Data intensive Research in the field of Biological Sciences

Dr G P S Raghava, Head Bioinformatics Centre

Institute of Microbial Technology, Chandigarh, India

Bioinformatics Drug Informatics Vaccine Informatics Chemoinformatics

Email: raghava@imtech.res.in http://crdd.osdd.net/

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

ponential Growth of Data: Source of Knowled

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/)

OMIC tools

Home | Reviews | News | FAQ | Media | About | Submit tools | Search

Browse by omic applications

- Sequencing (2277)
- Microarray (507)
- Mass spectrometry (349)
- NMR spectroscopy (98)
- PCR (113)
- nCounter System (4)
- Bioimaging (101)
- Cytometry (68)
- Common omic tools (290)

Browse by functional analysis

- Drug discovery (419)
- Functional analysis (1368)
- Epigenomics (35)
- Genome editing (38)
- Transcriptomics (354)
- Proteomics (369)
- Gene prioritization (16)
- Fluxomics (49)
- Metabolic engineering (25)
- Biomolecular structure (391)
- Health & Diseases (205)
- Immunology (80)
- Educational resources (198)

bioinformatics.ca
links directory

Computer Related (83)

Education (75)

Human Genome (238)

Model Organisms (378)

Protein (1009)

Expression (395)

DNA (604)

Other Molecules (378)

Literature (86)

RNA (203)

Sequence Comparison (271)

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

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

Chemoinformatics and Pharmacoinformatics

Web Server	Description
DrugMint	A Server for Identification of Drug-like Molecules
ABMPred	Prediction of AntiBacterial Compounds against MurA Enzyme
MDRIpred	Prediction of Inhibitor against Drug Resistant M.Tuberculosis
DMKpred	Prediction of Drug molecules for kinase protein
KiDoQ	Prediction of inhibition constant of a molecule against Dihydrodipicolinate synthase enzyme
TOXIpred	Prediction of aqueous toxicity of small chemical molecules in T. pyriformis.
MetaPred	Prediction of cytochrome P450 isoform responsible for metabolizing a drug molecule.
GDoQ	Model for prediction of GLMU inhibitors using QSAR and docking apprach.
KetoDrug	Binding affinity prediction of ketoxazole derivatives against fatty acid amide hydrolase.
WebCDK	Web Interface for CDK libraries
TLR4HI	SVM based model for computing inhbitors against human TLR4 (Toll like receptor).
DMKPred	A webserver for the prediction of binding of chemical molecules with specific kinases.
ntEGFR	Predicting and designing imidazothiazoles/pyrazolopyrimidines based inhibitors against wild/mutant EGFR.
CancerIn	Classification and designing of anti-cancer inhibitors.
EGFRpred	Prediction of inhibitor of anti-EGFR molecules of diverse class.
DiPCell	Designing of inhibitors against pancreatic cancer cell lines.
HIVfin	Prediction of fusion protein inhibitors against HIV.

Molecular Interactions

Software name	Description
ADPint	Prediction of ADP interacting residues in a protein.
ATPint	Identification of ATP binding sites in ATP-binding proteins.
DOMprint	SVM based model for predicting domain-domain interaction (DDI).
GlycoEP	Prediction of C-, N- and O-glycosylation site in eukaryotic proteins.
GlycoPP	Prediction of potential N-and O-glycosites in prokaryotic proteins.
GTPbinder	Identification of GTP binding residue in protein sequences.
MYCOprint	A tool fort exploration of the interactome of Mycobacterium tuberculosis.
NADbinder	Prediction of NAD binding proteins and their interacting residues.
Pprint	ANN based method for identification of RNA-interacting residues in a protein.
PreMieR	Identification of mannose interacting residues (MIRs) in protein sequences.
PROprint	Prediction of physical/functional interaction between two protein molecules.
RNApin	A server for the prediction of protein interacting nucleotides in RNA sequences.
tRNAmod	Prediction of post transcriptional modifications in transfer-RNA (tRNA) sequence.
VitaPred	Identification of different class of vitamin interacting residues in a protein.

Biological Databases

Database name	Description
MHCBN	A curated database of MHC-binding, Non-binding peptides and T-cell epitopes.
Bcipep	A database of B-cell epitopes.
HaptenDB	A database of hapten molecules that can not activate immune system.
PolysacDB	Compilation of antigenic polysaccharides found on surface of microbial organism.
TumorHope	A database of experimentally characterized tumor homing peptides.
AntigenDB	Information about a wide range of experimentally-validated antigens.
OXDBase	Compilation of oxygenases involved in the biodegradation on xenobiotic compounds.
HMRBase	A manually curated database of hormones and their Receptors.
CPPsite	Compilation of experimentally validated Cell Penetrating Peptides (10-30 amino acids).
BIAdb	Information about Benzylisoquinoline Alkaloid molecules
HIVsir	A manually curated database of anti-HIV siRNAs.
CCDB	Catalog of genes involved in the different stages of cervical carcinogenesis.
ProGlycProt	Repository of experimentally characterized eubacterial and archaeal glycoproteins.
NPACT	A database of plant derived natural compounds that exhibit anti-cancerous activity.
CancerDR	Compilation of anticancer drugs and their effectiveness against various cancer cell lines.
ccPDB	Compilation and creation of datasets from PDB for structural/functional annoation of proteins.
ParaPep	HIPdb is a manually curated database of experimentally validated antiparasite peptides.
EGFRindb	Collection of EGFR inhibitors from literature.
CancerPPD	Collection and compilation of experimentally validated anticancer peptides
PCMdb	Pancreatic cancer methylation database provides large scale collection of metylated genes.
HerceptinR	Information about assays performend to test sensitivity/resistance of Herceptin Antibody.
HemolytiK	A resource of experimentally tested hemolytic peptides.
CancerTope	A database of epitopes found in protein involved in cancer.
AHTPDB	AHTPDB is an ideal platform for complete & relevant information for large number of antihypertensive peptides

