Open source software and web services for designing therapeutic molecules

G. P. S. Raghava, Head Bioinformatics Centre,
Institute of Microbial Technology, Chandigarh

Important Web Sites

Open Source Drug Discovery(OSDD) **http://www.osdd.net/**

CRDD Site: **http://crdd.osdd.net/**

Raghava's Group: **http://www.imtech.res.in/raghava/**

Software Development for Drug Discovery

Importance of Open Source for Drug Discovery

- Discovery of Drug by Public for Public
- Drugs for Disease Specific to Developing Countries (like India)
- Development of Drugs for diseases of poor persons
- Process of Discovery will be fast (few to many contributors)
- Academic institutes/universities and small industry may afford

Examples of open source software

- Operating Systems
  - Linux
  - FreeBSD, OpenBSD, and NetBSD
- Internet
  - Apache (> 50% of the world's web servers)
  - BIND: DNS for the entire Internet.
  - Sendmail (Most email servers)
  - OpenSSL (standard for secure communication)
- Programming Tools
  - Languages (Perl, Python, PHP)
  - GNU compilers and tools (GCC, Make)

OPEN SOURCE DRUG DISCOVERY

Search this site

- Home
- What is OSDD
- Who we are
- How OSDD works
- OSDD Portfolio
- Join the movement
- Publications and Presentations
- Media Centre
- News updates
- Contact Us
- FAQ's
- Sitemap

A CSIR TEAM INDIA CONSORTIUM WITH GLOBAL PARTNERSHIP

CSIR INDIA

Message from Chief Mentor, OSDD

"I believe that affordable healthcare is a right for all. But, pragmatically speaking, when it comes to health, we need to have a balanced view between health as a right and health as a business" Read Message

I am a Student I am an Organisation I am a Researcher

SysBorg 2.0 Sysborg 2.0 is OSDD's cyber infrastructure for collaborative research. Sysborg has over 6700 participants from 130 countries across the world. Join Sysborg 2.0 in order to know more about the ongoing activities of OSDD, view results of experiments and participate in research projects **Login** (Forgot Password) **Register**

OSDD Chem OSDDChem is the web interface for large scale synthesis of diverse chemical compounds to screen them against TB and Malaria.Log into OSDDChem using your Sysborg OpenID to know more about submitting molecules and project proposals. Login

Community Developed Resources:

Current Events and Updates !!!

CSIR-OSDD Signs MoU with Royal Society of Chemistry In pursuit of common aims and to raise awareness of the importance of cheminformatics to accelerate the discovery of novel therapies for neglected diseases like TB and Malaria, CSIR-OSDD ...

Posted Jan 30, 2013, 12:18 AM by Anshu Bhardwaj

OSDD Research Unit @ IISc An OSDD Research Unit will be opened at the Indian Institute of Science (IISc), Bengaluru, on 9th January 2013 to coordinate various ongoing activities at IISc and other centers.The ...

Posted Jan 5, 2013, 3:04 AM by Anshu Bhardwaj

Showing posts **1 - 2** of **20**. View more »

News Updates

Award Winners of CSIR-OSDD-VP Short Film Competition-2013 The final screening of the 23 shortlisted videos of CSIR-OSDD-VP Short Film Competition-2013 on "Need of New Drugs for TB" was successfully conducted by a national jury ...

Posted Jan 18, 2013, 2:16 AM by Anshu Bhardwaj

Dr. Kalam on OSDD at Pharma Vision 2020

Dr Kalam, who launched the Pharma Vision 2020 programme in 2003, said currently India has one million pharmacists and pharmaceutical scientists.

al Conference on Open Source for Computer Aided Drug Discovery (March 22-26, 2009

Computational Resources for Drug Discovery

OSDD Forum is an initiative with a vision to provide affordable healthcare to the developing world. The OSDD concept aims to synergize the power of genomics, computational technologies and facilitate the participation of young and brilliant talent from Universities and industry. It seeks to provide a global platform where the best brains can collaborate and collectively endeavor to solve the complex problems associated with discovering novel therapies for neglected diseases like Tuberculosis.

CRDD (Computational Resources for Drug Discovery) is an important module of the *in silico* module of OSDD. The CRDD web portal provides computer resources related to drug discovery on a single platform. Following are major features of CRDD:

- **CRDD provides computational resources for researchers in the field of computer-aided drug design.**
- **CRDD allows users to discuss their problem with other members.**
- **CRDD gives equal opportunity to those willing to solve these problems.**
- **CRDD Wiki maintain wikipedia related to drug discovery.**
- **Contributors may host their database or web server on CRDD portal.**

Thus, CRDD provides a platform for researchers having limited resources.

India specific Genomes Sequenced, Assembled and Annotated

Genomes sequence/assemble/annotate at CSIR Institutes

Organism	Discription	Institute	Publication
Acinetobacter baumannii MSP4-16	Isolated from mangrove soil sample from Parangipettai (11°30 N,79° 47'E), Tamil Nadu, India.	CSIR-IMTECH	23558533
Streptomyces gancidicusÂ Strain BKS 13-15	Isolated from mangrove sediment samples collected from the Bhitar Kanika Mangrove Reserve Forest, Odissha, India.	CSIR-IMTECH	23599292
Serratia fonticolaÂ Strain AU-AP2C	Isolated from the Pea Rhizosphere	CSIR-IMTECH	24309742
PantoeaÂ sp. Strain AS-PWVM4	Isolated from the rhizosphere ofÂ Punica granatum, exhibits phosphate solubilization	CSIR-IMTECH	24309733

Current Topics in Medicinal Chemistry, 2013, *13*, 1172-1191

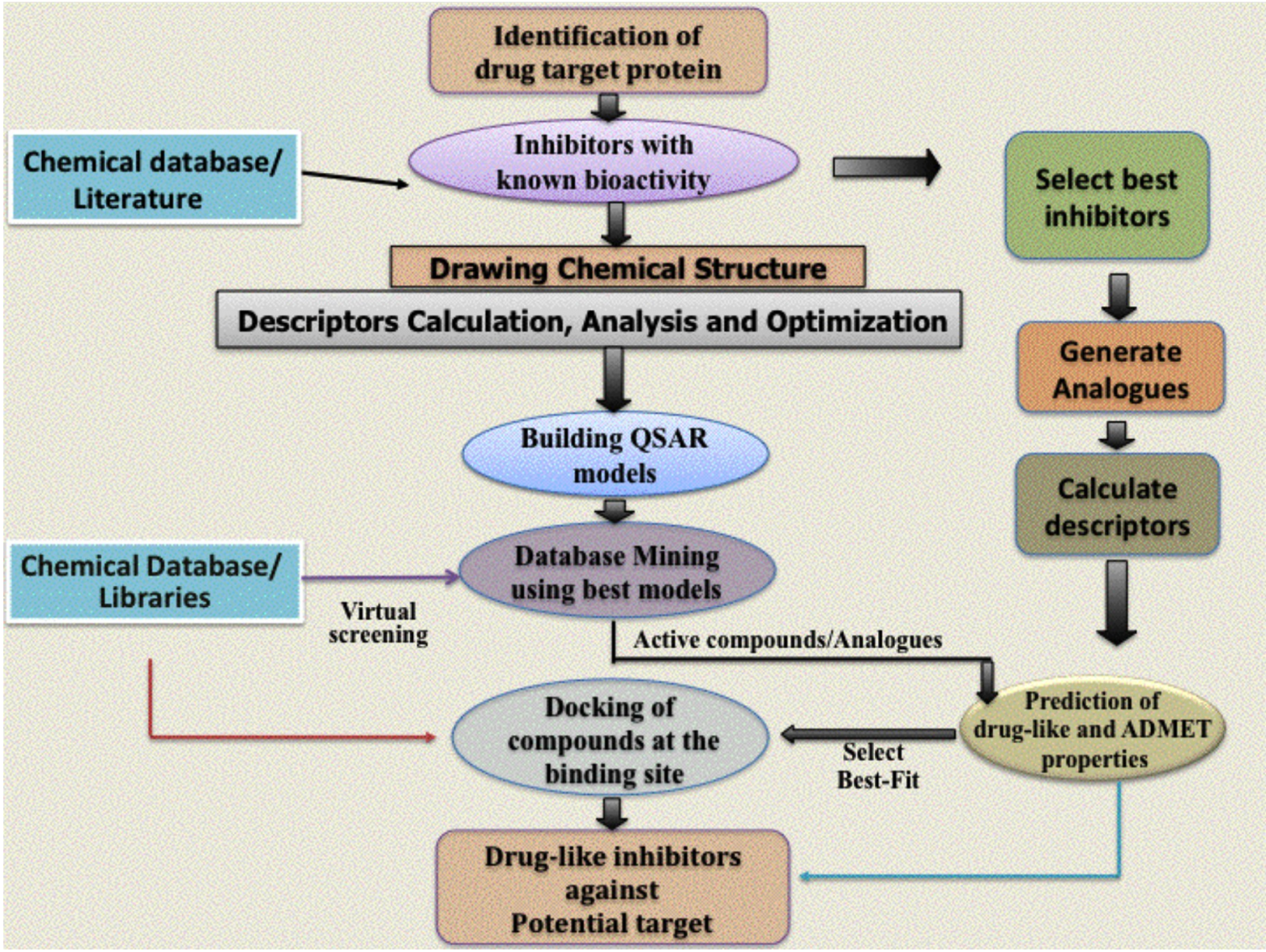
Open Source Software and Web Services for Designing Therapeutic Molecules

Deepak Singla[1,2], Sandeep Kumar Dhanda[1], Jagat Singh Chauhan[1], Anshu Bhardwaj[3], Samir K. Brahmachari[3,4], Open Source Drug Discovery Consortium[3] and Gajendra P.S. Raghava[1,*]

[1]*Bioinformatics Centre, CSIR-Institute of Microbial Technology, Chandigarh, India;* [2]*Centre for Microbial Biotechnology, Panjab University, Chandigarh, India;* [3]*CSIR-Open Source Drug Discovery Unit, New Delhi, India;* [4]*CSIR-Institute of Genomics and Integrative Biology, New Delhi, India*

Abstract: Despite the tremendous progress in the field of drug designing, discovering a new drug molecule is still a challenging task. Drug discovery and development is a costly, time consuming and complex process that requires millions of dollar and 10-15 years to bring new drug molecules in the market. This huge investment and long-term process are attributed to high failure rate, complexity of the problem and strict regulatory rules, in addition to other factors. Given the availability of 'big' data with ever improving computing power, it is now possible to model systems which is expected to provide time and cost effectiveness to drug discovery process. Computer Aided Drug Designing (CADD) has emerged as a fast alternative method to bring down the cost involved in discovering a new drug. In past, numerous computer programs have been developed across the globe to assist the researchers working in the field of drug discovery. Broadly, these programs can be classified in three categories, freeware, shareware and commercial software. In this review, we have described freeware or open-source software that are commonly used for designing therapeutic molecules. Major emphasis will be on software and web services in the field of chemo- or pharmaco-informatics that includes *in silico* tools used for computing molecular descriptors, inhibitors designing against drug targets, building QSAR models, and ADMET properties.

