OSDDlinux: Operating System for Drug Discovery

G P S Raghava, Head Bioinformatics Ce

Institute of Microbial Technology, Chandigarh, India

Bioinformatics **Drug Informatics** **Vaccine Informatics**

Email: raghava@imtech.res.in

http://osddlinux.osdd.net/

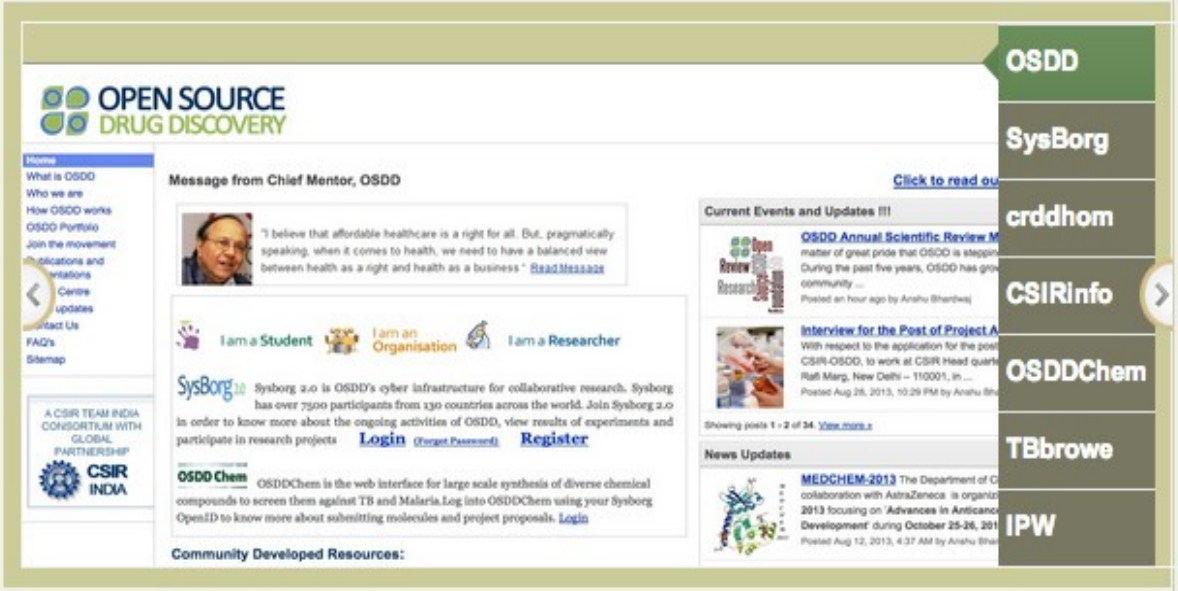
http://ww.imtech.res.in/raghava/

OSDDlinux: A Customized Operating System for Drug Discovery

General Information
Software Packages
Install/Download
OSDDLinux Online
Important Resources

Welcome to OSDDlinux

Open Source Drug Discovery (OSDD), a mission to provide affordable drugs for poors, is in the process of creating an *in silico* plateform for designing, discovering and simulating drugs. OSDD have initiate number of projects to support *in silico* drug discovery, including OSDDlinux and computational resources for drug discovery (CRDD).



Webservers

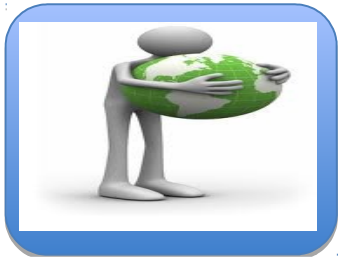
Galaxy

Standalone

All-in-one

OSDDLINUX

Customized operating environment for drug discovery pipeline



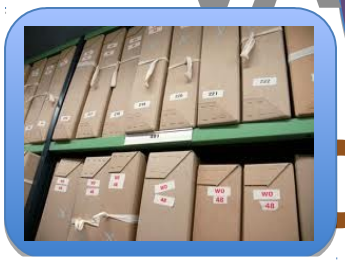
Live Server

Live CD

BIOINFORMATICS

VACCINE INFORMATICS

CHEMINFORMATICS



Pkg Repository

Installation

Webserver

Standalone

Galaxy platform

All in ONE

Computer-Aided Drug Discovery
Searching Drug Targets: Bioinformatics

Genome Annotation

FTGpred: Prediction of Prokaryotic genes
EGpred: Prediction of eukaryotic genes
GeneBench: Benchmarking of gene finders
SRF: Spectral Repeat finder

Comparative genomics

GWFASTA: Genome-Wide FASTA Search
GWBLAST: Genome wide BLAST search
COPID: Composition based similarity search
LGEpred: Gene from protein sequence