Genome Annotations

Server Name	Description
FTG	Locating probable protein coding region in nucleotide sequence using FFT based algorithm.
GWBLAST	Genome wide similarity search using BLAST
GWFASTA	Genome Wise Sequence Similarity Search using FASTA.
EGPred	Prediction of gene (protein coding regions) in eukaryote genomes that includes introns/exons.
SVMgene	SVM based approach to identify the protein coding regions in human genomic DNA.
SRF	Find repeats through an analysis of the power spectrum of a given DNA sequence.
MyPattern	A program for detection of a 'motif' in DNA sequence using an exact search method.
GeneBench	A suite of datasets and tools for evaluating gene prediction methods.
FTGPred	A web server for predicting genes in a DNAsequence.
PHDcleav	Prediction of Human Dicer cleavage sites.
PolyApred	Prediction of polyadenylation signal (PAS) in human DNA sequence.
siRNAPred	Predicting actual efficacy of both 21mer and 19mer siRNAs with high accuracy.
ECGPred	Analsis of expresion data and correlation between gene expression and nucleotides composition of genes.
desiRam	Designing of highly efficient siRNA with minimum mutation approach
MARSpred	Discriminating between Mitochondrial and Cytosolic Aminoacyl tRNA Synthetases
Icaars	Identification & Classification of Aminoacyl tRNA Synthetases.
LGEpred	Prediction of correlation between amino acid residue and gene expression level.

Immunoinformatics or Vaccine Informatics

Software name	Description
T-Helper Epitopes or MHC/HLA Class II binders (Adaptive Immunity, Exogenous Antigen)	
MHCBN	A database of MHC-Binding, Non-binding peptides and T-cell epitopes.
ProPred	Identification of promiscuous MHC Class-II binding regions in an antigen sequence
HLA-DR4Pred	Identification of HLA-DRB1*0401(MHC class II alleles) binding peptides.
MHC	Matrix Optimization Technique for identification of binding core in MHC II binding peptides
MHC2pred	The MHC2Pred is an SVM based method for prediction of promiscuous MHC class II binding peptides.
MHCBENCH	Benchmarking of MHC binding peptide prediction algorithms.
FDR4	Prediction of binding affinity of HLA-DRB*0401 binders in an antigenic sequence.
IL4pred	In silico platform for designing and disovering of interleukin-4 inducing peptides.
IFnepitope	Designing of interferon-gamma inducing epitopes.
CTL Epitopes or MHC/HLA Class I binders (Adaptive Immunity, Endogenous Antigens)	
PROPRED1	Prediction of promiscuous binders for 47 MHC/HLA class I alleles using quantitative matrices;
Pcleavage	Identification of protesosomal cleavage sites in a protein sequence.
TPPred	Prediction of TAP binding peptides for understanding of peptide internalization to endoplasmic reticulum
CTLPred	A direct method for prediction of CTL epitopes.
nHLApred	This is a comprehensive method for prediction of MHC binding peptides or CTL epitopes of 67 MHC class alleles.
MMBPred	Prediction of mutated MHC class I binders in an antigen, having high affinity and promiscuousity.
HLAPRED	The method can identify and predict HLA (both class I & II) binding regions in an antigen sequence.
Linear & Conformational B-cell Epitopes	
BCIPEP	Collection & compilation of B-cell epitopes from literature
BCEPRED	Prediction of linear B-cell epitopes, using Physico-chemical properties
ABCPred	Mapping of B-cell epitope(s) in an antigen sequence, using artificial neural network.

Functional Annotation of Proteins

Server name	Description
NRpred	Prediction and classification of nuclear receptors, SVM models based on composition.
GPCRpred	Prediction of families and superfamilies of G-protein coupled receptors (GPCR)
ESLpred	Subcellular localization of the eukaryotic proteins using dipeptide compostion and PSI-BLAST.
PSLPred	Prediction of subcellular localization of bacterial proteins
BTXPred	It predicts bacterial toxins and their function from primary amino acid sequence.
GPCRsclass	This webserver predicts amine type of G-protein coupled receptors
Mitpred	Specifically trained to predict mitochondrial proteins with high accuracy
Oxypred	Classification and prediction of oxygen binding proteins.
VGIchan	Classification and prediction of proteins involved in voltage gated ion channels.
HSLpred	Subcellular localization of human proteins with high accuracy
DNAsize	Compute length of DNA or protein fragments from gel using a graphical method.
GSTpred	SVM-based method for predicting Glutathione S-transferase protein.
Mango	A server for predicting functional class of a protein.
LGEpred	Calculate correlation coefficient between amino acid residue and gene expression level.
NTXpred	Identification of neurotoxins their source and function from primary amino acid sequence.
VICMpred	Classification of bacterila proteins particularly virulent proteins
ALGpred	Prediction of allergenic proteins and mapping of IgE epitopes in antigens.
PseaPred	Prediction of proteins secreted by Malarial Parasite P. falciparum into infected-erythrocyte.
RSLPred	SVM based method for subcellular localizaton of rice proteins.
COPid	Composition based identification and classification of proteins.
ESLPred2	Advanced method for subcellular localization of eukaryotic proteins.
ISSpred	Identification of Inteins hiding in their protein sequences.
CyclinPred	CyclinPred is a SVM based prediction method to identify novel cyclins.

