# Role of Informatics in Designing and Discovering Drugs/Vaccines


G P S Raghava, Head Bioinformatics Ce

Institute of Microbial Technology, Chandigarh, India

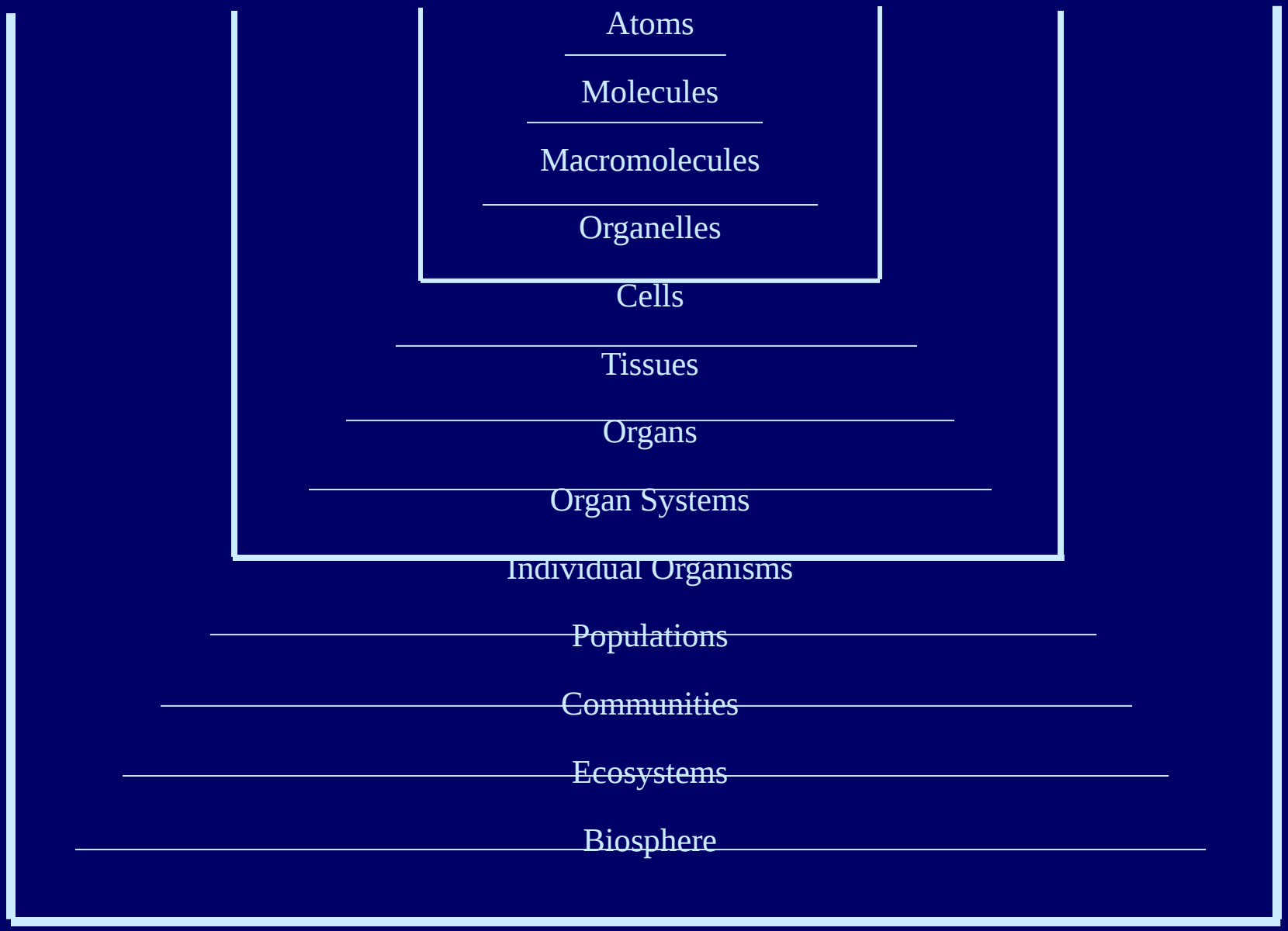