Subcellular Localization Methods

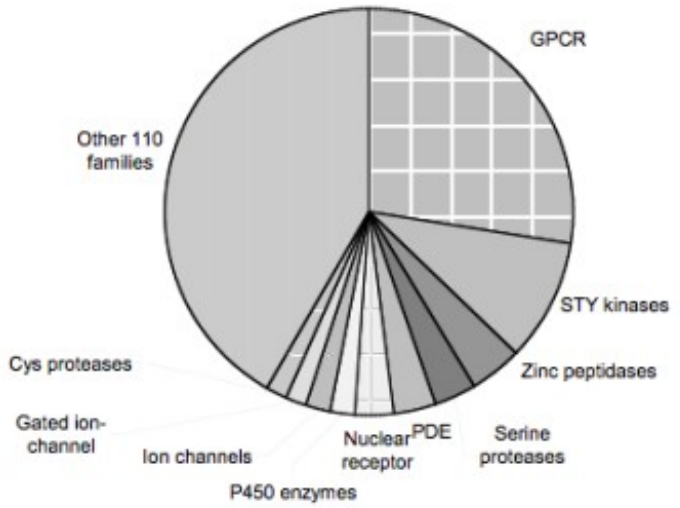
PSLpred: localization of prokaryotic proteins
ESLpred: localization of Eukaryotic proteins
HSLpred: localization of Human proteins
MITpred: Prediction of Mitochndrial proteins
Bpred: Localization of mycobacterial proteins

Prediction of drugable proteins

Nrpred: Classification of nuclear receptors
GPCRpred: Prediction of G-protein-coupled receptors
GPCRsclass: Amine type of GPCR
VGIchan: Voltage gated ion channel
Pprint: RNA interacting residues in proteins
GSTpred: Glutathione S-transferases proteins

Protein Structure Prediction

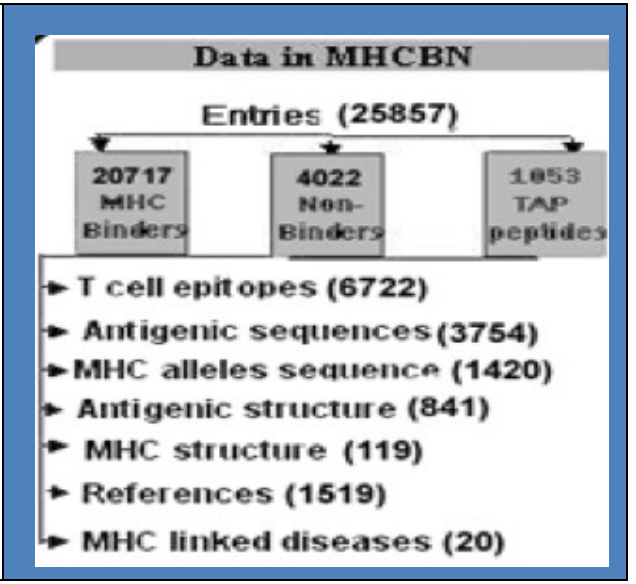
APSSP2: protein secondary structure prediction
Betatpred: Consensus method for β-turns prediction
Bteval: Benchmarking of β-turns prediction
BetaTurns: Prediction of -turn types in proteins
Turn Predictions: Prediction of α/ β/γ -turns in proteins
GammaPred: Prediction of-turns in proteins
BhairPred: Prediction of Beta Hairpins
TBBpred: Prediction of trans membrane beta barrel proteins
SARpred: Prediction of surface accessibility (real accessibility)
PepStr: Prediction of tertiary structure of Bioactive peptides



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

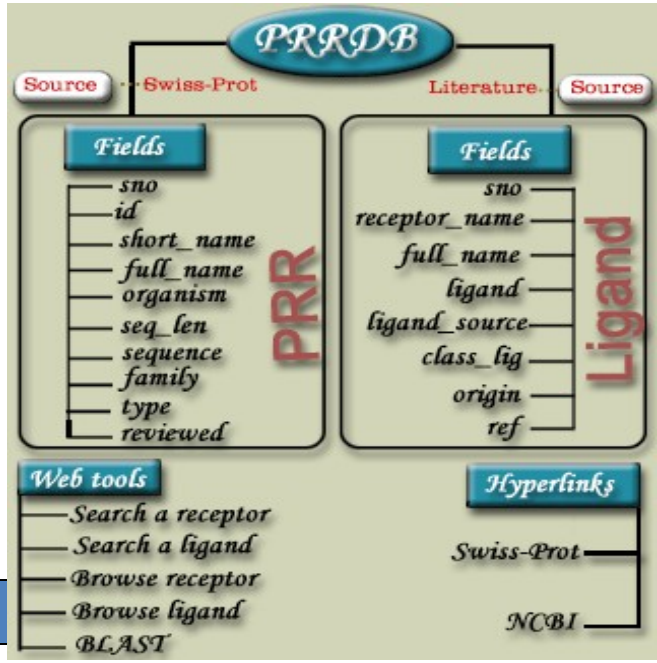
Pcleavage: for proteome cleavage sites
TAPpred: for predicting TAP binders
Propred1: for promiscuous MHC I binders
CTLpred: Prediction of CTL epitopes