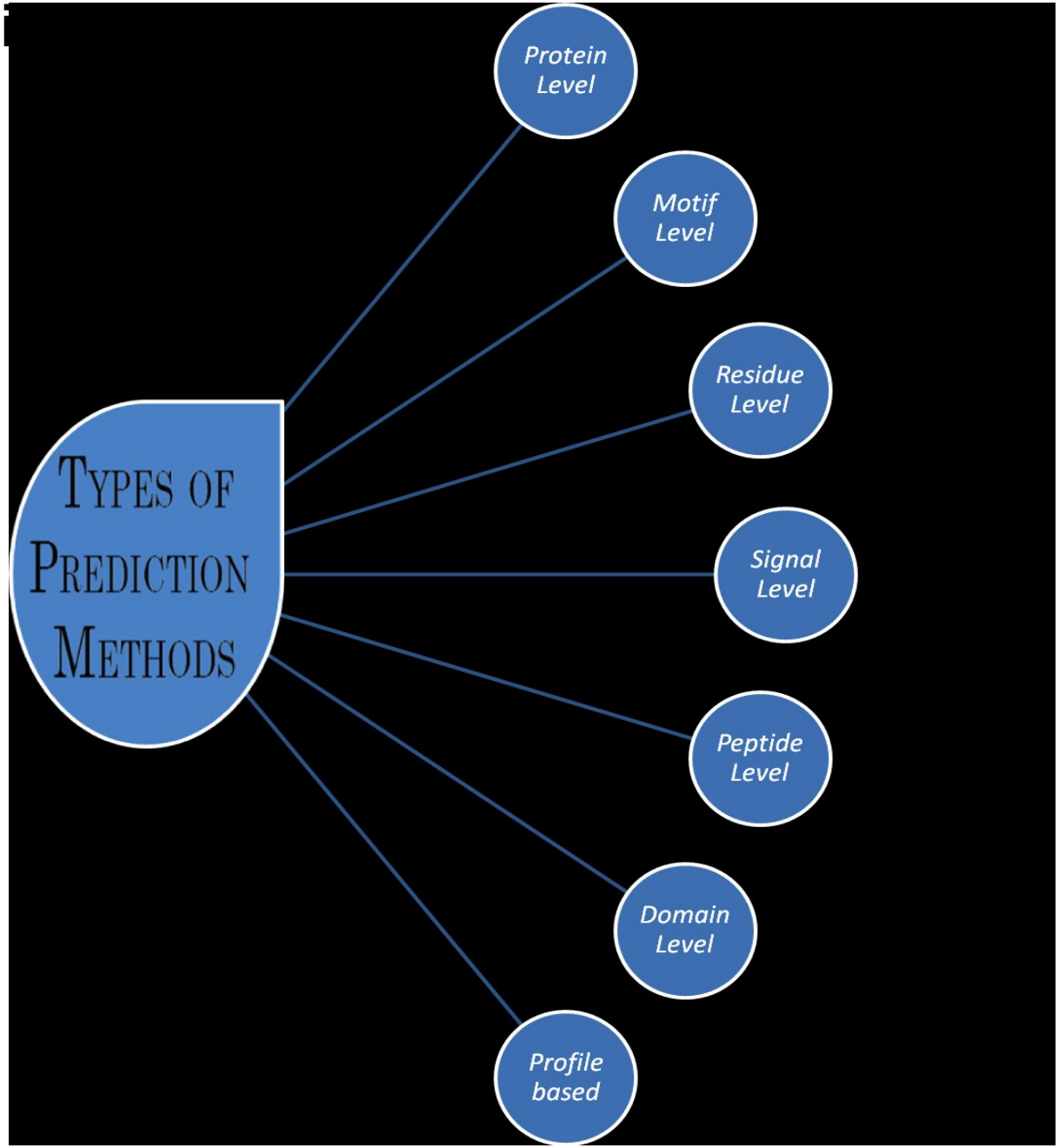
Proteins Structure Prediction

Web Server	Description
AlphaPred	A neural network based method for predicting alpha-turn in a protein.
APSSP2	Prediction of secondary structure of proteins from their amino acid sequence.
AR_NHPred	Identification of aromatic-backbone NH interaction in protein residues.
BetatPred	Statistical-based method for predicting Beta Turns in a protein.
Betatpred2	Prediction of Beta-turns with high accuracy using multiple sequence alignment.
BetaTurns	It predict different types of beta-turns (e.g., Types I/II/IV/VIII) in a protein.
BhairPred	Prediction of beta hairpins in proteins using ANN and SVM techniques.
CHpredicts	Prediction of CH-O, CH-PI interactions in backbone residues of a protein
GammaPred	Identification of gamma-turn containing residues in a given protein sequence.
PEPstr	Prediction of tertiary structure of small peptides (7 to 25 residues).
Proclass	Classification of proteins based on secondary structure contents.
PSA	Analyze the amino acid sequence and multiple sequence alignment of proteins.
RPFOLD	A fold recognition server for searching protein fold in PDB.
SARpred	ANN-model for redicting real-value of surface acessibility of protein residues.
TBBpred	This server predict Transmembrane Beta Barrel regions in a protein.
PEP2D	This server allows you to predict secondary structure of peptides.

