His, Asn, and Pro are abundant as well as preferred at various positions in toxic peptides. We developed models based on machine learning technique and quantitative matrix using various properties of peptides for predicting toxicity of peptides. The performance of dipeptide-based model in terms of accuracy was 94.50% with MCC 0.88. In addition, various motifs were extracted from the toxic peptides and this information was combined with dipeptide-based model for developing a hybrid model. In order

			to evaluate the over-optimization of the best model based on dipeptide composition, we evaluated its performance on independent datasets and achieved accuracy around 90%. Based on above study, a web server, ToxinPred has been developed, which would be helpful in predicting (i) toxicity or non-toxicity of peptides, (ii) minimum mutations in peptides for increasing or decreasing their toxicity, and (iii) toxic regions in proteins.
Screening strategies to identify new chemical diversity for drug development to treat kinetoplastid infections	ROB DON and JEAN-ROBERT IOSET	January 2014	The Drugs for Neglected Diseases *initiative* (DND*i*) has defined and implemented an early discovery strategy over the last few years, in fitting with its virtual R&D business model. This strategy relies on a medium- to high-throughput phenotypic assay platform to expedite the screening of compound libraries accessed through its collaborations with partners from the pharmaceutical industry. We review the pragmatic approaches used to select compound libraries for screening against kinetoplastids, taking into account screening capacity. The advantages, limitations and current achievements in identifying new quality series for further development into preclinical candidates are critically discussed, together with attractive new approaches currently under investigation

A search for medications to treat COVID-19 via *in silico* molecular docking models of the SARS-CoV-2 spike glycoprotein and 3CL protease	Donald C.HallJr. Hai-FengJi	12 April 2020	Utilization of the available sequence information, homology modeling, and *in slico* docking a number of available medications might prove to be effective in inhibiting the SARS-CoV-2 two main drug targets, the spike glycoprotein, and the 3CL protease.

6. CONCLUSION:

During the technique of selection of novel drug candidates many vital steps are taken to put off such compounds that have aspect consequences and also show interplay with other drugs. In-silico drug designing softwares play an vital function to design revolutionary proteins or drugs in biotechnology or the pharmaceutical field. The drug designing softwares and packages are used to study molecular modelling of gene, gene expression, gene sequence analysis and 3D structure of proteins. In-silico methods were of top notch importance in target identity and in prediction of novel drugs.

In silico drug design is a powerful method, in particular whilst used as a device inside an apparatus, for coming across new drug leads in opposition to critical targets. After a target and a structure of that target on are defined, new leads can be designed from chemical ideas or selected from a subset of small molecules that scored nicely when docked in silico towards the target. After a preliminary evaluation of bioavailability, the candidate leads continue in an iterative method of re-entering structural determination and re-evaluation for optimization. Focused libraries of synthesized compounds primarily based on in silico approach can create a very promising lead which can maintain to scientific trials. As structural genomics, bioinformatics, cheminformatics, proteomics and computational energy continue to explode with new advances, similarly

successes in in silico drug design are probably to follow. Each year, new targets are being diagnosed, structures of these targets are being determined at an astonishing rate, and our functionality to capture a quantitative print of the interactions among macromolecules and ligands is accelerating.

7. REFERENCES:

- https://www.sciencedirect.com/science/article/pii/S1477893920301150
- https://www.futuremedicine.com/doi/10.4155/ebo.13.272
- https://www.healthline.com/health/crohns-disease#and-uc
- https://microbenotes.com/in-silico-drug-designing-and-role-of-bioinformatics/
- https://pymol.org/2/
- http://snpanalyzer.uthsc.edu/
- https://www.cgl.ucsf.edu/chimera/cgi-bin/secure/chimera-get.py?file=win64/chimera-1.14-win64.exe
- https://www.frontiersin.org/articles/10.3389/fphar.2017.00298/full
- http://www.oapublishinglondon.com/article/1119#
- https://bioinfo3d.cs.tau.ac.il/PatchDock/php.php
- https://sift.bii.a-star.edu.sg/
- https://www.ncbi.nlm.nih.gov/2
- D. Kitchen, H. Decornez, J. Furr and J. Bajorath. Docking and scoring in virtual screening for drug discovery: methods and applications. Nature Reviews Drug Discovery, 3, 2004, 935–949
- V. Rao and K. Srinivas. Modern drug discovery process: An in silico approach. Journal of Bioinformatics and Sequence Analysis, 2, 2011, 89-94.
- H. Böhm, G. Schneider, H. Kubinyi, R. Mannhold and H. Timmerman. Virtual screening for bioactive molecules (Methods and principles in medicinal chemistry) (Wiley-VCH, Weinheim, Germany, 2000).
- H. Noori and R. Spanagel. In silico pharmacology: drug design and discovery's gate to the future. In silico Pharmacology, 1, 2013, 1-2.

- R. Taylor, P. Jewsbury and J. Essex. A review of protein-small molecule docking methods. J. Comput. Aided Mol. Des., 16, 2002, 151–166.
- D. Joseph-McCarthy. Computational approaches to structure-based ligand design. Pharmacol. Ther., 84, 1999, 179–191.
- R. Bohacek and C. McMartin. Modern computational chemistry and drug discovery: structure generating programs. Curr. Opin. Chem. Biol., 1, 1997, 157–161.
- H. Carlson and J. McCammon. Accommodating protein flexibility in computational drug design. Mol. Pharmacol., 57, 2000, 213– 218.
- B. Shoichet, S. McGovern, B. Wei and J. Irwin. Lead discovery using molecular docking. Curr. Opin. Chem. Biol., 6, 2002, 439–446.
- G. Klebe and H. Bohm. Energetic and entropic factors determining binding affinity in protein-ligand complexes. J. Recept. Signal Transduct. Res., 17, 1997, 459–473.
- U. Madsen, P. Krogsgaard-Larsen and T. Liljefors. Textbook of drug design and discovery (Taylor and Francis, Washington, USA, 2002).
- S. Arlington. An industrial revolution in R&D. Pharmaceutical Executive, 20, 2000, 74-84.
- D. Plewczynski, A. Philips, M. von Grotthuss, L. Rychlewski and K. Ginalski. HarmonyDOCK: the structural analysis of poses in protein-ligand docking. Journal of Computational Biology, 18, 2010, 1-10.
- K. Gunjan, S. Dinesh, V. Yogesh and S. Vishal (2013). A review on drug designing, methods, its applications and prospects. International Journal of Pharmaceutical Research and Development, 5, 2013, 15-30.
- T. Sterling, J.J. IrwinZINC 15--ligand discovery for everyone
  J Chem Inf Model, 55 (11) (2015), pp. 2324-2337

- U. Pieper, B.M. Webb, G.Q. Dong, D. Schneidman-Duhovny, H. Fan, S.J. Kim, N. Khuri, Y.G. Spill, P. Weinkam, M. Hammel, J.A. Tainer, M. Nilges, A. Sali
  ModBase, a database of annotated comparative protein structure models and associated resources
  Nucleic Acids Res, 42 (Database issue) (2014), pp. D336-346