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

Computational Resources for Drug Discovery

CSIR

Target Identification
Genome Annotation
Proteome Annotation
Potential Targets
Protein Structure

Virtual Screening
QSAR Techniques
Docking & QSAR
Chemoinformatics
siRNA/miRNA

Drug Design
Lead Optimization
Pharmainformatics
ADMET
Clinical Informatics

How to Contribute?
Expermentalists
Virtual Trainees/Jobs
Software Developers

Computational Resources
Library Interfaces
Meta Servers
Publishing Document
Data on *M.tb.*

**** Prediction of cytochrome P450 isoform responsible for metabolizing a drug molecule ****

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:

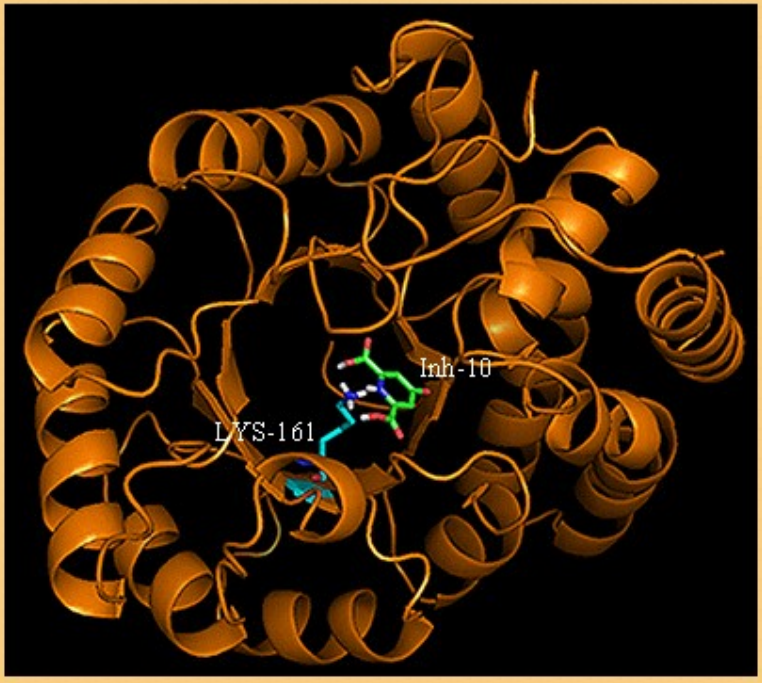
- **CRDD provides computational resources for researchers in the field of computer-aided drug design.**
- **CRDD allows users to discuss their problem with other members.**
- **CRDD gives equal opportunity to those willing to solve these problems.**
- **CRDD Wiki maintain wikipedia related to drug discovery.**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

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

KiDoQ

Prediction of inhibition constant using docking and qsar

2010) KiDoQ: using docking based energy scores to develop ligand based model for predicting antibacterials. BMC Bioinformatics 2010, 11:125 ===



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.

GDoQ

Prediction of GLMU inhibitors using QSAR and AutoDock

Home OSCADD

Inhibitor Prediction

KiDoQ
GDoQ
KetoDrug
DMKPred
TLR4HI
ABMpred
eBooster
MDRIpred
HIVFin
MetaServer (Comming Soon)

Antigenic Properties

Carbotope

ADMET Properties

MetaPred
ToxiPred
DrugMint
QED
MetaServer (Comming Soon)

Descriptors

, D., Anurag, M., Dash, D., Raghava, G. P. S. (2011). "A Web Server for Predicting Inhibitors against Bacterial Target GlmU Prot