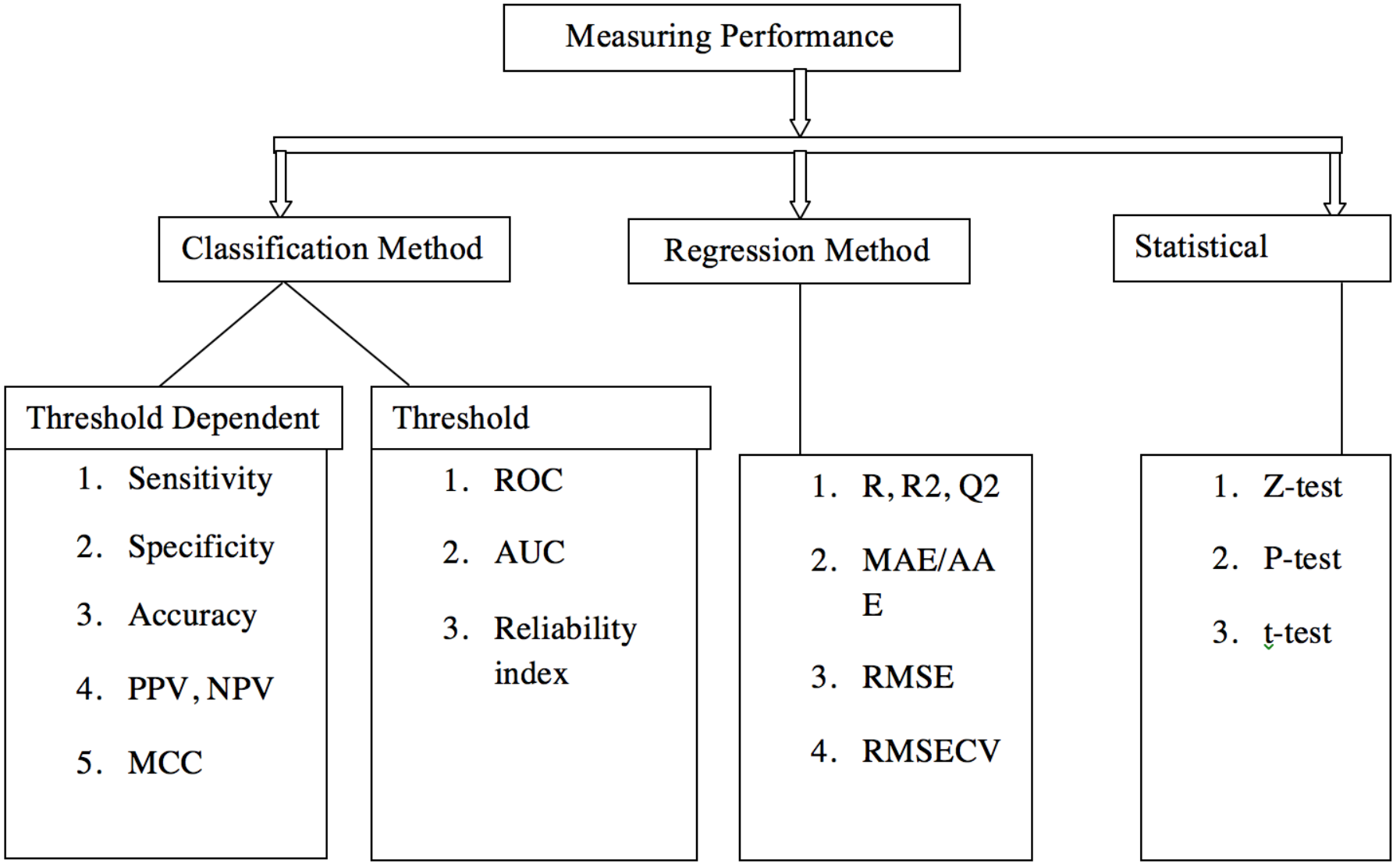
GPSR: A Resource for Genomics Proteomics and Systems Biology