An overview of the workflow of *in silico* drug designing process



Important Points

1. Source of Molecules (databases or repositories)
2. Molecular Editors (editing & viewing existing molecules)
3. Analog Generators (software used to generate analogs)
4. Structure Optimization (Energy/geometry of molecules)
5. Calculation of Molecular Descriptors
6. Chemical Similarity Search
7. Development of QSAR/QSPR Models
8. Classification and Clustering of Small Molecules
9. Docking Small Molecules in Macromolecules
10. Pharmacophore Tools/Search
11. Software for ADMET Techniques
12. Designing of Inhibitors
13. Major Initiatives towards affordable drugs

Databases and resources managing and hosting chemical compounds

Database	Brief Description with URL
PubChem	**A comprehensive database of bioassays, compounds and substances (http://pubchem.ncbi.nlm.nih.gov/)**
ChEMBL	**Database of drug like molecules (https://www.ebi.ac.uk/chembldb)**
Zinc	**Maintain commercially-available compounds for virtual screening (http://zinc.docking.org/)**
ChemDB	**Collection of small-molecules (http://cdb.ics.uci.edu/)**
ChemSpider	**A chemical database (http://www.chemspider.com/)**
MMsINC	**Commercial compounds (http://mms.dsfarm.unipd.it/MMsINC/)**
KEGG	**Maintain comprehensive information (http://www.genome.jp/kegg/)**
SMPDB	**Small molecule Pathway database (http://www.smpdb.ca)**
HMDB	**Human Metabolites (http://www.hmdb.ca/)**
PDBeChem	**Dictionary of chemical components refered in PDB entries (http://www.ebi.ac.uk/pdbe-srv/pdbechem/)**
PDB-Bind	**Binding affinity information for PDB Ligands (http://sw16.im.med.umich.edu/databases/pdbbind/index.jsp)**
BindingDB	**Binding affinity of PDB Ligands (http://www.bindingdb.org/)**
NCI	**Small molecules related to cancer (http://cactus.nci.nih.gov/ncidb2.1/)**

List of major molecular editors, frequently used for drawing and editing molecules

Editors	Brief description
BKchem	**Python based free 2D molecule editor (http://bkchem.ziral.org/)**
PubChem Sketcher [*117*]	**A web-based tool for sketching, integrated in PubChem (http://pubchem.ncbi.nlm.nih.gov/edit2/index.html)**
ChemSketch	**ACD/ChemSketch Freeware is a free software for drawing chemicals (**http://www.acdlabs.com/resources/freeware/chemsketch/ **)**
JChemPaint	**Editor for 2D chemical structures (http://jchempaint.github.com/)**
AccelrysDraw	**Draw and edit complex molecules, no fee for academic community (http://accelrys.com/products/informatics/cheminformatics/draw/index.html)**
XDrawChem	**Molecule drawing program (http://xdrawchem.sourceforge.net/)**
MedChem Designer	**Drawing molecules and integration with ADMET property. (http://simplus-downloads.com/)**
JME	**JME Molecular Editor (**http://www.molinspiration.com/jme/ **)**

Analogs generation softwares

Software	Brief description
SmiLib [*118*]	**Enumerates combinatorial libraries with very high rate (http://gecco.org.chemie.uni-frankfurt.de/smilib/)**
GLARE [*119*]	**Generate combinatorial library (http://glare.sourceforge.net/)**
Library synthesizer	**Virtual chemical enumeration (http://tripod.nih.gov/?p=370)**
CLEVER [*120*]	**Chemical Library Editing, Visualization and Enumerating Resource (http://datam.i2r.a-star.edu.sg/clever/)**
Newlead [*121*]	**Generate of combinatorial library from bioactive conformations (http://www.ccl.net/cca/software/MAC/index.shtml)**

List of software and web servers used for structure optimization of molecules

Software	Brief description
Openbabel [*38*]	**The Open Source Chemistry Toolbox (**http://openbabel.org/**)**
Frog [*39*]	**Generation of free online drug conformation (http://bioserv.rpbs.jussieu.fr/cgi-bin/Frog)**
SMI23D [*122*]	**Generation of 3D (http://www.chembiogrid.org/cheminfo/smi23d/)**
Cyndi [*123*]	**Generate geometrically extended or compact conformations (http://www.biomedcentral.com/1471-2105/10/101/additional/)**
Balloon	**Conformer Ensembles (http://web.abo.fi/~mivainio/balloon/index.php)**
DG-AMMOS [*41*]	**Generate 3D conformation using distance geometry (http://www.mti.univ-paris-diderot.fr/fr/downloads.html)**
TINKER [*124*]	**Software Tools for Molecular Design (http://dasher.wustl.edu/tinker/)**
MOPAC	**Semiempirical quantum chemistry program (http://openmopac.net/)**

Important software and webserver for computing molecular descriptors

Software	Brief description
PowerMV [*125*]	Window based calculation of descriptors (http://nisla05.niss.org/PowerMV/index.html)
PaDEL [*126*]	Fingerprints calculation (http://padel.nus.edu.sg/software/padeldescriptor)
Joelib	Descriptor calculation software (http://sourceforge.net/projects/joelib)
MOLD2 [*45*]	Calculating descriptors from a two-dimensional chemical structure (http://www.fda.gov/ScienceResearch/BioinformaticsTools/Mold2/default.htm)
Afgen	Fragment-based descriptors (http://glaros.dtc.umn.edu/gkhome/afgen/overview)
ISIDA-fragmentor	Calculate of Substructural Molecular Fragments and ISIDA Fragments (http://infochim.u-strasbg.fr/spip.php?rubrique49)
ODDesripotrs	Simple java-based command level tool for descriptor calculation (http://www.softpedia.com/get/Science-CAD/ODDescriptors.shtml)
CDK [*127*]	Chemistry Development Kit (http://cdk.sourceforge.net)
Filter-it	Filter-it is used for calculating descriptors and filtering drug-like molecules. (http://silicos-it.com/software/software.html)
MODEL [*128*]	A webserver for molecular descriptor based upon 3D structure http://jing.cz3.nus.edu.sg/cgi-bin/model/model.cgi

Similarity search algorithms and their web links

Software	Description
JCsearch	**JC search used for searching similar structure, substructureas well as super structure from a given database. (http://www.chemaxon.com/jchem/doc/user/Jcsearch.html)**
PubChem	**PubChem provide the facility to search similar chemical in PubChem database using PubChem based binary fingerprints. (http://pubchem.ncbi.nlm.nih.gov/search/)**
SIMCOMP [*129*]	**Chemical structure similarity search against KEGG COMPOUND, KEGG DRUG, and other databases. SIMCOMP is based on 2D graph representation. (http://www.genome.jp/tools/simcomp/)**
SUBCOMP [*129*]	**SUBCOMP is based on bit-string representation of chemical structures. (http://www.genome.jp/tools/subcomp/)**
SMSD [*130*]	**SMSD is a Java based software library for calculating Maximum Common Subgraph (MCS) between small molecules. This will help us to find similarity/distance between two molecules. (http://www.ebi.ac.uk/thornton-srv/software/SMSD/)**

Machine learning and feature selection techniques in cheminformatics

Software	Brief description
Software used for developing QSAR model	
SVM	SVM is a supervised learning technique, used for classification and regression analysis. The QSAR models can be optimized using different SVM parameters and kernels. (http://www.cs.cornell.edu/People/tj/svm_light/)
ANN	ANN is based on supervised learning, unsupervised learning and reinforcement learning. SNNS (Stuttgart Neural Network Simulator) is a free software simulator for neural networks. (http://www.ra.cs.uni-tuebingen.de/SNNS/)
kNN	The k-nearest neighbor algorithm (k-NN) is a method for classifying objects based on closest training examples. TiMBL is an open source software package implementing k-nearest neighbor classification. (http://ilk.uvt.nl/timbl/)
Weka [53]	Weka is a collection of visualization tools and algorithms for data analysis and predictive modeling, It contains libSVM, SMO, NaiveBayes, LMT, Random Forest etc learning algorithms. (http://www.cs.waikato.ac.nz/~ml/weka/)
Feature Selection techniques	
Weka	Weka is a popular java based tool used in feature selection. There are various feature selection methods and evaluators are available in Weka package. (http://www.cs.waikato.ac.nz/~ml/weka/)
Rapidminer	RapidMiner is a freely available software. It contains Brute force, Forward selection and Backward elimination algorithms for feature selection. (http://rapid-i.com/content/view/181/190/)
Orange	orngFSS (Orange.feature.selection) module provides

List of chemical clustering tools and their web addresses

Software	Brief description
ChemMine	Chemical clustering and analysis (http://chemmine.ucr.edu/)
ChemMineR	It is R based open source chemical clustering tool. (http://manuals.bioinformatics.ucr.edu/home/chemminer)
jcluster	ChemAxon Cluster (http://www.chemaxon.com/products/jklustor/)
ChemBioserver [57]	Chemical clustering webserever (http://bioserver-3.bioacademy.gr/Bioserver/ChemBioServer/)

Different types of pharmocophore searching softwares

Softwares	Brief description
Pharmapper [*69*]	**Ligand based Pharmacophore search (http://59.78.96.61/pharmmapper/)**
PharmaGist [*70*]	**Ligand based Pharmacophore search (**http://bioinfo3d.cs.tau.ac.il/PharmaGist/**)**
Pharmer [*71*]	**Both PDB and ligand based pharmacophore search (http://smoothdock.ccbb.pitt.edu/pharmer/)**
ZincPharma [*133*]	**Both PDB and ligand based pharmacophore search (http://zincpharmer.csb.pitt.edu/pharmer.html)**
Boomer	**Pharmacokinetic drug monitering (http://www.boomer.org/)**
Cyber Patient	**A software for pharmacokinetic simulations (http://www.labsoft.com/www/software.html)**
PKfit	**A tool for pharmacokinetic modeling (http://cran.csie.ntu.edu.tw/web/packages/PKfit/index.html)**
JPKD	**Therapeutic drug monitoring (http://pkpd.kmu.edu.tw/jpkd/)**
tdm	**Therapeutic drug monitoring (http://pkpd.kmu.edu.tw/tdm/)**
mobilePK	**(http://pkpd.kmu.edu.tw/mobilepk/)**

Some open source initiatives for drug discovery with their research area

Project	Research Areas
Drugs for Neglected Diseases Initiative	**Sleeping sickness, visceral leishmaniasis, Chagas disease (http://www.dndi.org/)**
Infectious Disease Research Institute	**Tuberculosis, leishmaniasis, Chagas disease, malaria, leprosy and Buruli ulcer (http://www.idri.org/)**
Blue Obelisk	**Provides open source cheminformatics tools. (http://sourceforge.net/apps/mediawiki/blueobelisk/index.php?title=Main_Page)**
OSDD	**Promoting open source for neglected disease. (http://www.osdd.net/)**
OpenTox	**Toxicology (http://www.opentox.org/)**
Global Alliance for TB	**Tuberculosis (http://www.tballiance.org/)**

Computational Resources for Drug Discovery

CSIR

Home OSCADD

Inhibitor Prediction

KiDoQ (Mtb target)

GDoQ (Mtb target)

ABMpred (Mtb target)

eBooster (Mtb target)

MDRIpred (Mtb cell)

CancerIN (Cancer)

ntEGFR (Cancer EGFR)

EGFRpred (Cancer EGFR)

DiPCell (Pancreatic Cancer)