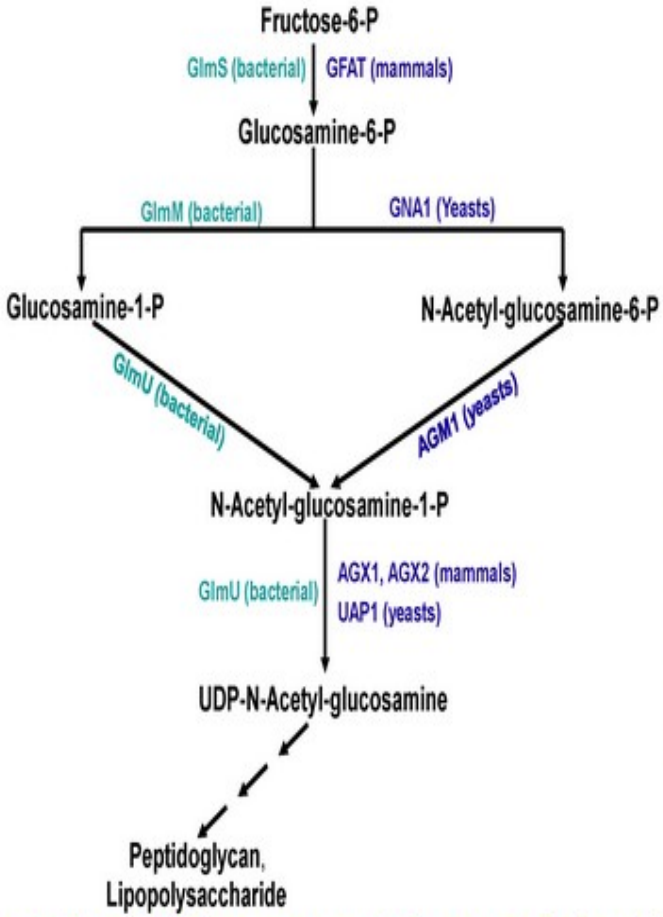


Figure shows the importance of Glmu protein in cell

A open source Webserver for prediction of Mycobacterium T inhibitors using QSAR and Docking statergies.GDoQ wet developed to serve scientific community working in the field for designing inhibitors against N-acetylgluco uridyltransferase (GLMU) protein, a potential drug target in cell wall synthesis. This server uses molecular docking and predict inhibitory activity value (IC_{50}) of chemical comp protein. The algorithm behind the server includes the docking the active site of enzyme followed by QSAR modeling where, energy based scoring values (for the best conformer), v-life d Cdk descriptors are used as input variables to QSAR base Regression (MLR) model for prediction of IC_{50} value. implemented on the server has been trained on the dive inhbitors of GLMU Protein and predict the IC_{50} value wit values of 0.79/0.62.

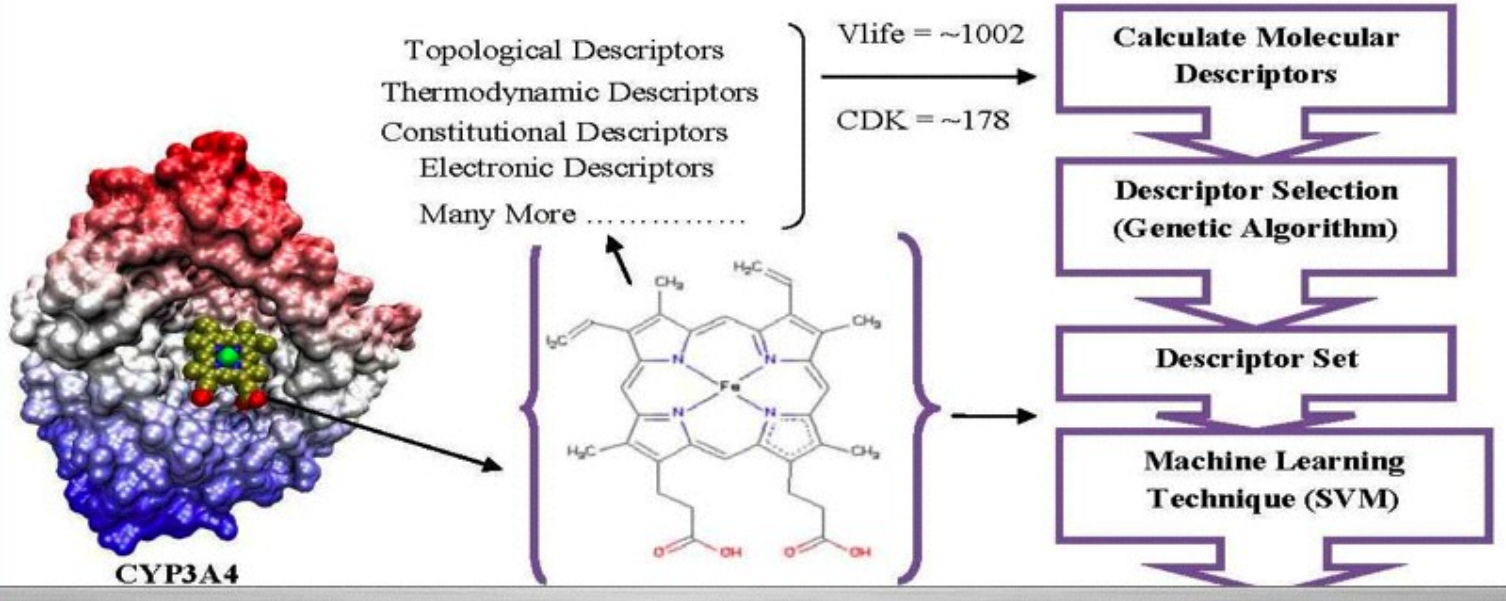
MetaPred:A webserver for the Prediction of Cytochrome P450 Isoform responsible for Metabolizing a Drug Molecule

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

» 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.