- **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**

Types of Predicti



Measuring Performance



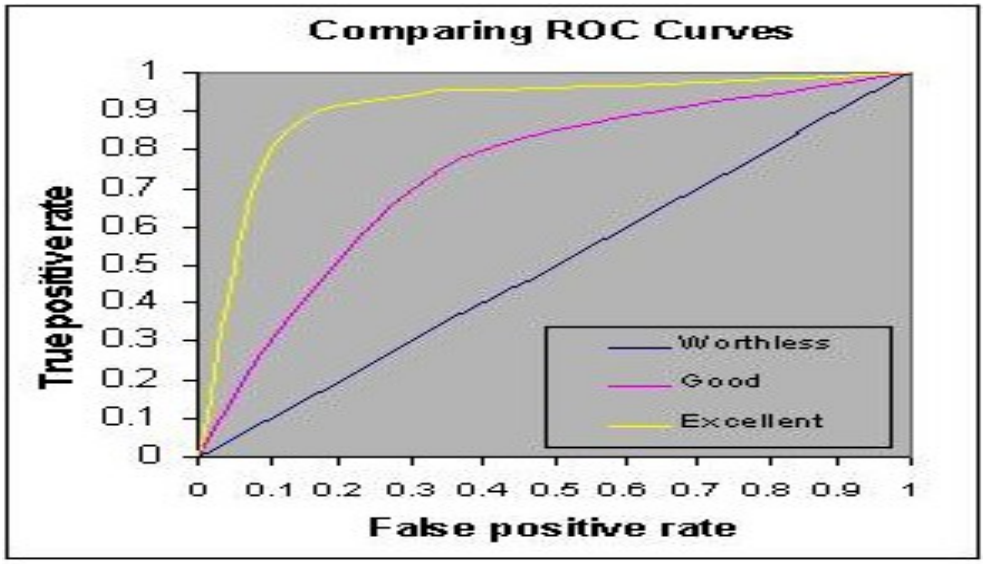
$$\text{Senstivity} = \frac{TP}{TP+FN} \times 100$$

$$\text{Specifity} = \frac{TN}{TN+FP} \times 100$$

		Actual		
		Positive	Negative	
Predicted	Positive	TP	FP	**PPV**
	Negative	FN	TN	**NPV**
		Sensitivity	**Specificity**	

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100$$

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \times 100$$



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)

GPSR

A resource for genomics, proteomics and systems biology

OSDDLINUX

Customized operating environment for drug discovery pipeline

BIOINFORMATICS

VACCINE INFORMATICS

CHEMINFORMATICS

Live Server

Live CD

Pkg Repository

Installation

Webserver

Standalone

Galaxy platform

All in ONE

Major Features of OSDDlinux

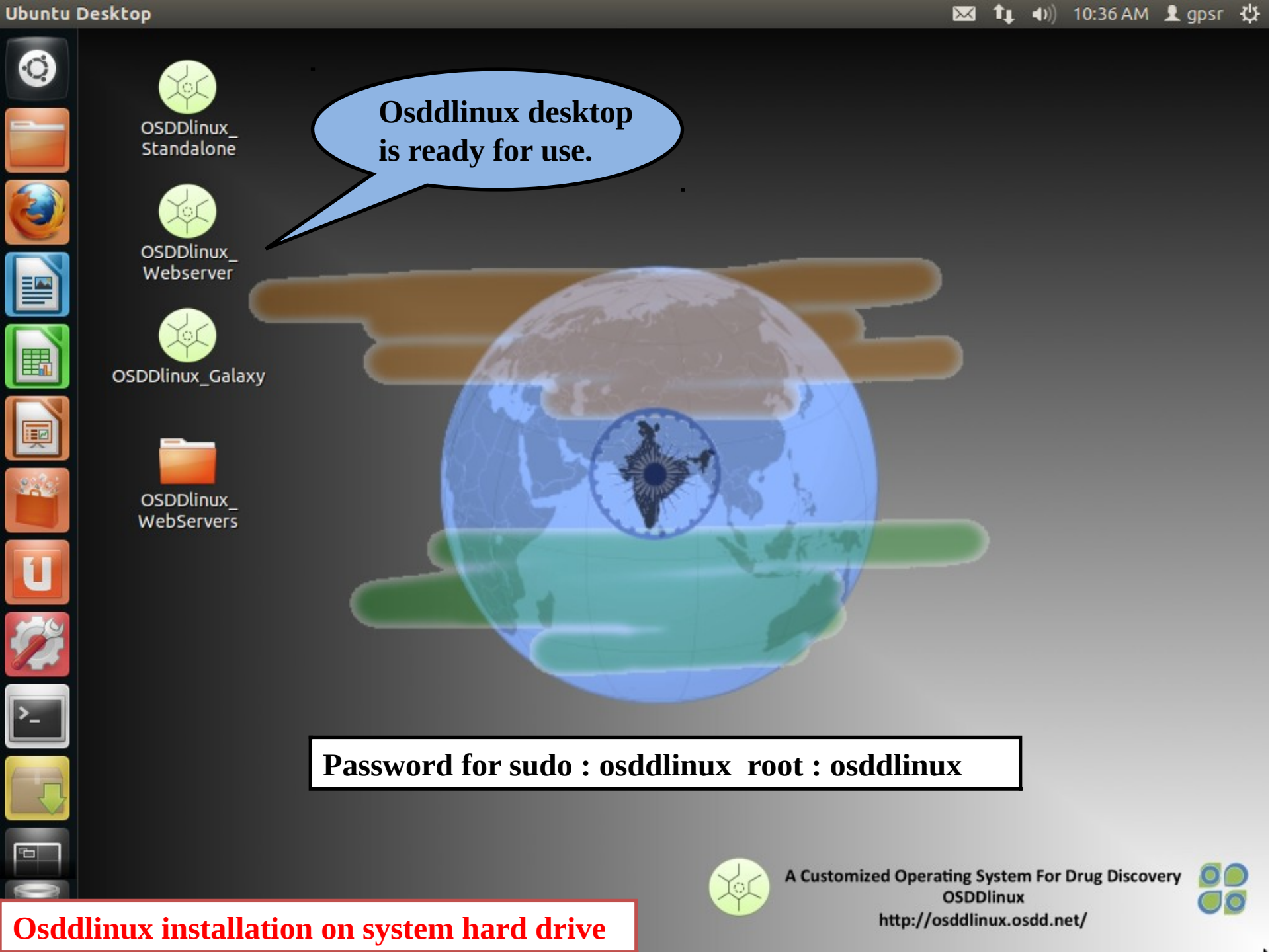
Service for Scientific Community
Online for Occasional Users
LiveDVD/USB on local computer with Data Security