Human Genome  Drug Discovery  Vaccine Design


**Email: raghava@imtech.res.in**

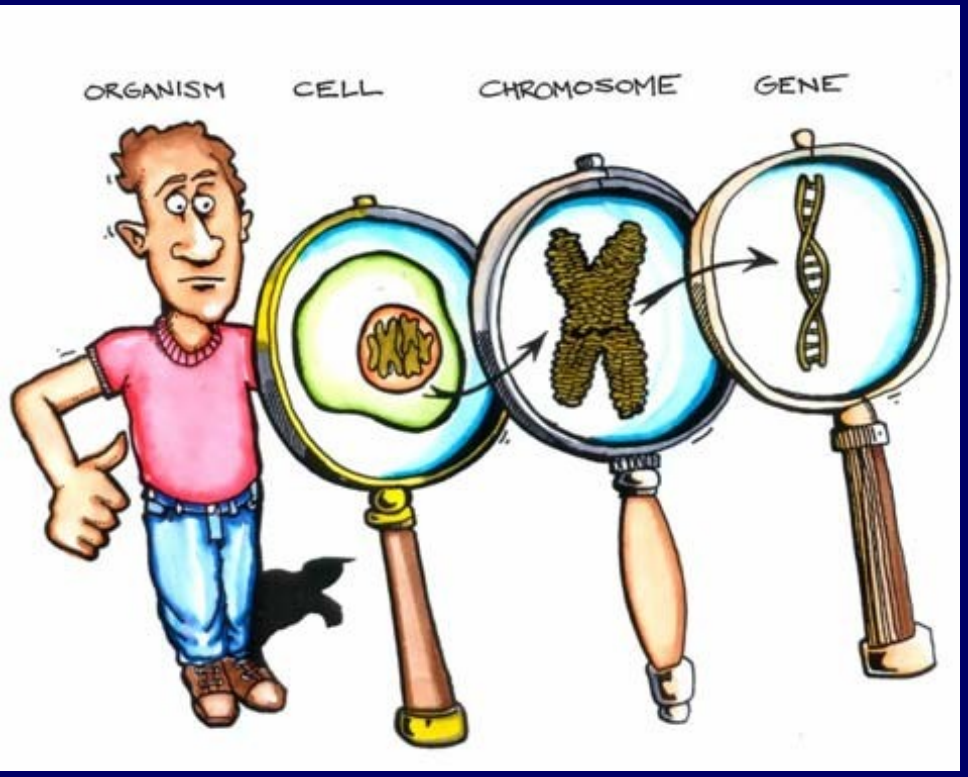

**http://crdd.osdd.net/**

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