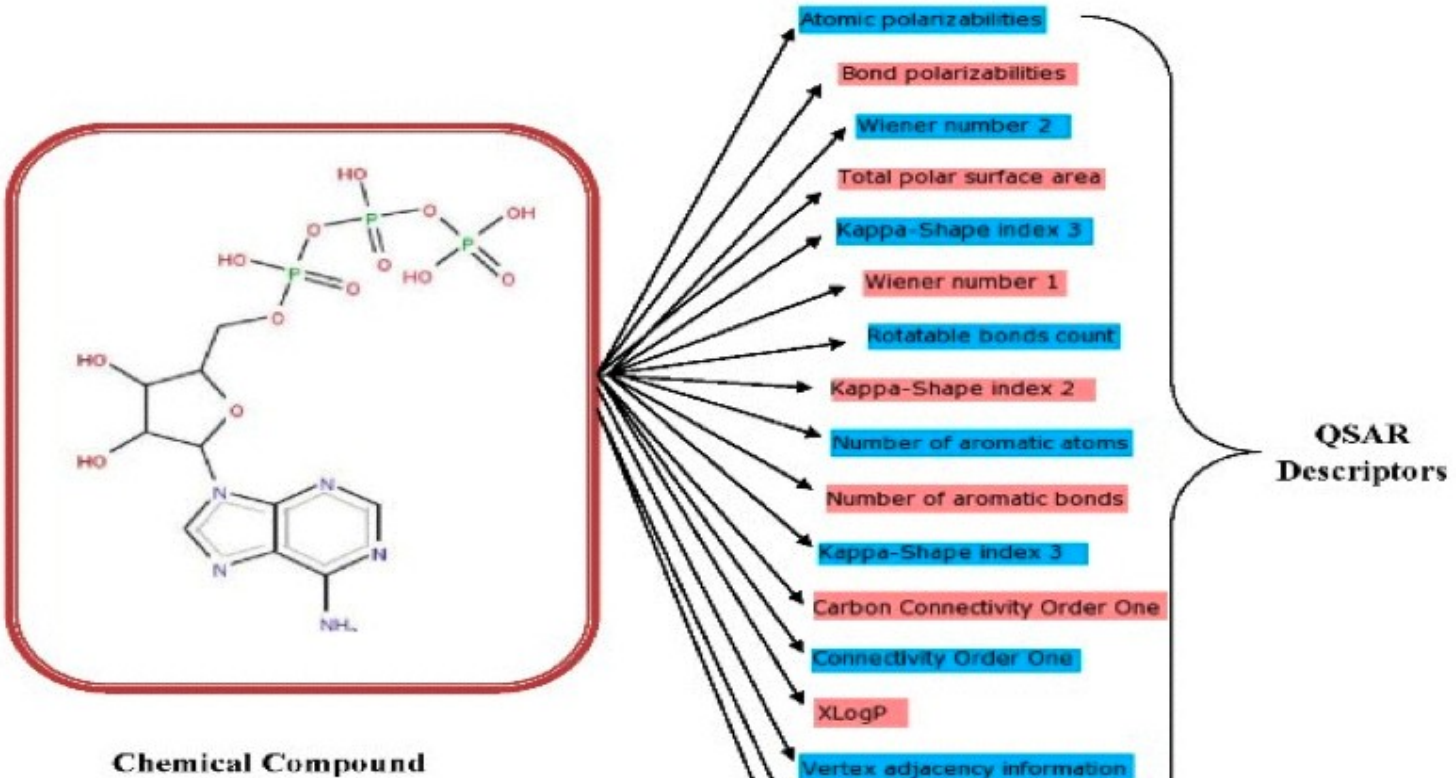
MetaPred Server predict metabolizing CYP isoform of a drug molecule/substrate, based on SVM models developed using CDK descriptors.This server will be helpful for researcher working in the field of drug discovery.This study demonstrates that it is possible to develop free web servers in the field of 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

Identification of non-toxic drug design is a major challenge in the field of drug design, most of the drug failure due to toxicity being found in late development or even in clinical trials. Thus the use of predictive toxicology is called for. Keeping this problem in view, several QSAR methods have been employed previously but we have started this study with latest dataset and apply different machine learning classifiers including non-linear method (Support Vector machine (SVM)) and linear method (Multiple linear regressions (MLR)).

Toxipred is a server where user can submit chemical molecules in the commonly used format (mol/SMILE/sdf) and after descriptors calculation our server would predict the pIGC50 value of the molecule. We hope that present model will aid in the area of drug designing.



DrugMint

A SERVER FOR IDENTIFICATION OF DRUG-LIKE MOLECULE

Home OSCADD

Inhibitor Prediction

KiDoQ
GDoQ
KetoDrug
DMKPred
TLR4HI
ABMpred
eBooster
MDRIpred
HIVFin
MetaServer (Comming Soon)

Antigenic Properties

Carbotope

ADMET Properties

MetaPred
ToxiPred
DrugMint
QED
MetaServer (Comming Soon)