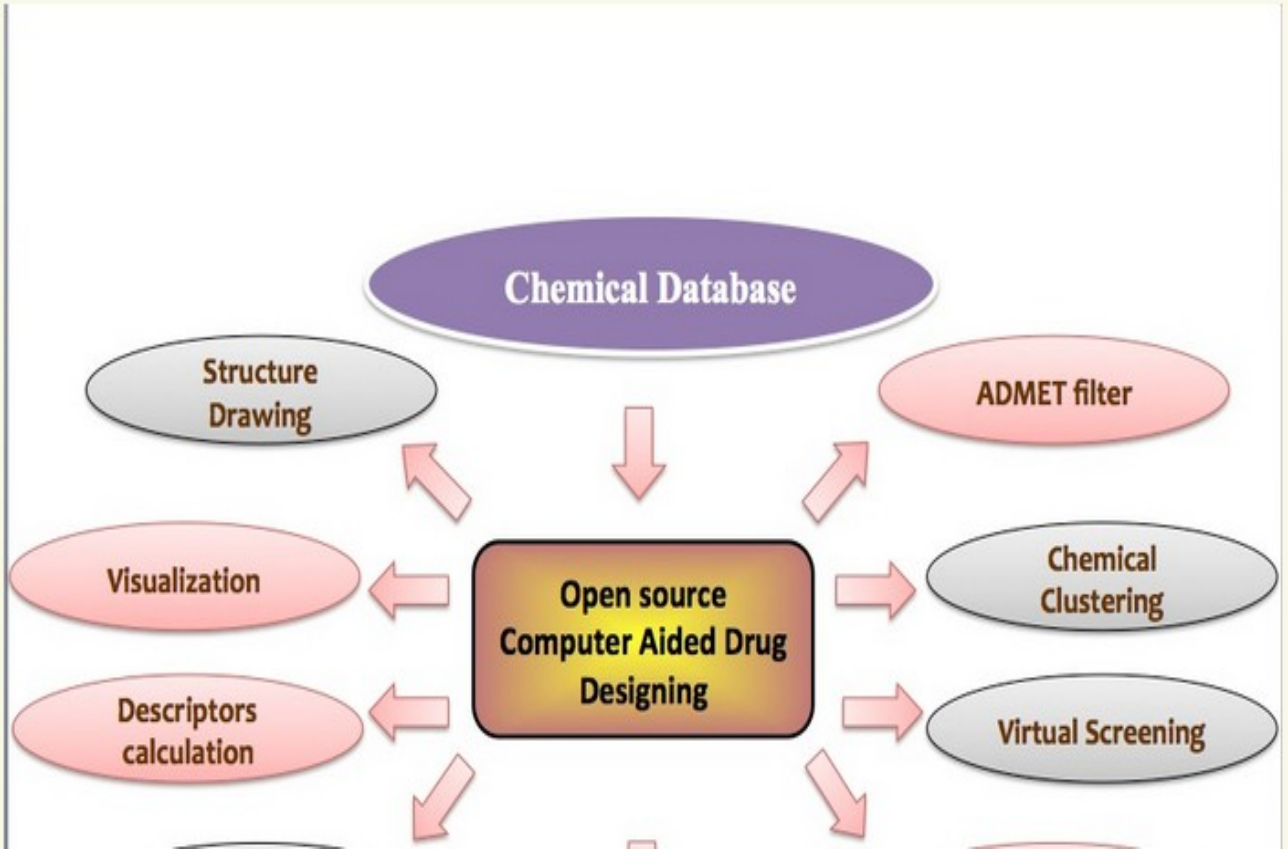
DMKPred (Human Kinases)

TLR4HI (Human TLR4)

HIVFin (HIV)

Open Source for Computer-Aided Drug Discovery

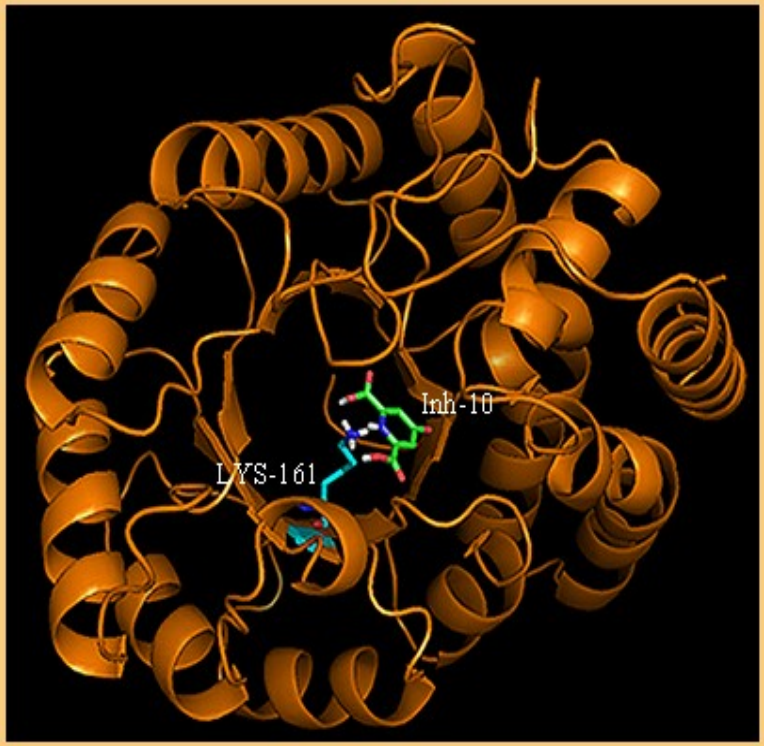
Aim of this page is to provide service to community involved in drug discovery. We are promoting open source in drug disc
possible service we are planning to provide, If you have any idea, please send to us at raghava@imtech.res.in, attempt w
implement your idea.



KiDoQ

Prediction of inhibition constant using docking and qsar

. S. (2010) KiDoQ: using docking based energy scores to develop ligand based model for predicting antibacterials. BMC Bioinformatics 2010, 11:1



KiDoQ, a web server has been developed to serve scientific community working in the field of designing inhibitors against Dihydrodipicolinate synthase (DHDPS), a potential drug target enzyme of a unique bacterial DAP/Lysine pathway. The server has employed the molecular docking and ligand based QSAR strategies to predict inhibitory activity value (Ki) of small compounds for DHDPS enzyme. The algorithm behind the server includes the docking of compounds to the active binding site of enzyme followed by QSAR modeling where, docking generated energy based scoring values (for the best conformer) are cascaded as input variables to QSAR based SVM model for prediction of Ki value. The QSAR model implemented on the server has been trained on the dataset of 23 inhbitors of DHDPS and predict the Ki value with correlation R/q2 values of 0.93/0.80 and MAE of 1.89.

MDRIpred: A webserver for predicting inhibitor against drug tolrent *M.Tuberculosis*

Home OSCADD

Inhibitor Prediction

KiDoQ (Mtb target)

GDoQ (Mtb target)

ABMpred (Mtb target)

eBooster (Mtb target)

MDRIpred (Mtb cell)

CancerIN (Cancer)

ntEGFR (Cancer EGFR)

EGFRpred (Cancer EGFR)

DiPCell (Pancreatic Cancer)

Submission Form

This server allows users to predict inhibitor against different phase of drug tolerant *M.tuberculosi*
methods.
1. Sketch using JME editor.
2. Paste molecules in the box.
3. Upload file containing moleclues in standard format.
Option 2 and 3 allow users to submit more than one molecule (upto 10).

Job Name (Optional)

Email Address (
(Please enter you
via email)

Method 1. **Sketch Structure using JME editor**

Method 2. **Past**

use an example

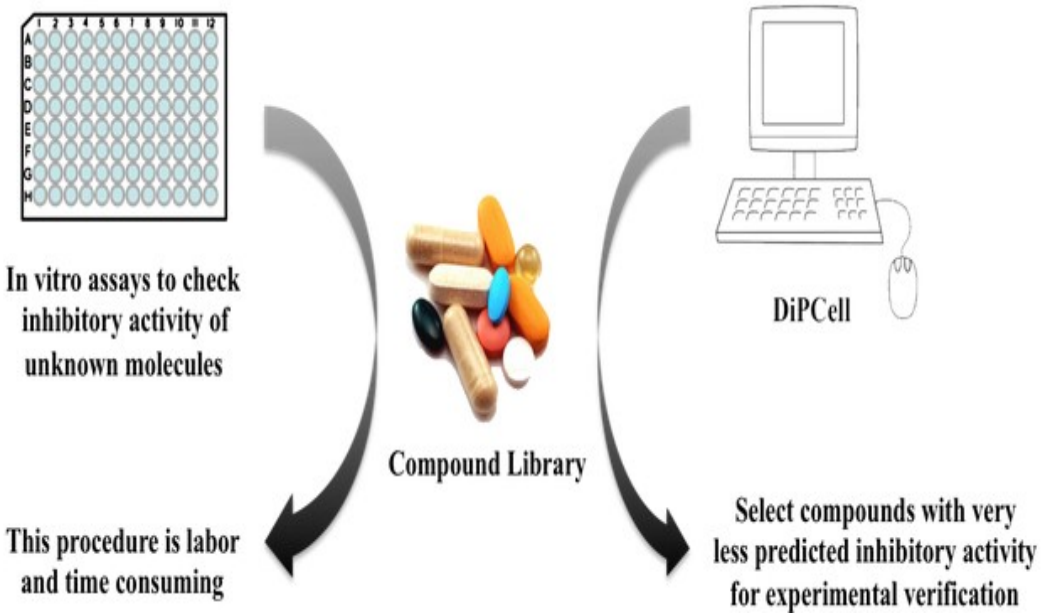
(Example Test.s

DiPCell: Designing of inhibitors of pancreatic cancer cell

CSIR – Institute of Microbial Technology, India

Welcome to DiPCell

DiPCell is a webserver for the predicting inhibitory activity of unknown molecules and designing their analogs against pancreatic cancer cell lines. DiPCell implements the QSAR models, which were developed by using SMOreg machine learning algorithm on high throughput drug screening data. This high throughput screening data is obtained from the Genomics of Drug Sensitivity in Cancer (GDSC) database.



DrugMint

A SERVER FOR IDENTIFICATION OF DRUG-LIKE MOLECULE

Home OSCADD

Inhibitor Prediction

KiDoQ (Mtb target)

GDoQ (Mtb target)

ABMpred (Mtb target)

eBooster (Mtb target)

MDRIpred (Mtb cell)

CancerIN (Cancer)

ntEGFR (Cancer EGFR)

EGFRpred (Cancer EGFR)

DiPCell (Pancreatic Cancer)

Welcome to Design Analogs Module

Lead optimization is a time consuming process in drug discovery. This tool generate all the possible analogs c
user-defined or identified R groups. Finally, the server will generate the virtual chemical library, which wo
properties. After screening, the results will be displayed in tabular format with the facility to sort them as per u
click help.