# Hierarchy in Biology

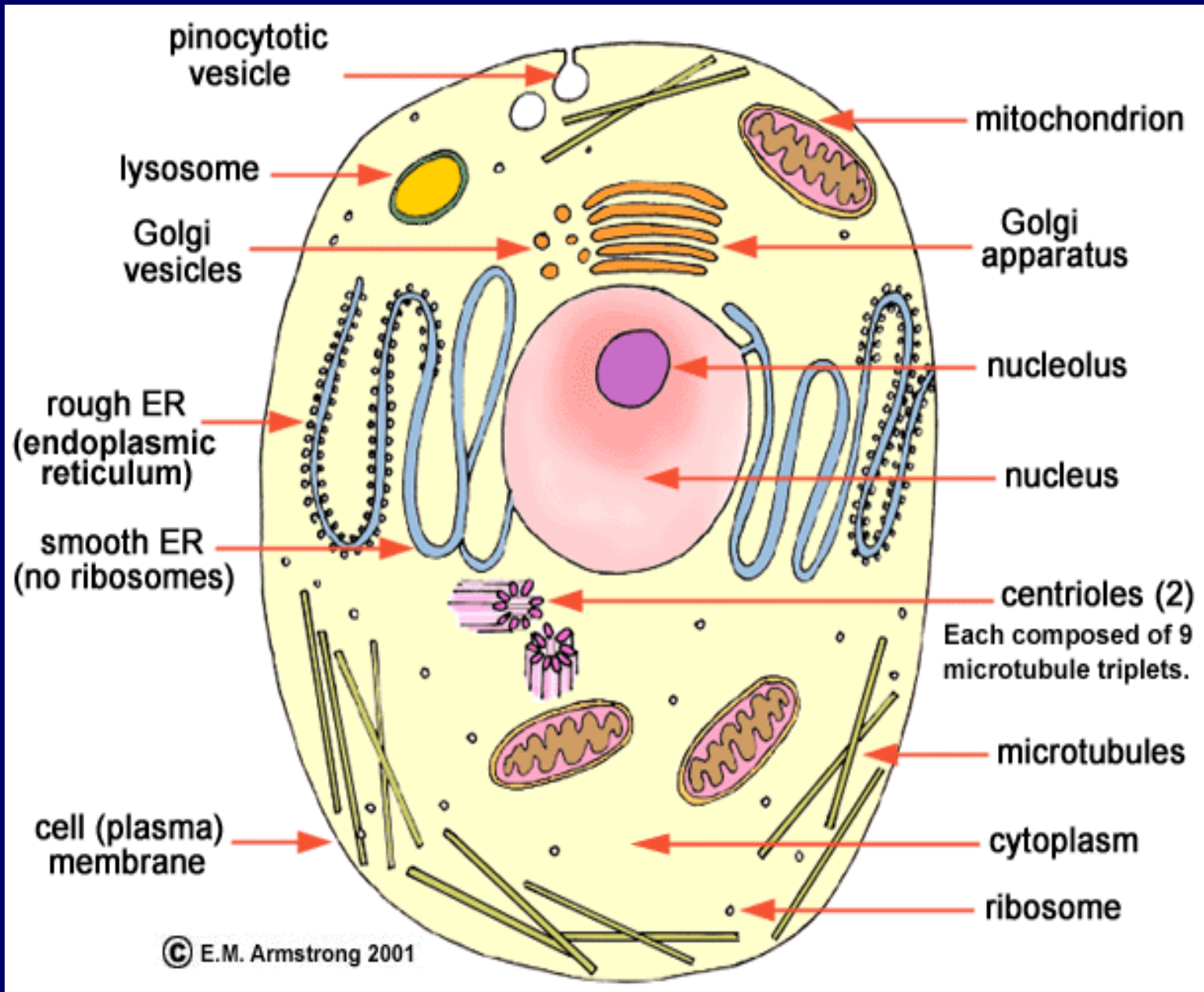


# Cells and DNA

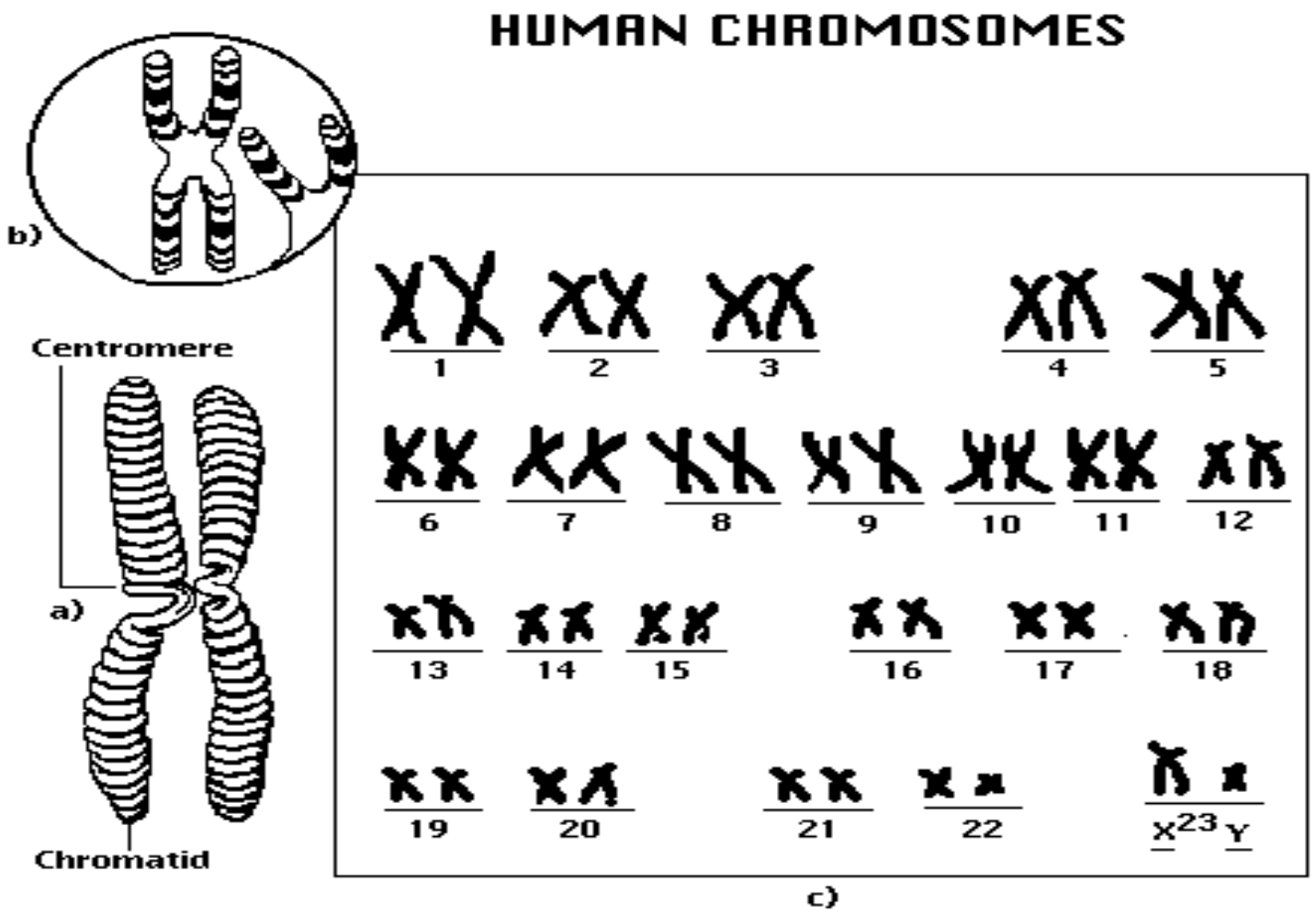



- Human body contains ~100 trillion cells
- Each cell contains 23 pairs of chromosomes (= **genome**)
- Chromosomes contain DNA
  - DNA is made of 4 nucleotide bases (Adenine, Guanine, Cytosine & Thymine) = AGCT sequence
- Every cell (except a few) in an individual contains the same exact genome