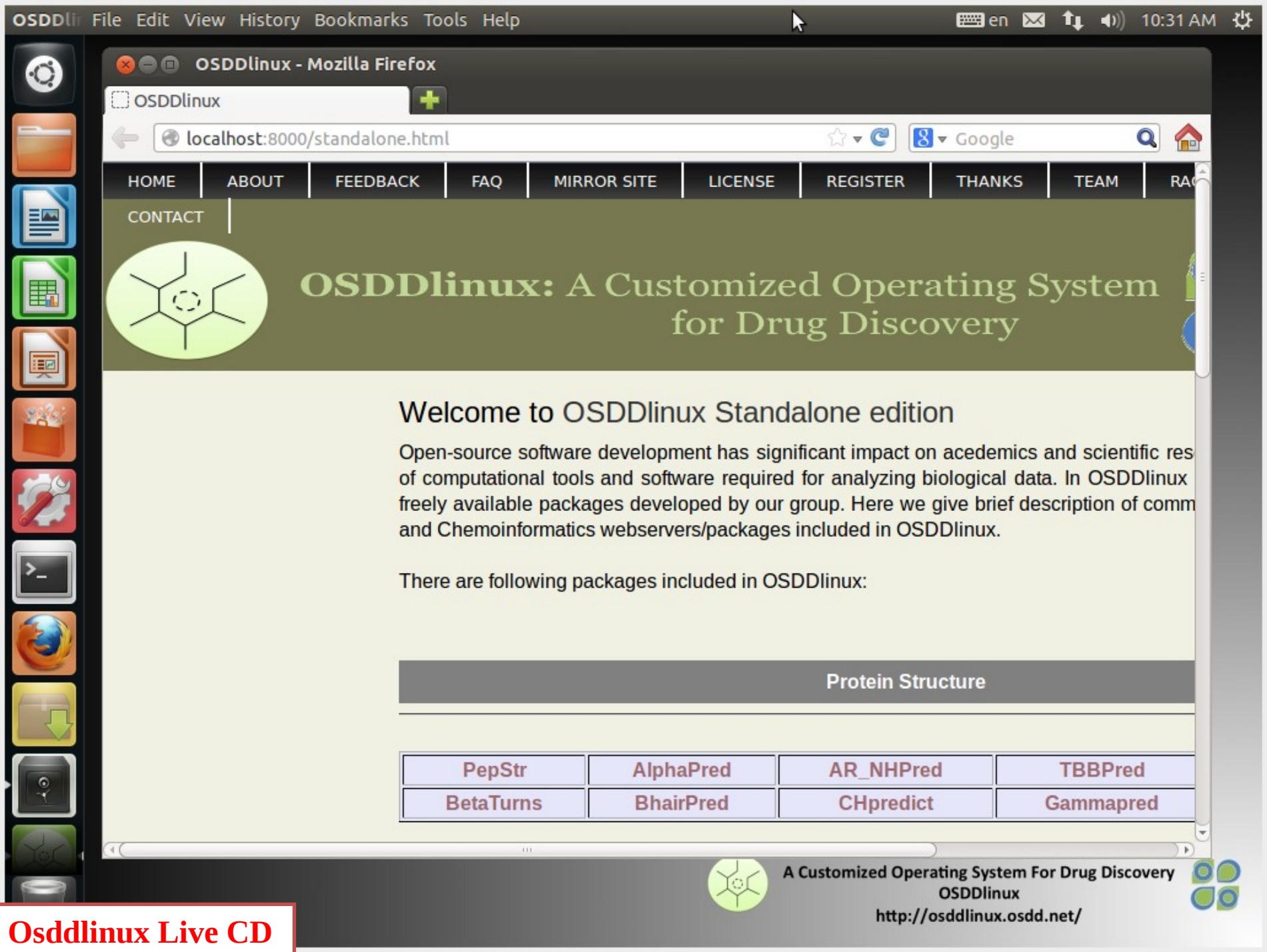
Platform for Developers
Developers with no infrastructure
Infrastructure on existing linux setup