Paste Scaffold structure:
Use Scaffold Example

Paste Building Blocks structure:
Use Building Blocks Example

Paste linkers :
Use Linker Example

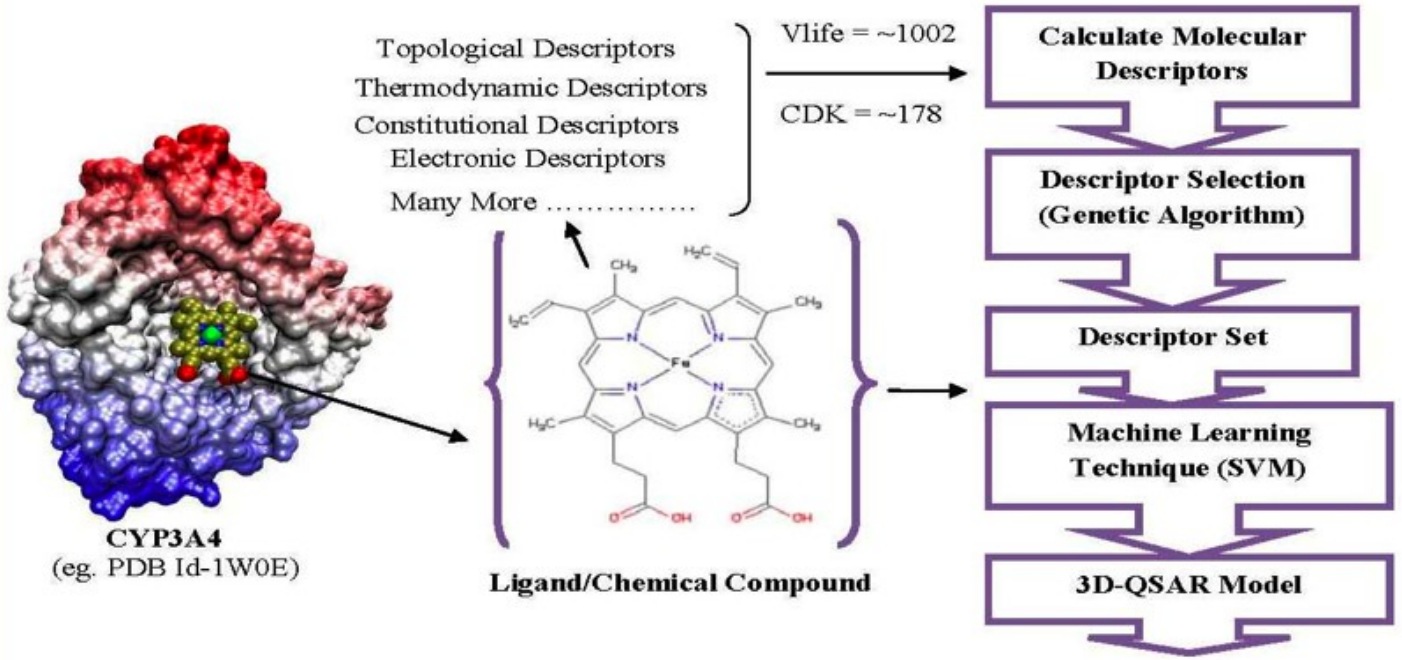
MetaPred:A webserver for the Prediction of Cytochrome P450 Isoform responsible for Metabolizing a Drug Molecule

ver, please cite:: **Prediction of cytochrome P450 isoform responsible for metabolizing a drug molecule** BMC Pharmacolo

- » Home
- » Submit
- » Algorithm
- » Developers
- » Contact Us
- » Help
- » Dataset

Cytochrome P450 enzymes (CYPs) are a multi gene family of heme-containing isoenzymes that are involved in oxidative metabolism of drug, steroids and carcinogens. About sixty CYPs are reported in human genome, but more than 90% of all therapeutic drugs are metabolized by five isoforms i.e. CYP1A2, CYP2C9, CYP2C19, CYP2D6 and CYP3A4.

MetaPred Server predict metabolizing CYP isoform of a drug molecule/substrate, based on SVM models developed using CDK descriptors.This server will be helpful for researcher working in the field of drug discovery.This study demonstrates that it is possible to develop free web servers in the field of chemoinformatics. This will encourage other researchers to develop web server for public use, which may lead to decrease the cost of discovering new drug molecules. In the following flow digaram we have given the example of CYP3A4, how this study will be helpful in drug design.



ToxiPred: A server for prediction of aqueous toxicity of small chemical molecules in T. pyriformis

Home OSCADD

Inhibitor Prediction

KiDoQ (Mtb target)

GDoQ (Mtb target)

ABMpred (Mtb target)

eBooster (Mtb target)

MDRIpred (Mtb cell)

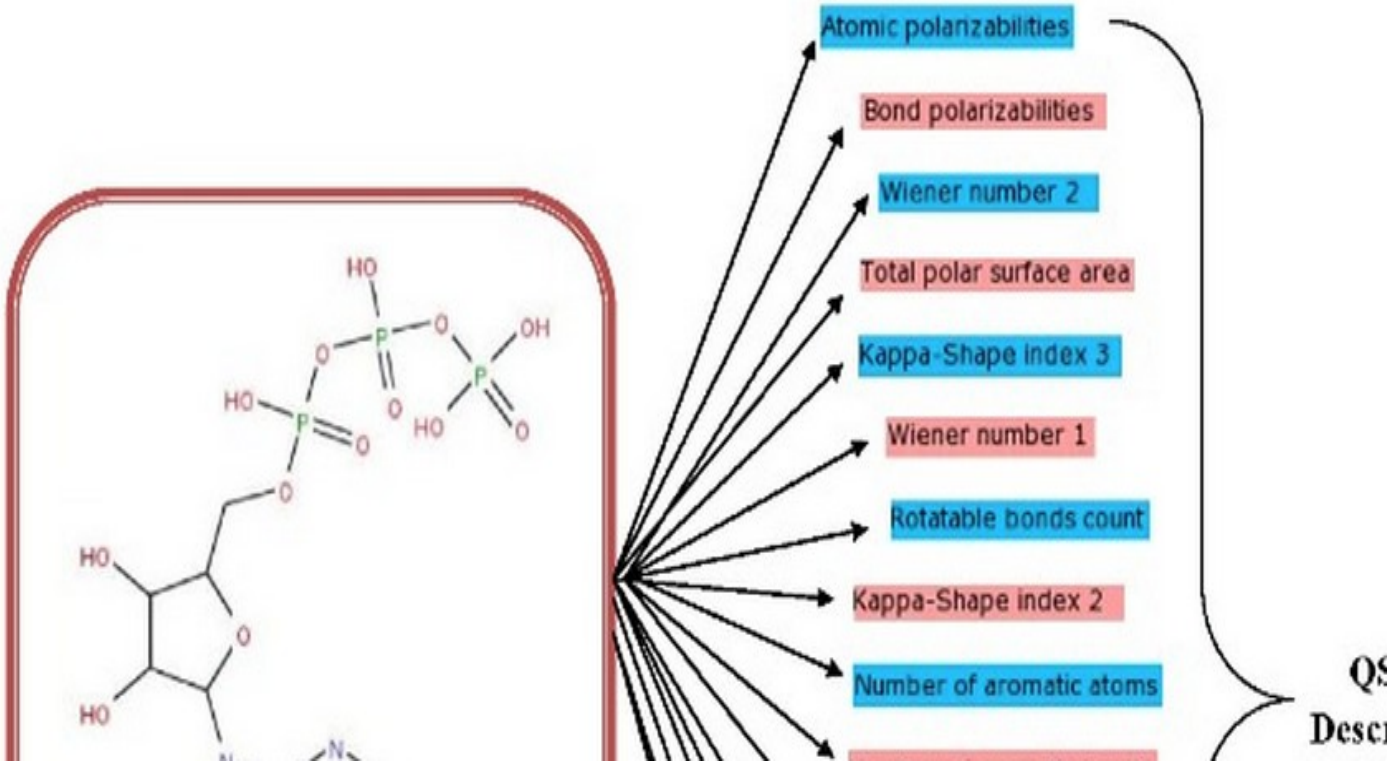
CancerIN (Cancer)

ntEGFR (Cancer EGFR)

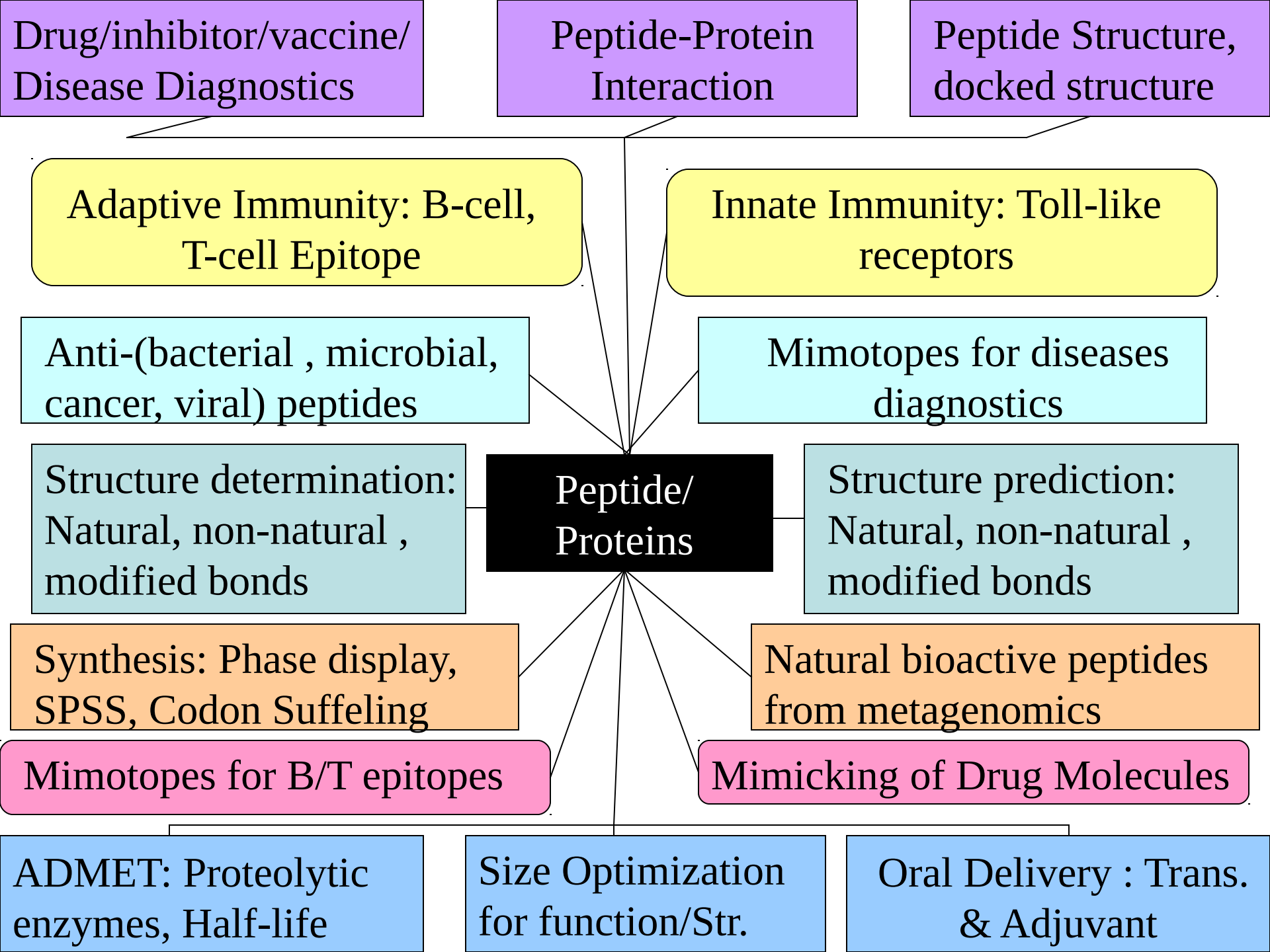
EGFRpred (Cancer EGFR)

DiPCell (Pancreatic Cancer)

Identification of non-toxic drug design is a major challenge in the field of drug design, most of th
development or even in clinical trials. Thus the use of predictive toxicology is called for. Keeping this
employed previously but we have started this study with latest dataset and apply different machin
implemented in WEKA and linear method (Multiple linear regressions (MLR)) using R-package. To
molecules in the commonly used format (mol/SMILE/sdf) and after descriptors calculation our server w
hope that present model will aid in the area of drug designing.