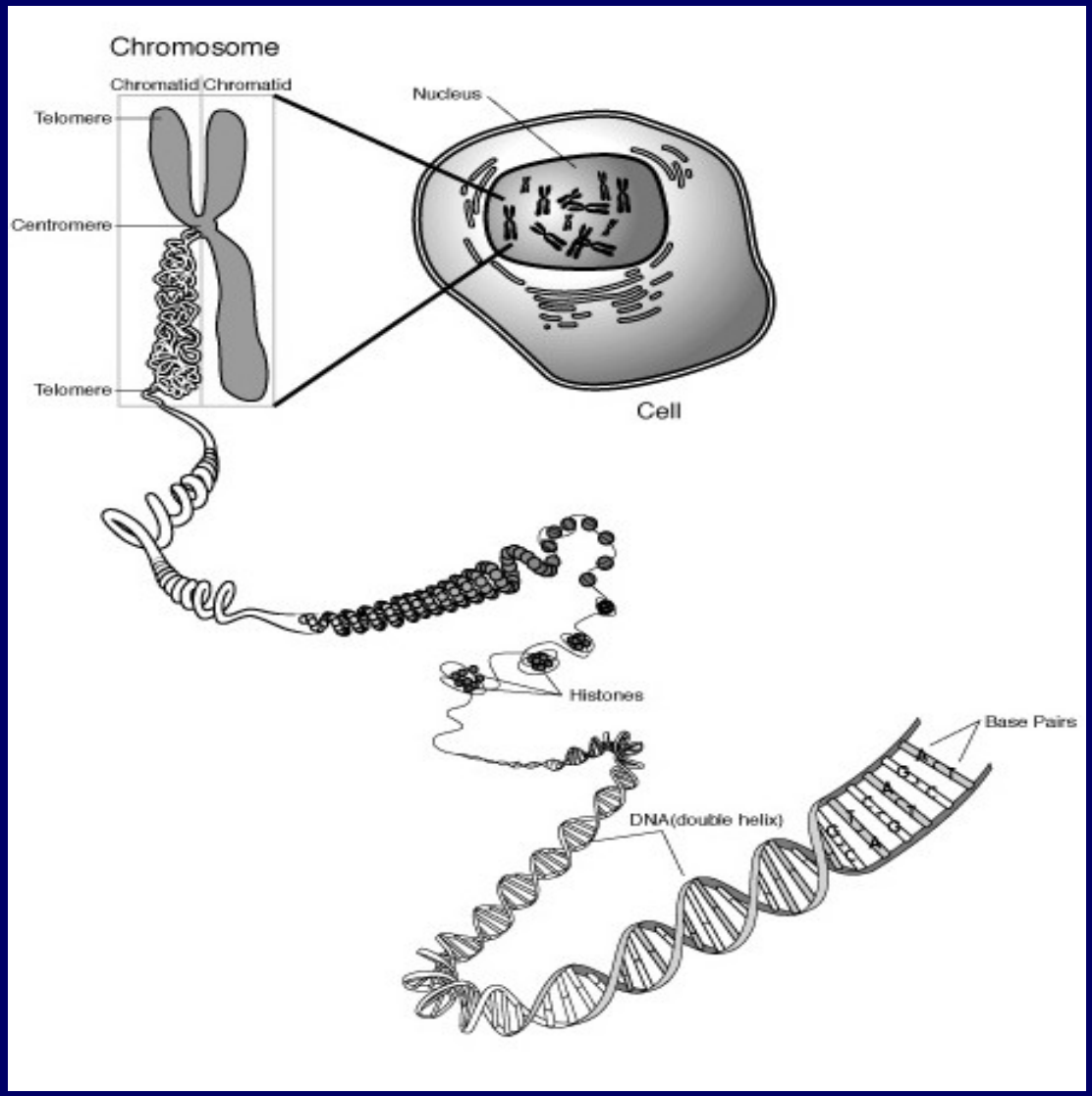
# Animal cell


pinocytotic vesicle
mitochondrion
lysosome
Golgi vesicles
Golgi apparatus
nucleolus
rough ER (endoplasmic reticulum)
nucleus
smooth ER (no ribosomes)
centrioles (2)
Each composed of 9 microtubule triplets.
microtubules
cell (plasma) membrane
cytoplasm
ribosome
© E.M. Armstrong 2001