Descriptors

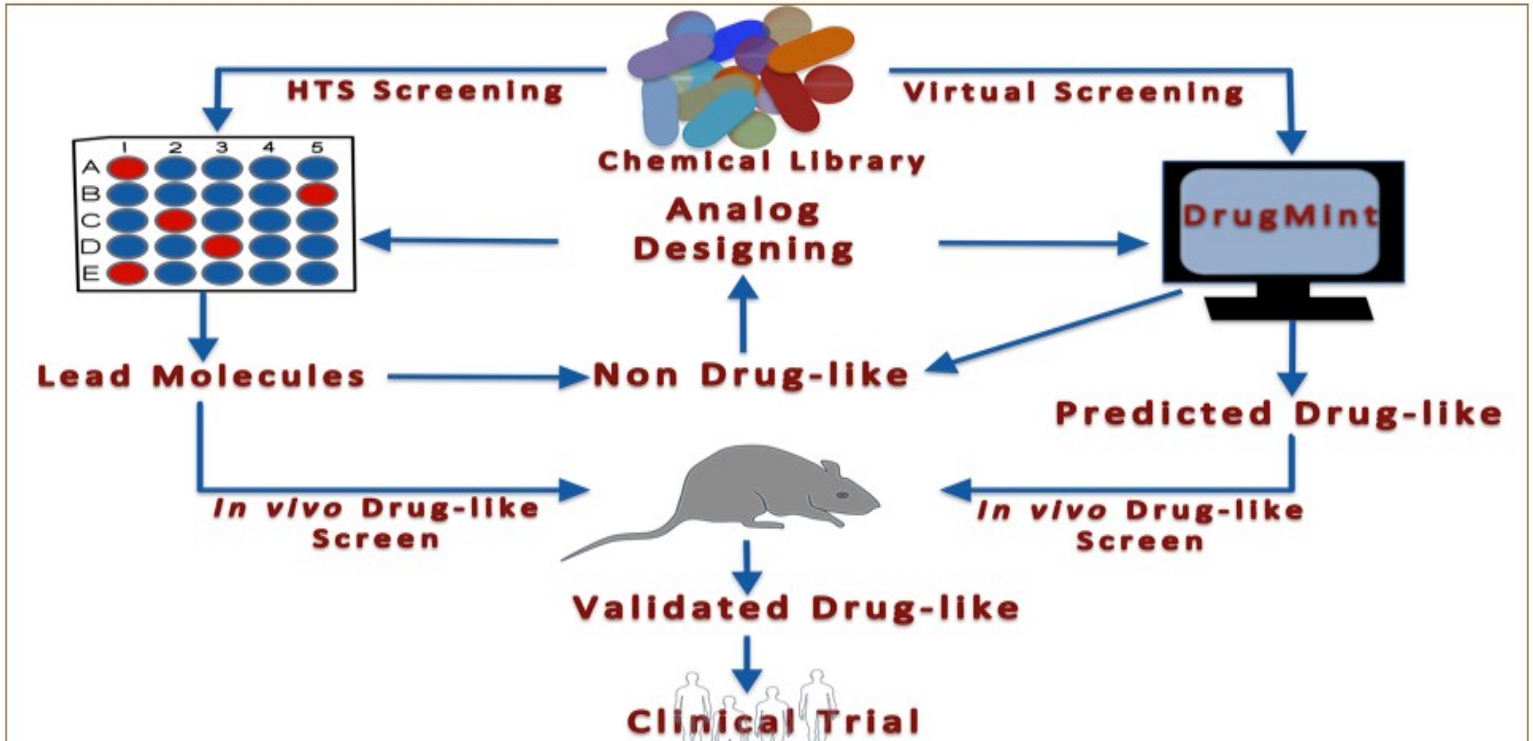
Format Conversion
WebCDK
MetaServer

Welcome to DrugMint server

DrugMint is a web server developed for predicting drug-likelihood of a compound. All models were trained, tested and evaluated on a dataset contain 1347 approved drugs obtained from DrugBank 2.5 and 3206 experimental drugs. All QSAR models were developed using open source software packages like PaDEL, WEKA, SVM_Light. Overall objective of this server is to provide service to drug development community in predicting, screening and designing drug-like molecules.

DrugMint is also available in following platform:

1. Mirror Site at http://osddlinux.osdd.net/oscadd/drugmint
2. Standalone version at http://osddlinux.osdd.net>
3. Galaxy version at http://osddlinux.osdd.net:8001



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

Home OSCADD

Inhibitor Prediction

KiDoQ
GDoQ
KetoDrug
DMKPred
TLR4HI
ABMpred
eBooster
MDRIpred
HIVFin
MetaServer (Comming Soon)

Antigenic Properties

Carbotope

ADMET Properties

MetaPred
ToxiPred
DrugMint
QED
MetaServer (Comming Soon)

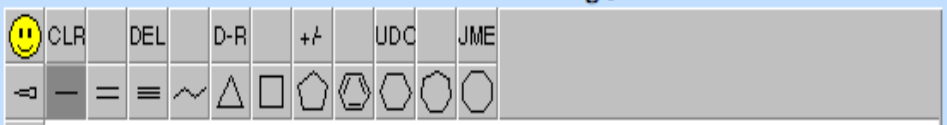
Submission Form

This server allows users to predict inhibitor against different phase of drug tolerant *M.tuberculosis*. You may submit molecules using any
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 (Optional):
(Please enter your email address if you want to r
email)