Software name	Description
AntiCP	Prediction and design of anticancer peptides.
Toxinpred	Prediction and designing of toxic/non-toxic peptides.
AntiBP	Mapping of antibacterial peptides in a protein sequence.
AntiBP2	Advanced server for predicting antibacterial peptides with high precision.
CPPsite	CPPsite is a database of experimentally validated cell penetrating Peptides
CellPPD	Computer-aided Designing of efficient cell penetrating peptides.
TumorHoPe	A comprehensive database experimentally characterized tumor homing peptides.
TumorHPD	Server dedicated for designing tumor homing peptides.
Destamp	Designing of stable antibacterial peptides
HEMOLYTIK	A repository of experimentally validated hemolytic and non-hemolytic peptides.
ParaPep	Compilation of experimentally validated anti-parasitic peptides and their structure


Drug/inhibitor/vaccine/ Disease Diagnostics
Peptide-Protein Interaction
Peptide Structure, docked structure
Adaptive Immunity: B-cell, T-cell Epitope
Innate Immunity: Toll-like receptors
Anti-(bacterial , microbial, cancer, viral) peptides
Mimotopes for diseases diagnostics
Structure determination: Natural, non-natural , modified bonds
Peptide/ Proteins
Structure prediction: Natural, non-natural , modified bonds
Synthesis: Phase display, SPSS, Codon Suffeling
Natural bioactive peptides from metagenomics
Mimotopes for B/T epitopes
Mimicking of Drug Molecules
ADMET: Proteolytic enzymes, Half-life
Size Optimization for function/Str.
Oral Delivery : Trans. & Adjuvant

Adaptive Immunity

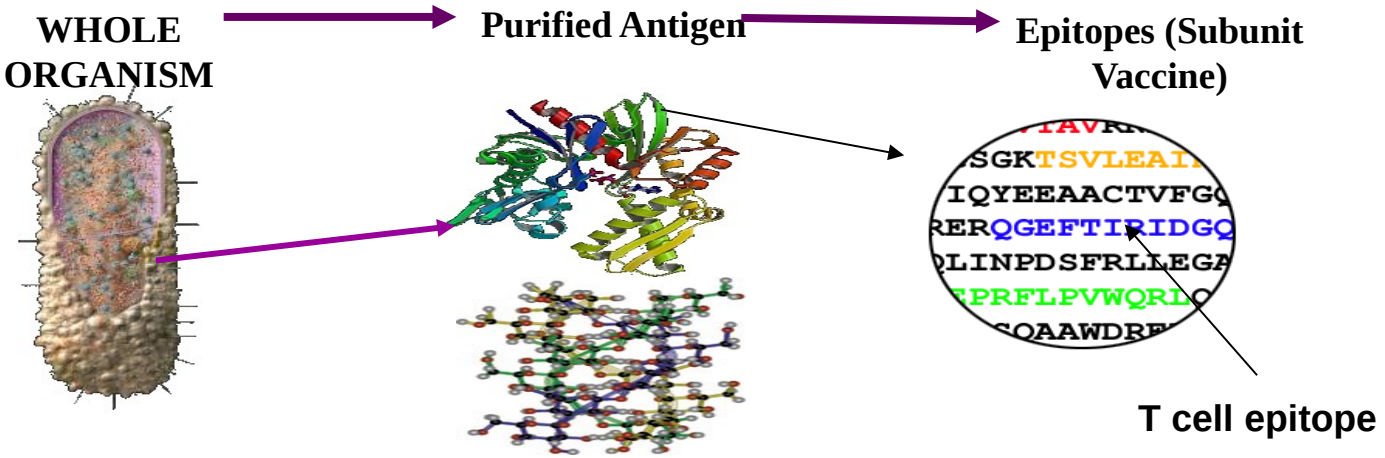
Innate Immunity

Vaccine Informatics

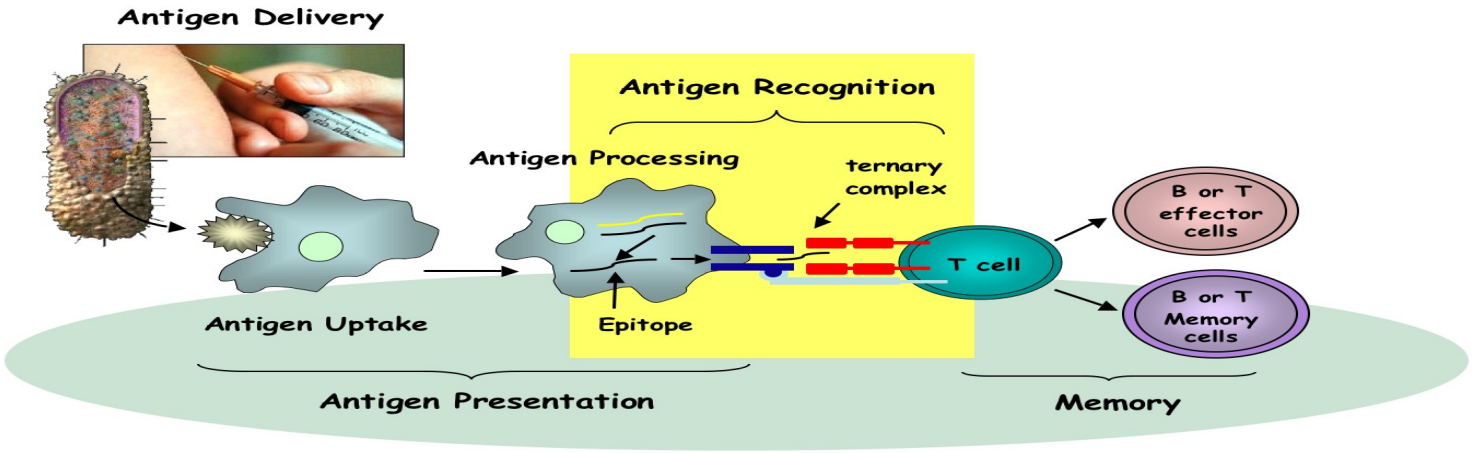
Bioinformatics Centre
IMTECH, Chandigarh

Protective Antigens

Vaccine Delivery



Attenuated



Adaptive Immunity

Innate Immunity

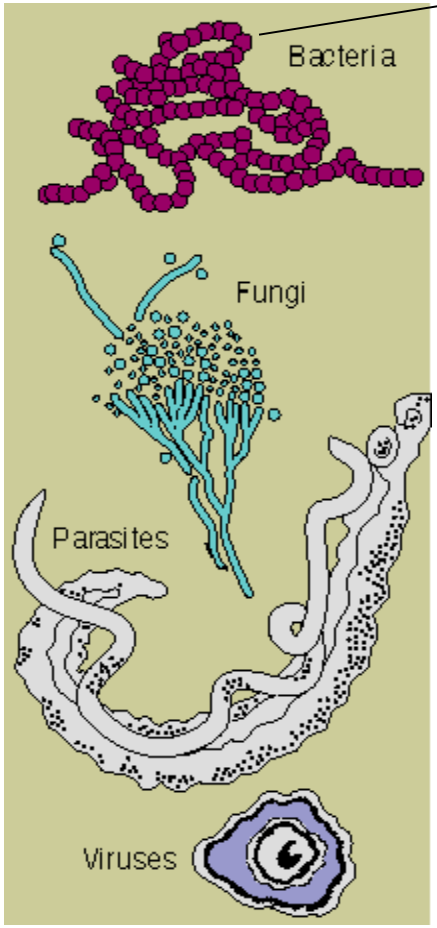
Vaccine Informatics

Bioinformatics Centre
IMTECH, Chandigarh

Protective Antigens

Vaccine Delivery

Disease Causing Agents



Pathogens/Invaders

Host Immune Defense System
For Microbial (Pathogens) Attack

Non-Specific Defense
(Innate Immunity)

Specific Defense
Adaptive Immunity
Responsible for Generating
Memory B & T Cells

TOLL LIKE RECEPTORS
PATHOGEN ASSOCIATED
MOLECULAR PATTERNS

Humoral Immunity
(B-Lymphocytes)

Cell-Mediated Immunity
(T-Lymphocytes)

B-Cell Epitopes
Antibody Prod.
Neutralization

Endogenous Antigen Processing

Exogenous Antigen Processing

CTL Epitopes
MHC I binders
TAP Binders
P. Cleavage Sites

T-Helper Epitopes
MHC II binders

Modelling of Immune System for Designing Epitope-based Vaccines

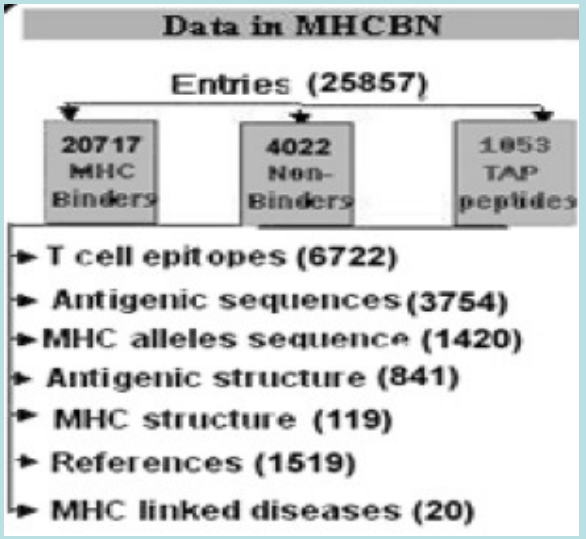
Adaptive Immunity (Cellular Response) : T_{helper} Epitopes

Propred: for promiscuous MHC II binders

MMBpred:for high affinity mutated binders

MHC2pred: SVM based method

MHCBN: A database of MHC/TAP binders and non-binders



Adaptive Immunity (Cellular Response) : CTL Epitopes

Pcleavage: for proteome cleavage sites

TAPpred: for predicting TAP binders

Propred1: for promiscuous MHC I binders

CTLpred: Prediction of CTL epitopes

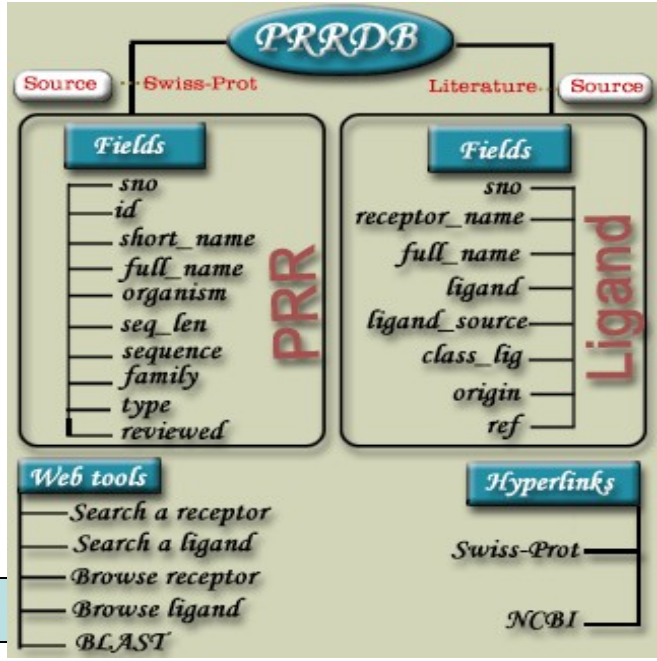
Adaptive Immunity (Humoral Response) :B-cell Epitopes

BCIpep: A database of B-cell eptioes;