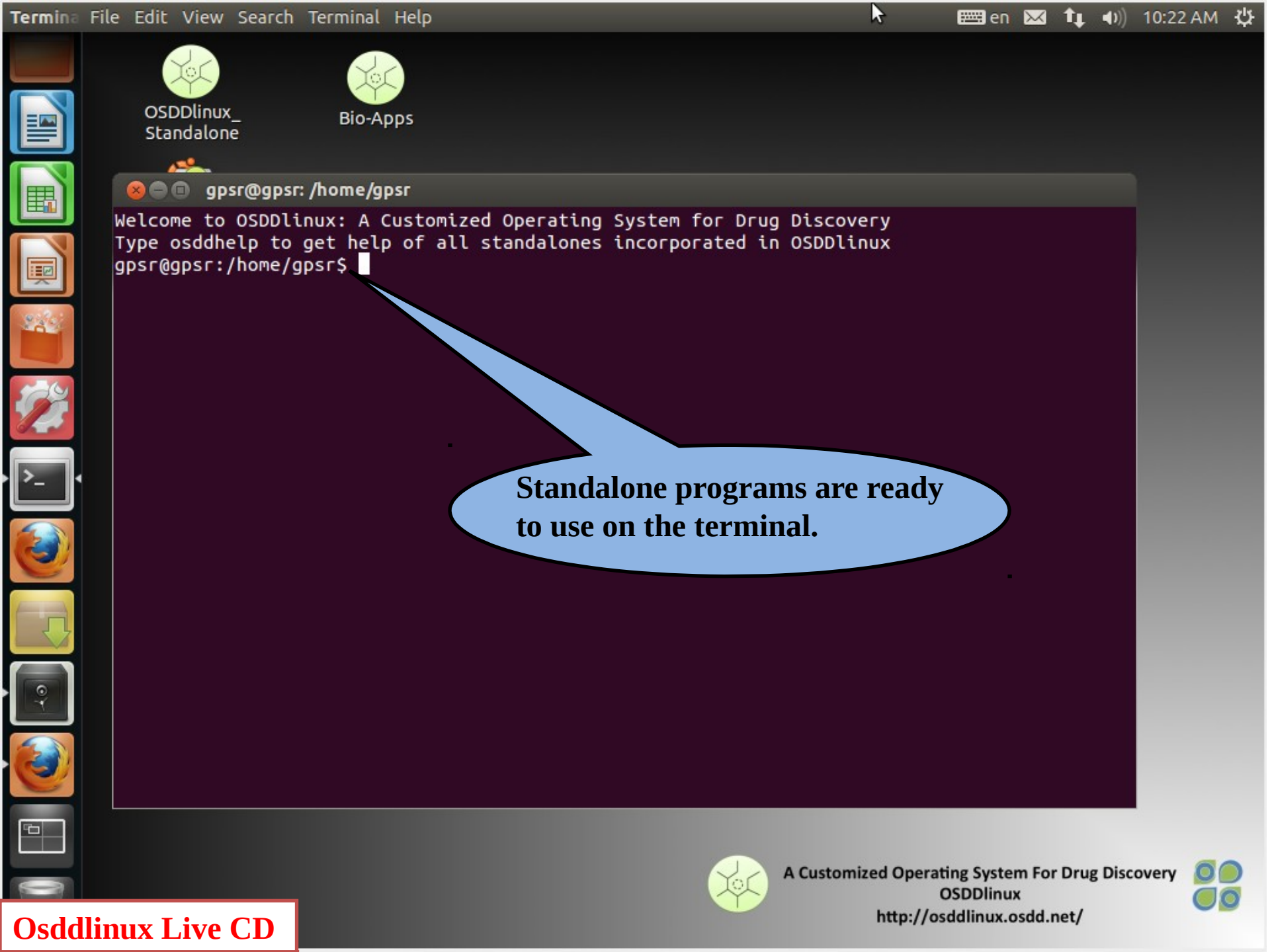
Linux for Students
Bootable LivedDVD for occasional learning
OSDDlinux on Windows/MAC Users
Online hand-on experience on Linux & PERL programming