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



Method 2. **Paste structure in Mol/SMILE/SD**

use an example

(Example Test.sdf Test.mol Test.smi)

Limitations of existing web services

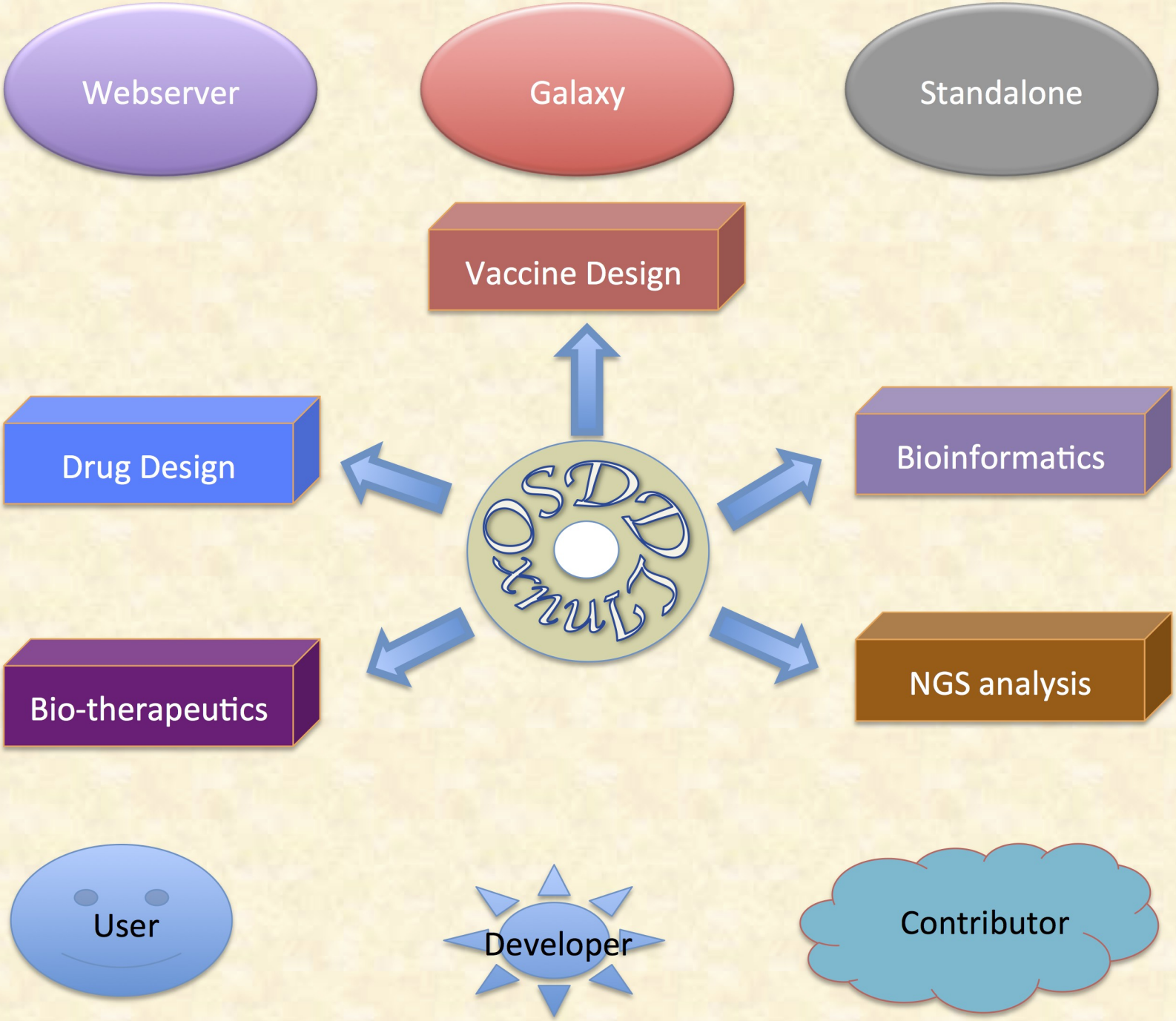
- Uploading or downloading large data
- Serving too many user from single source
- Difficult to provide computer intensive job
- Depend on internet and its bandwidth
- Security of data in transition
- Maintain confidentiality of data
- Difficult to analyze graphical data