ABCpred: for predicting B-cell epitopes

ALGpred: for allergens and IgE eptopes

HaptenDB: A datbase of haptens



Innate Immunity : Pathogen Recognizing Receptors and ligands

PRRDB: A database of PRRs & ligands

Antibp: for anti-bacterial peptides

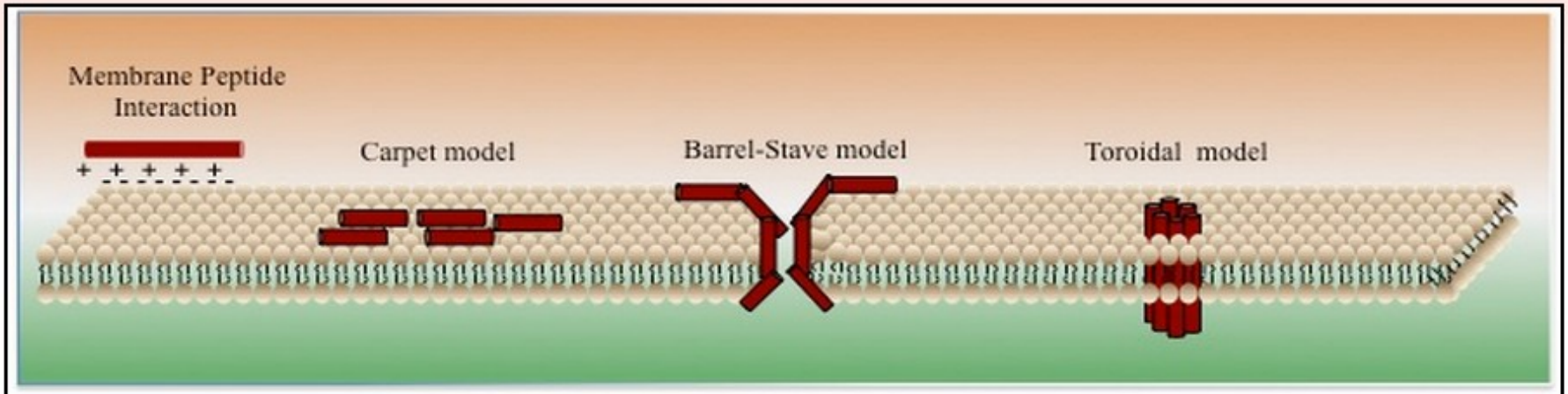
Signal transduction in Immune System

Cytopred: for classification of Cytokines

AntiCP:- Designing of Anticancer Peptides

Institute of Microbial Technology, Chandigarh India

Welcome to AntiCP



Various models of membrane permeation by anti-cancer peptides

AntiCP is web based prediction server for Anticancer peptides. SVM models developed are based on amino acid composition and binary profile features. Positive dataset consists of 225 antimicrobial peptides with anticancer properties. This server is extremely useful for the researchers working in the field of Anticancer peptides. This server allows the users to design ACPs and their mutants with

ParaPep -A Database of Anti-parasitic peptides

Home | Search | Browse | Similarity | Downloads | Important | General

Home

Home Page of ParaPep

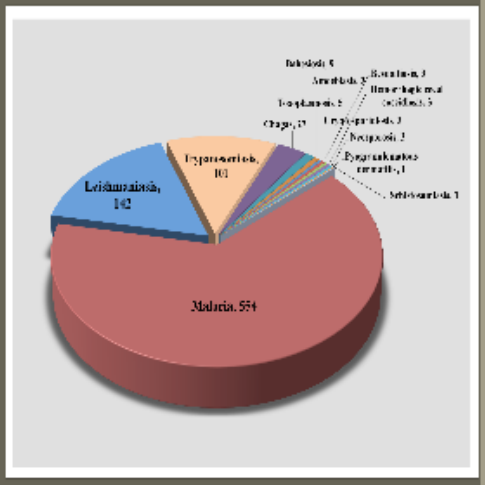
ParaPep: It is a manually curated repository of experientially validated anti-parasitic peptides and their structures. Data have been collected from research papers, published patents and other databases.

Peptide sequences: The current release of ParaPep contains 863 anti-parasite peptide entries, which have been tested against 12 different types of parasites. Most of the entries have been compiled for Malaria followed by Leishmaniasis and Trypanosomiasis.

Type of Peptides: ParaPep consists of various types of peptides, which includes linear peptides, cyclic peptides and peptides having L-amino acids, non-natural amino acids (e.g., D-amino acid, ornithine, etc.) and chemically modified residues.

Structure of Peptides: We determined secondary and tertiary structure of each peptide in ParaPep. using PepStr software. First, we scan PDB to identify all identical peptides to assign their tertiary structure. Structure of remaining peptides were predicted using PEPstr. Secondary structure of peptides were assigned using DSSP from their tertiary structure.

Parasitic Disease Covered



Quick View of ParaPep

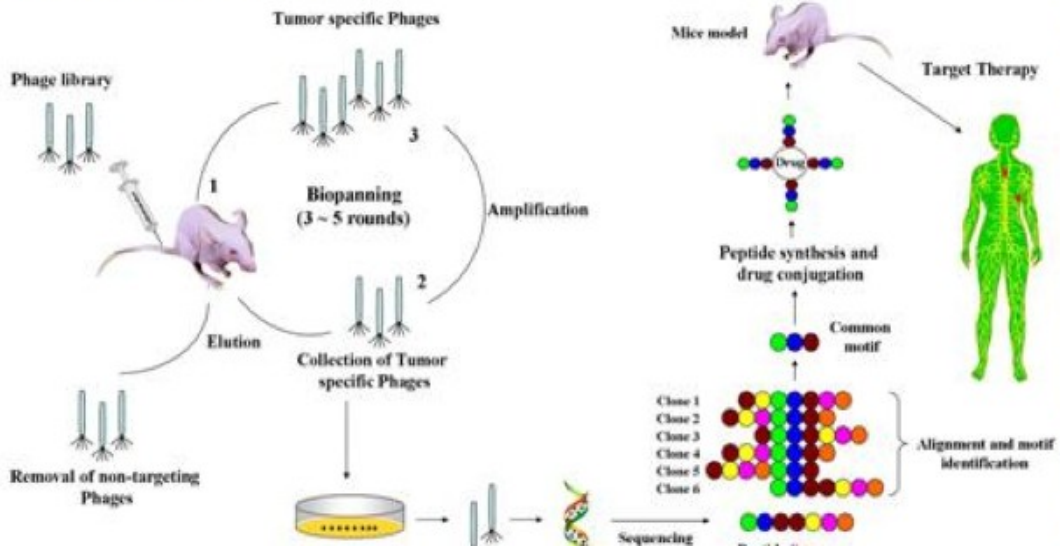
Peptide Resources/Databases

TumorHope - Tumor Homing Peptide Database

Home | Search | Browse | Structure | Tools | Important | Help | About Us | Contact Us

Welcome to TumorHope - A comprehensive database of Tumor Homing Peptides

TumorHope is a manually curated comprehensive database of experimentally characterized tumor homing peptides. These peptides recogninze tumor tissues and tumor associated micro environment, including tumor metastasis. Thus, they can be used to deliver drugs selectively in tumors.



Importance of Peptides: Poor selectivity of chemotherapeutic drugs for cancer is a major challenge for successful clinical outcome. Conjugation of drug with homing peptide may enhance the selectivity and efficacy of the therapy. Current efforts are being focused on tumor homing peptides that may target tumor tissues.

Information about Peptides: Tumor Homing Peptide Database has been developed using extensive literature search. It contains detailed information about the tumor targeting/homing peptides. Each entry contains following type of information about a peptide; its sequence, source, target tumor, target cell, biomarker, applications and clones. Experimental details like phage display libraries used, cell lines, *in*

TumorHPD: Designing of Tumor Homing Peptides

(Institute of Microbial Technology, Chandigarh, India)

Welcome to TumorHPD

Tumor homing peptides are the short peptides having average length between 7 to 12 residues. These peptides h
bind to tumor cells or tissues. These peptides can be used to deliver target specific drugs and as imaging agents for t
Thus prediction of tumor homing peptide is important for managing cancer treatment effectively.

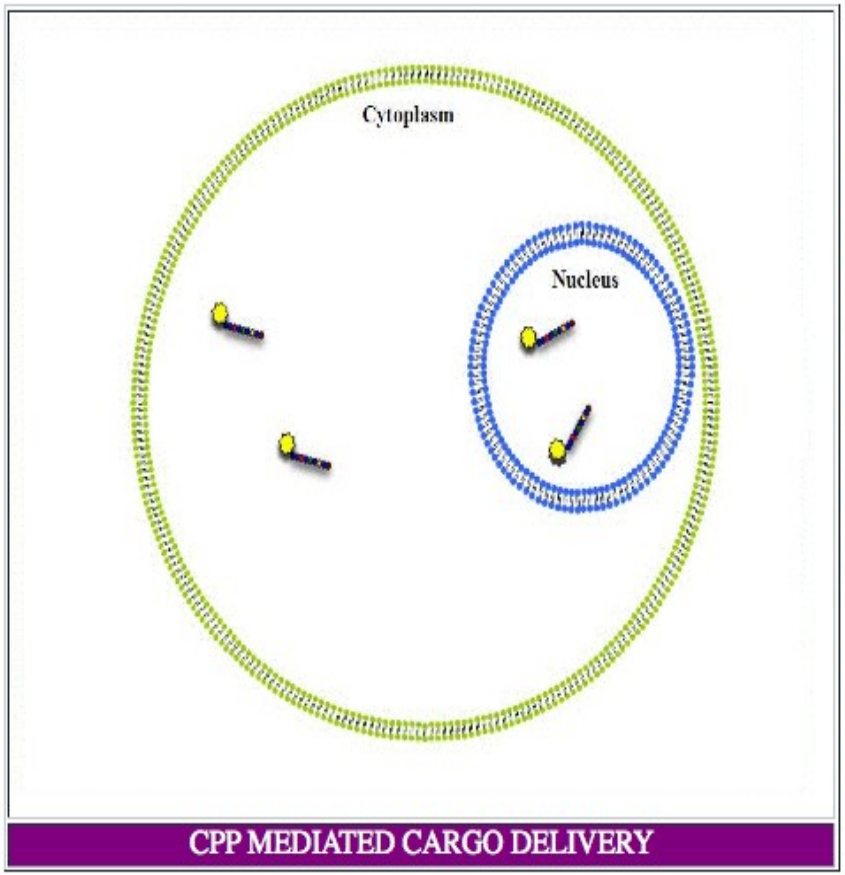
TumorHPD is a web server for predicting and designing tumor homing peptides. This server is extremely useful fo
the field of therapeutic peptides. This server allows the users to design tumor homing peptides and their mutants
and physicochemical properties.

ference: Sharma, A. et al. Computational approach for designing tumor homing peptides. Sci. Rep. 3, 1607; DOI:10.

CPPsite: a webSite for Cell Penetrating Peptides

Navigation

- Home
- Search
- Browse
- Structure
- Tools
- Important
- Help
- About us
- Contact us



CPP MEDIATED CARGO DELIVERY

CPPsite: CPPsite is a database of experimentally validated Cell Penetrating Peptides (10-30 amino acids).

Importance of CPPsite: CPPs have tremendous therapeutic applications. These are widely used to promote intracellular uptake of conjugated cargos (nucleic acids, peptide nucleic acids, proteins, drugs, liposomes etc.) and thus play role to overcome the problem of poor delivery and low bioavailability of therapeutic molecules. CPP conjugated drugs when delivered *in vivo* have s hown promising results with high efficacy. Many CPP-conjugated compounds are under clinical trials. CPPsite database provides comprehensive information on CPPs, which may be helpful to scientific community working in the area of peptide based drug discovery.

What type of information it has: CPPsite database's current version contains comprehensive information of 843 CPPs with multiple entries in terms of peptide sequence, source/origin, localization, uptake efficiency, uptake mechanism, hydrophobicity, charge etc.

Is it a manually curated database: Yes, we have collected and compiled all the information from published literature. In addition, we have also generated structural information of CPPs. We predicted tertiary and secondary structure of these peptides using PepStr and DSSP.