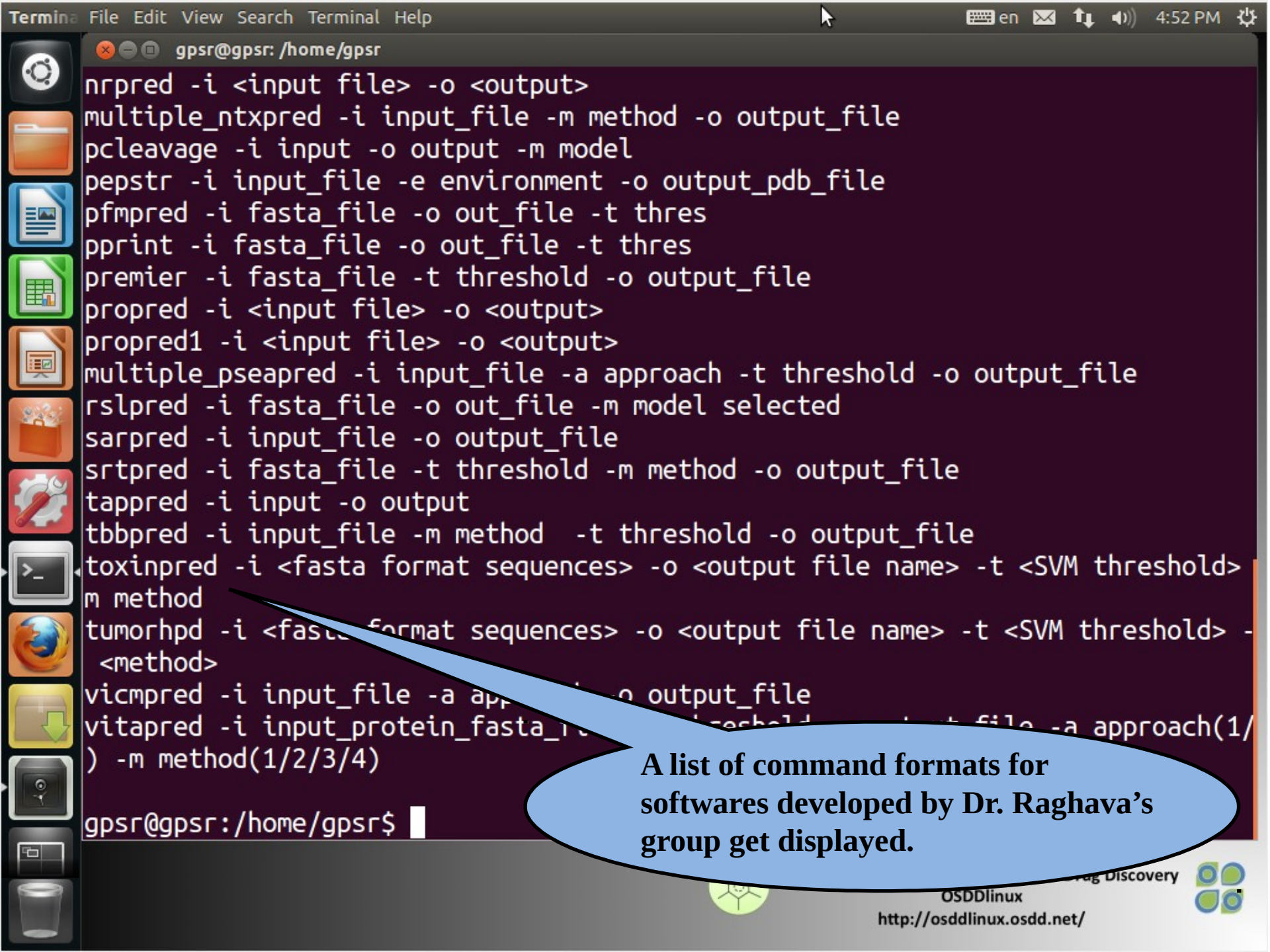
Installation of ODDlinux

1. **Live CD/DVD/USB no need to boot**
2. **Install on a new machine**
3. **Install via Virtual Box (Windows/MAC/Unix)**
4. **Install on existing Unix/MAC machine**


Ubuntu Desktop
10:36 AM
gpsr
OSDDlinux_ Standalone
OSDDlinux_ Webserver
OSDDlinux_Galaxy
OSDDlinux_ WebServers
Osddlinux desktop is ready for use.
Password for sudo : osddlinux root : osddlinux
A Customized Operating System For Drug Discovery
OSDDlinux
http://osddlinux.osdd.net/
Osddlinux installation on system hard drive


OSDDlinux - Mozilla Firefox
OSDDlinux
localhost:8000/standalone.html
HOME
ABOUT
FEEDBACK
FAQ
MIRROR SITE
LICENSE
REGISTER
THANKS
TEAM
CONTACT
OSDDlinux: A Customized Operating System for Drug Discovery
Welcome to OSDDlinux Standalone edition
Open-source software development has significant impact on acedemics and scientific res
of computational tools and software required for analyzing biological data. In OSDDlinux
freely available packages developed by our group. Here we give brief description of comm
and Chemoinformatics webservers/packages included in OSDDlinux.
There are following packages included in OSDDlinux:
Protein Structure
PepStr
AlphaPred
AR_NHPred
TBBPred
BetaTurns
BhairPred
CHpredict
Gammapred
A Customized Operating System For Drug Discovery
OSDDlinux
http://osddlinux.osdd.net/
Osddlinux Live CD


Termina File Edit View Search Terminal Help
en 10:22 AM
OSDDlinux_ Standalone
Bio-Apps
gpsr@gpsr: /home/gpsr
Welcome to OSDDlinux: A Customized Operating System for Drug Discovery
Type osddhelp to get help of all standalones incorporated in OSDDlinux
gpsr@gpsr:/home/gpsr$
Standalone programs are ready to use on the terminal.
A Customized Operating System For Drug Discovery
OSDDlinux
http://osddlinux.osdd.net/
Osddlinux Live CD


Termina File Edit View Search Terminal Help
en 4:52 PM
gpsr@gpsr: /home/gpsr
nrpred -i <input file> -o <output>
multiple_ntxpred -i input_file -m method -o output_file
pcleavage -i input -o output -m model
pepstr -i input_file -e environment -o output_pdb_file
pfmpred -i fasta_file -o out_file -t thres
pprint -i fasta_file -o out_file -t thres
premier -i fasta_file -t threshold -o output_file
propred -i <input file> -o <output>
propred1 -i <input file> -o <output>
multiple_pseapred -i input_file -a approach -t threshold -o output_file
rslpred -i fasta_file -o out_file -m model selected
sarpred -i input_file -o output_file
srtpred -i fasta_file -t threshold -m method -o output_file
tappred -i input -o output
tbbpred -i input_file -m method -t threshold -o output_file
toxinpred -i <fasta format sequences> -o <output file name> -t <SVM threshold>
m method
tumorhpd -i <fasta format sequences> -o <output file name> -t <SVM threshold> -
<method>
vicmpred -i input_file -a
o output_file
vitapred -i input_protein_fasta_f
-a approach(1/
) -m method(1/2/3/4)
gpsr@gpsr:/home/gpsr$
A list of command formats for softwares developed by Dr. Raghava's group get displayed.
OSDDlinux
http://osddlinux.osdd.net/

OSDDLINUX

Customized operating environment for drug discovery pipeline

BIOINFORMATICS

VACCINE INFORMATICS

CHEMINFORMATICS

Live Server

Live CD

Pkg Repository

Installation

Webserver

Standalone

Galaxy platform

All in ONE

Thanks