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

Title	Description
	pro2aac (To calculate amino acid composition of protein) The amino acid composition in a protein is simply the percentage of the different amino acids represented in a particular protein. The aim of calculating the composition of proteins is to transform the variable length of protein sequences to fixed length feature vectors. In addition the conversion of a protein sequence to a vector of 20 dimensions using amino acid composition will encapsulate the properties of the protein into the vector. The composition of all 20 natural amino acids were calculated by using the following equation $$\text{Composition of amino acid } i = \frac{\text{Total number of amino acid } i \times 100}{\text{Total number of all amino acids in protein}}$$ Where i can be any amino acid
Usage	*pro2aac -i seq.sfa -o seq.out*
-i	Input file name contains single fasta format
-o	Output file name gives amino acid composition
seq.sfa	>seq_1##MRNRGFGRRELLVAMAMLVSVTGCARHASGARPASTTLPAGADLADRFAEL ERRYDARLGVYVPATGTTAAIE >seq_2##ACGRGFGVKLACNMNNACRTYFSDVAMAMLVSVTGCARHASGARPASTTL PAGADLADIEYRADERFAFCSTF
seq.out	# Amino Acid Composition of proteins # A , C , D , E , F , G , H , I , K , L , M , N , P , Q , R , S , T , V , W , Y, 19.18, 1.37, 4.11, 5.48, 2.74, 9.59, 1.37, 1.37, 0.00, 9.59, 4.11, 1.37, 4.11, 0.00,13.70, 4.11, 8.22, 6.85, 0.00, 2.74, 19.18, 6.85, 5.48, 2.74, 6.85, 8.22, 1.37, 1.37, 1.37, 5.48, 4.11, 4.11, 2.74, 0.00, 8.22, 6.85, 6.85, 5.48, 0.00, 2.74,
Vector	20 dimension (i.e 20 types of amino acid composition is generated)


Webserver
Galaxy
Standalone
Vaccine Design
Drug Design
Bioinformatics
OSDD
Bio-therapeutics
NGS analysis
User
Developer
Contributor

Major Features of OSDDlinux

Service for Scientific Community

Online for Occasional Users

LiveDVD/USB on local computer with Data Security

Platform for Developers

Developers with no infrastructure

Infrastructure on existing linux setup

Linux for Students

Bootable LivedDVD for occasional learning

OSDDlinux on Windows/MAC Users

Online hand-on experience on Linux & PERL programming

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

Download base system set account and permission

Copy models, blastadata, webserver, galaxy from site

Set Vmbox on existing machine and install OSDDlinux

Download Options

Web-based download from web sites

- RSYNC for sync your data

OSDDLinux : A Platform for Open Source Drug Discovery

Introduction

OSDD-Linux integrates open source softwares, libraries , workflows and webservices for creating environment for drug discovery. First attempt made to customize linux to provide services to community of drug discovery. OSDD-Linux may bring down cost drug discovery.

Features

- A single platform for bioinformatics and cheminformatics.
- Tools available in three formats e.g. webservers, standalone and galaxy.

Softwares

GPSR packages , webservers, standalone, galaxy versions of tools developed in Dr.Rahghava's lab along with third party

python are integrated

Features

- User can **customize** according to need.
- Easy to install and free of cost.
- It can be launched using LiveCD, Liive server, USB and virttualldesk top like virtualbox.
- Provides source codes

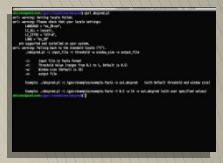
tools.

Webservers

A separate apache runs all web-servers as local host in the OSDD-Linux CD.

Standalone

Command line tools have been integrated for analysing large scale data.



Third party softwares

Open source softwares used in bioinformatics and cheminformatics have been integrated.

Galaxy

All softwares are integrated in galaxy servers for making the workflows.

System Requirements

The user may use OSDDLinux from portable devices like CD/DVD, USB drive etc or install on local machine or virtual machine depending upon requirement.

Future directions

More webservices, modules, debian packages will be updated on regular basis.

Drug Discovery manual

Installation Guide

GPSR manual

Required Package

Users Guide

License

Upgrade/New Packages

Full Installation

On existing machine

Live DVD/USB

Virtual Box

Package Repository

Platform for Developers

Linux for Students

Infrastructure for Developing World

Promoting Crowdsourcing in Drug development

Network-Based Coloration

Bioinformatics

GPSR Basic Scripts

Education & Research

Analysis of NGS data

Vaccine Informatics

Drug Informatics

Bio therapeutics

Open Source Drug Discovery (OSDD), a mission to provide affordable drugs for poors, is in the process of creating an in silico platform for designing,

workflows and web services in linux

HOME | ABOUT | FEEDBACK | FAQ | MIRROR SITE | LICENSE | REGISTER | THANKS | TEAM | RAGH

OSDD LINUX

A Custo
fo

http://osddlinux.osdd.net/

- General Inoformation
- Software Packages
- Install/Download
- Service to Community
- OSDDlinux Online
- Important Resources

Install/Download

- Live DVD/USB
- Full Installation
- Virtual Box
- On existing machine
- Package Repository
- Upgrade/New Packages

General Inoformation

- Major Features
- Installation Guide
- Users Guide
- Drug Discovery manual
- GPSR manual
- Required Package
- List of Packages

Service to Community

- Command Mode
- Web Services
- Galaxy Portal
- GUI-based Software

Software Packages

- Bioinformatics
- Vaccine Informatics
- Drug Informatics
- Biotherapeutics
- Analysis of NGS data
- Education & Research
- Basic Scripts