Work in Progress

1. Prediction of CPP 2. Designing CPP 3. Scanning in proteins

CellPPD: Designing of Cell Penetrating Peptides

Welcome to CellPPD

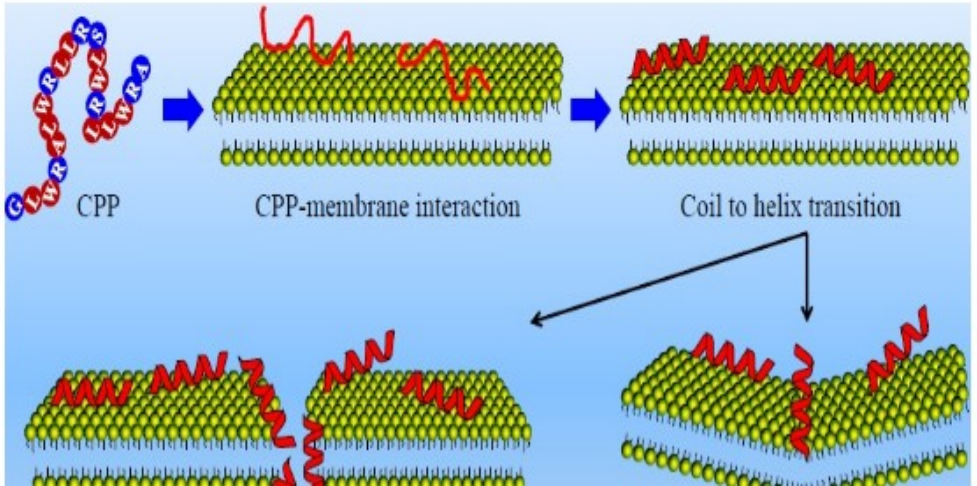
utam *et al.*: *In silico* approaches for designing highly effective cell penetrating peptides. *Journal of Translational Medicine* 13 11:74.Link

CellPPD is an *in silico* method, which is developed to predict and design efficient cell penetrating peptides (CPPs). The main dataset used in this method consists of 708 experimentally validated CPPs.

Major Features include:

(1) **Desing Peptide:** This module allows user to generate all possible single mutant analogues of their peptides and predict whether the analogue is cell penetrating or not.

(2) **Multiple Peptides:** This module of CellPPD allows user to predict number of CPPs in peptides submitted by the user

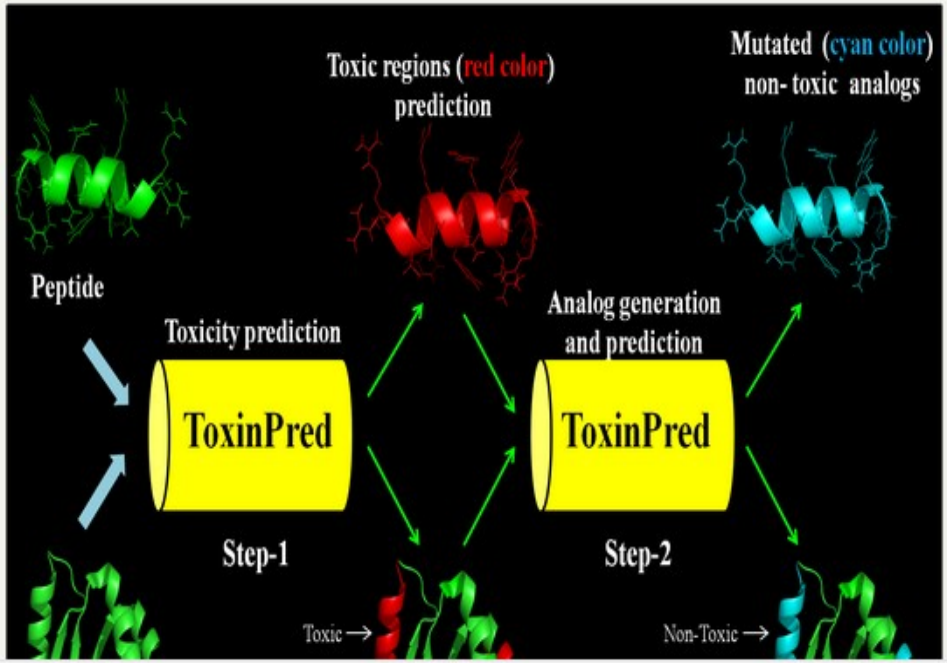


ToxinPred

Designing and prediction of toxic peptides

Welcome to ToxinPred

ToxinPred is an *in silico* method, which is developed to predict and design toxic/non-toxic peptides. The main dataset used in this method consists of 1805 toxic peptides (<=35 residues).



Major Features include:

(1) **Desing Peptide:** This module allows user to generate all possible single mutant analogs of their peptides and predict whether the analog is toxic or not.

(2) **Batch Submission:** This module of ToxinPred allows user to predict number of toxic peptides submitted by the user.

(3) **Protein Scanning:** This module generates all possible overlapping peptides and their single mutant analogs of protein submitted by the user. It also predicts whether overlapping peptide/analog is toxic or not.

(4) **QMS Calculator:** This tool allows the users to submit query peptide in FASTA format and to optimize the peptide sequence to get maximum/minimum/desired toxicity based upon the Quantitative Matrix based position specific scores. It will help the user to tweak any residue from the predecessor peptide to attain the analog with desired property (highest/lowest toxicity).

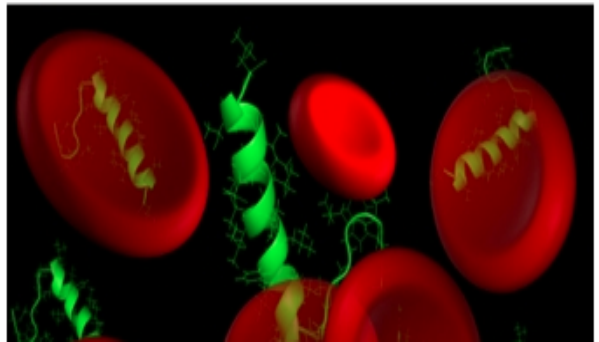
HEMOLYTIK: A Database of Hemolytic and Non-hemolytic Peptides

Welcome to Homepage

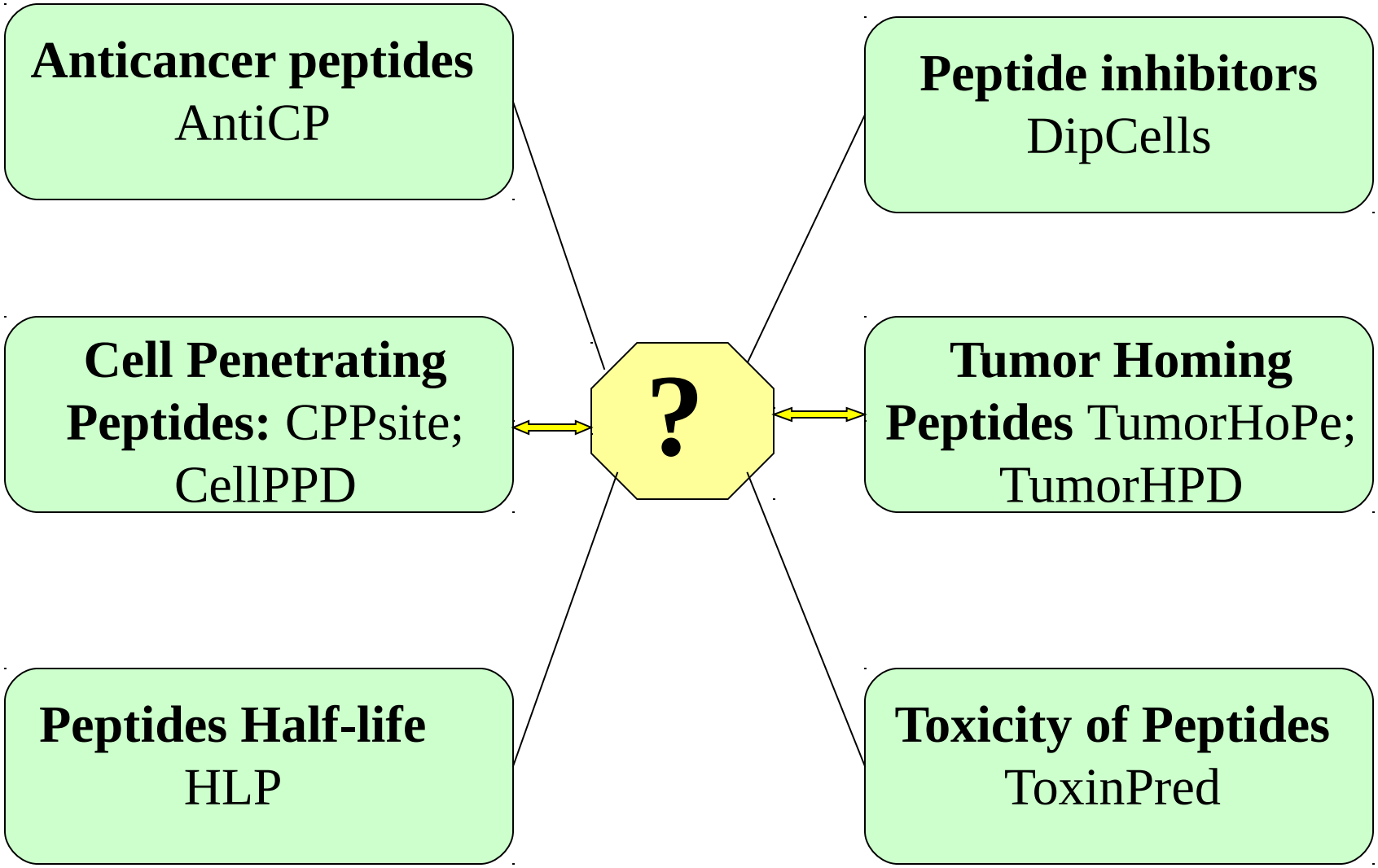
Hemolytik is a manually curated database of experimentally validated Hemolytic and Non-hemolytic peptides. In this database, peptides have been collected from both published articles as well as from other repositories like CAMP, DAMPD, APD2 and Swiss-Prot. In addition, tertiary structure of peptides have been predicted using PEPstr and secondary structure states are assigned using DSSP. First time, structure of modified peptides (containing Non-natural, D-amino acids, Modified-amino acid like Ornithine, Terminal modifications like Acetylation/Amidation) have also been predicted. In order to provide comprehensive information, peptides were searched and linked with important peptide and protein databases such as IEDB, PDB, Swiss-Prot and TrEMBL.

Major Features

(1) **Resource:** It provides comprehensive information about hemolytic peptides that include their Sequence, Name, Origin, Type (Linear/Cyclic), Chirality, End modification, Chemical modification, Source of RBCs, Hemolytic activity and Function. Data is collected from wide sources like literature and various other databases. Basic and Conditional Search facility enables the users to search a specific peptide/query in Hemolytik database.



Designing of therapeutic peptides against cancer



HerceptinR: Herceptin Resistance Database

Bioinformatics Centre, CSIR-Institute of Microbial Technology, India

- **Home**
- **Search Assays**
  - Simple Search
  - Advance Search
- **Browse Assays on**
  - Cell lines
  - Suppl. Drugs
  - Alterations in Cells
  - Publications
- **Cell Line Data**
- **Data Retrieval**
  - Mutaion Search
  - Summary of Cell lines
  - Browse on Cell Lines
  - Relative gene function
- **Web Tools**
  - Compare Genes
  - Alignment of Mutants
  - Align My Sequence
- **General Information**
  - Resistant Genes
  - Guide/Help/FAQ
  - Data downloads
  - Acknowledgement
  - Who are we?

Welcome to Home Page of HerceptinR

HerceptinR is a database of assays performed to test sensitivity or resistance of Herceptin Antibodies towards breast cancer cell lines. This database provides comprehensive information about experimental data perform to understanding factors behind herceptin resistance aas well as assays peformed for improving Herceptin sensitivity with the help of supplementary drugs. Best of our knowledge this is the first database developed to understand herceptin resistance, which can be used for designing herceptin sensitive biomarkers.

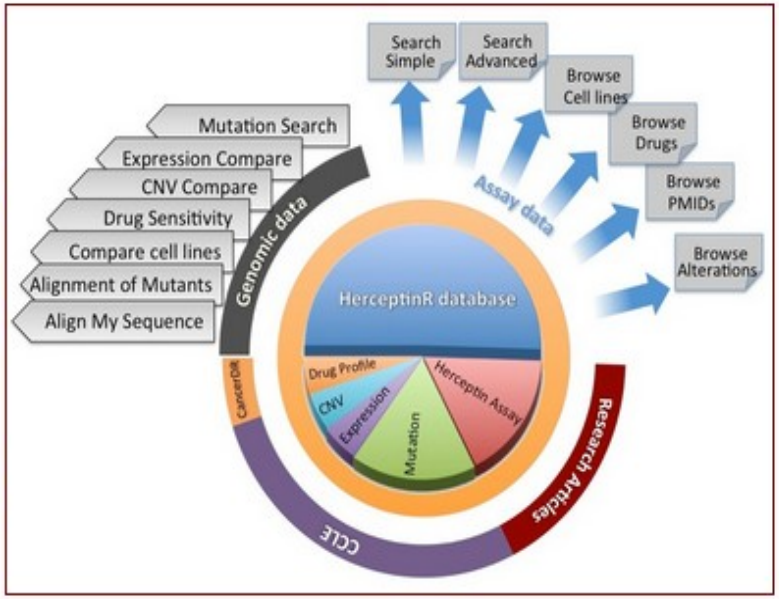
Major features of HerceptinR

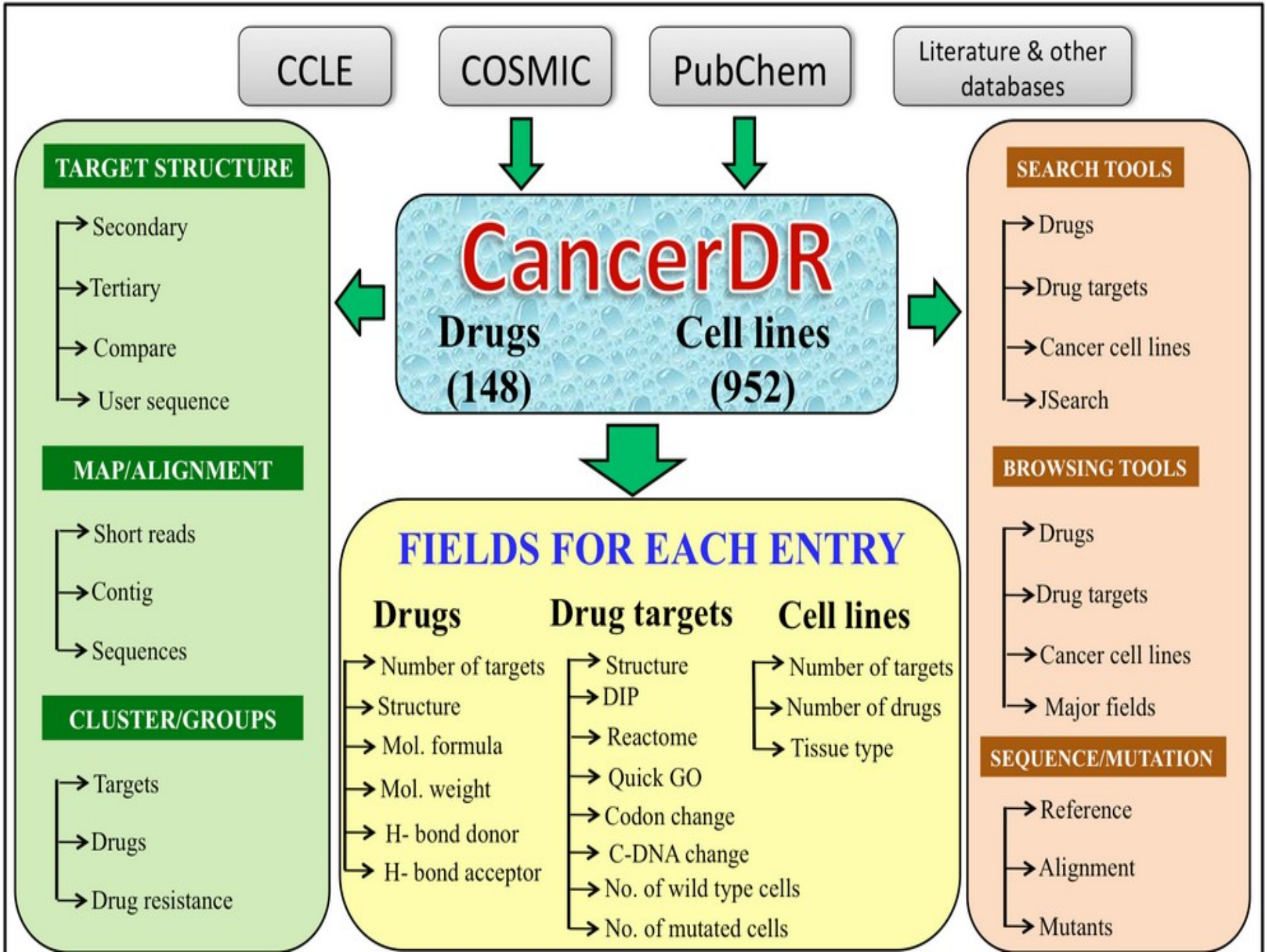
Herceptin Assays: It provides information about 2500 herceptin assays performed to test efficacy of herceptin on various breast cell lines (~ 30 unique cell lines) with and without supplementary drugs (~100 unique drugs).

Cell Line Data: This database also provides comprehensive information about breast cancer cell lines (e.g., mutation, copy number variation, expression of genes), in order to facilitate user to design herceptin biomarkers.

Important Genes: Comprehensive information about herceptin resistant genes (due to their mutation or altered expression), reported in literature.

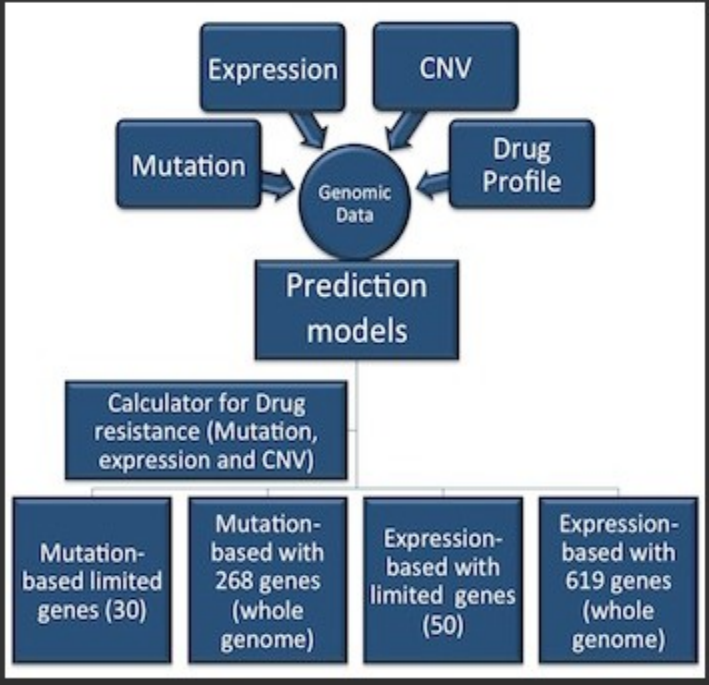
Web Tools: Number of tools have been integrated that include comparison of gene functions (expression, CNVs, mutations) of sensitive and resistant cell lines.




CCLE
COSMIC
PubChem
Literature & other databases
TARGET STRUCTURE
Secondary
Tertiary
Compare
User sequence
MAP/ALIGNMENT
Short reads
Contig
Sequences
CLUSTER/GROUPS
Targets
Drugs
Drug resistance
CancerDR
Drugs (148)
Cell lines (952)
FIELDS FOR EACH ENTRY
Drugs
Number of targets
Structure
Mol. formula
Mol. weight
H- bond donor
H- bond acceptor
Drug targets
Structure
DIP
Reactome
Quick GO
Codon change
C-DNA change
No. of wild type cells
No. of mutated cells
Cell lines
Number of targets
Number of drugs
Tissue type
SEARCH TOOLS
Drugs
Drug targets
Cancer cell lines
JSearch
BROWSING TOOLS
Drugs
Drug targets
Cancer cell lines
Major fields
SEQUENCE/MUTATION
Reference
Alignment
Mutants

CancerDRG

Cancer Drug Resistance for Genome



Drug Prioritization Prediction

Mutation1: Prediction based on limited (30) number of genes.

Mutation2: Prediction based on whole genome.

Expression: Prediction based on expression of whole genome.

Drug resistance Calculator: Interactive calculation of drug resistance based on probability. .

Introduction to CanDpred

The designing of a novel drug for cancer therapy has become progressively sophisticated nonetheless the cancers have also devised itself to resist the action of therapeutic drugs by several mechanisms. Among several such factors contributing in drug resistance, the major ones comprise mutations, expression and copy number variations. The association of these factors with drug resistance has

GPSR: A Resource for Genomics Proteomics and Systems Biology

- **A journey from simple computer programs to drug/vaccine informatics**
- **Limitations of existing web services**
  - **History repeats (Web to Standalone)**
  - **Graphics vs command mode**
- **General purpose programs**
  - **Small programs as building unit**
- **Integration of methods in GPSR**

Program	Purpose
❖ fasta2sfasta	Convert fasta format to single fasta format
❖ pro2aac	To calculate amino acid composition of protein
❖ pro2aac_nt	To calculate amino acid composition of N-terminal (nt) residues of a protein
❖ pro2aac_ct	To calculate amino acid composition of C-terminal (ct) residues of a protein
❖ pro2aac_rest.pl	To calculate amino acid composition of a protein after removing N-, and C-terminal residues
❖ pro2aac_split	To calculate split amino acid composition (SSAC) of a protein
❖ pro2dpc	To calculate dipeptide composition of protein
❖ pro2dpc_nt	To calculate dipeptide composition of N-terminal (nt) residues of a protein
❖ pro2dpc_ct	To calculate dipeptide composition of C-terminal (ct) residues of a protein
❖ pro2tpc	To calculate tripeptide composition of protein
❖ add_cols	To add columns of two files
❖ col2svm	To generating SVM_light input format
❖ col_mult	To multiplying each column of input file with a number
❖ col_mult_sel	To multiplying selective columns with a number
❖ perl col_rem	To remove selective columns from a file
❖ col_ext	To extract selective columns from a file
❖ col_corr	To compute correlation co-efficient between two column
❖ col_avg	To calculate average column of two files
❖ seq2pssm_imp	To calculate PSSM matrix in column format without any normalization
❖ pssm_n1	To normalize pssm profile based on 1/(1+e-x) formula

OSDDLINUX

Customized operating environment for drug discovery pipeline

BIOINFORMATICS

VACCINE INFORMATICS

CHEMINFORMATICS

Live Server

Live CD

Pkg Repository

Installation

Webser

Standal

Galaxy

All in

Major Features of OSDDlinux

Infrastructure for Developing World

- Infrastructure for Bioinformatics
- Assembling & Annotation of Genomes
- Computer-aided drug design
- Platform for launching services

Promoting Crowdsourcing

- Example-based learning of Web servers
- Galaxy-based platform for sharing

Network-Based Collobration

Installation of OSDDlinux

LiveDVD/USB

- **Download ISO image from web site**
- **Create bootable DVD/USB from ISO image or send request**
- **Boot your system from bootable DVD/USB**
- **Select Install option for setting OSDDlinux on your system**

Install on Existing System

1.Download base system set account and permission
2.Copy models, blastadata, webserver, galaxy from site
3.Set Vmbox on existing machine and install OSDDlinux

Download Options

- **Web-based download from web sites**
- **RSYNC for sync your data**


Thank you
wiseGEEK