# Human Chromosomes

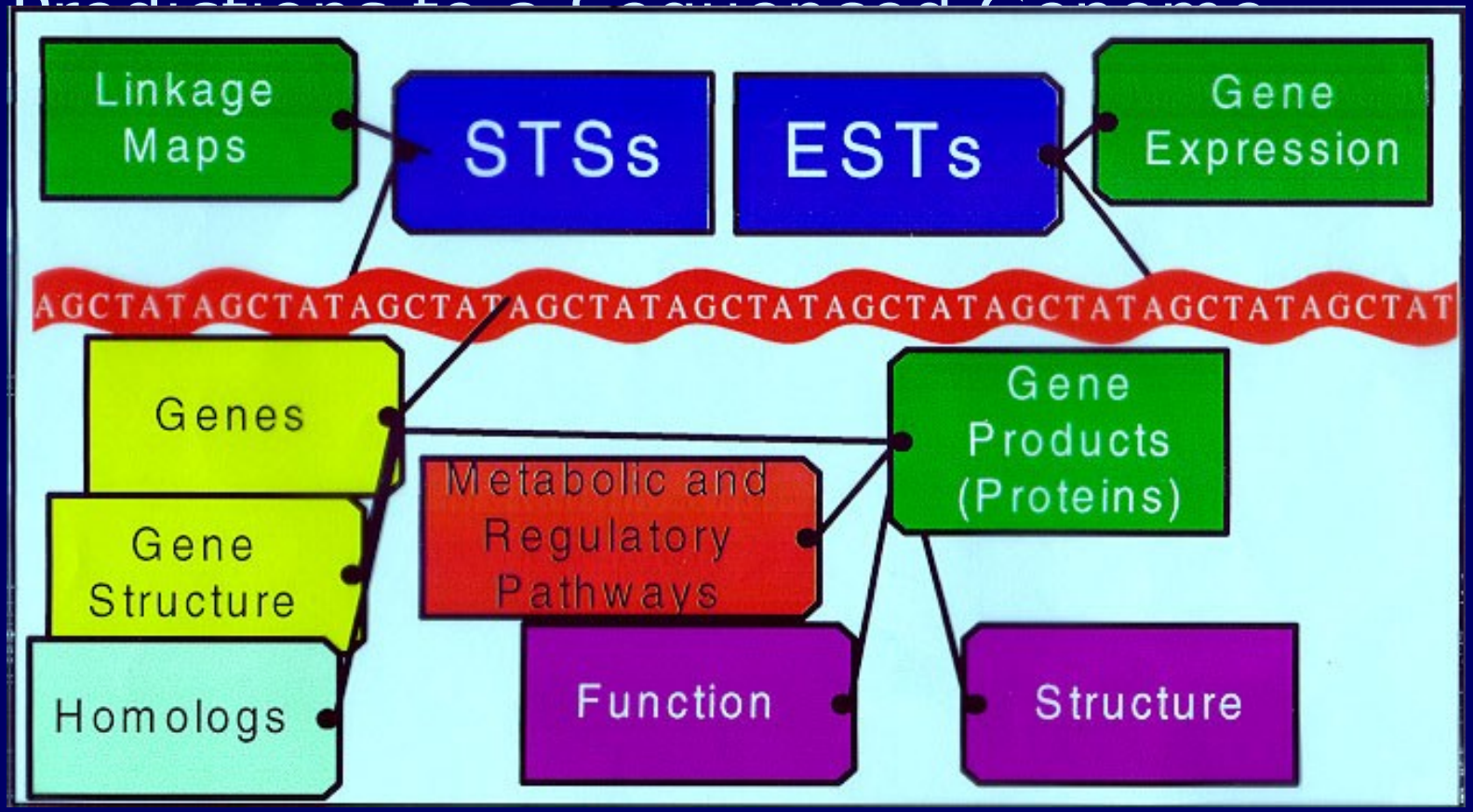


# Chromosomes and DNA



# Genome Annotation

The Process of Adding Biology Information and

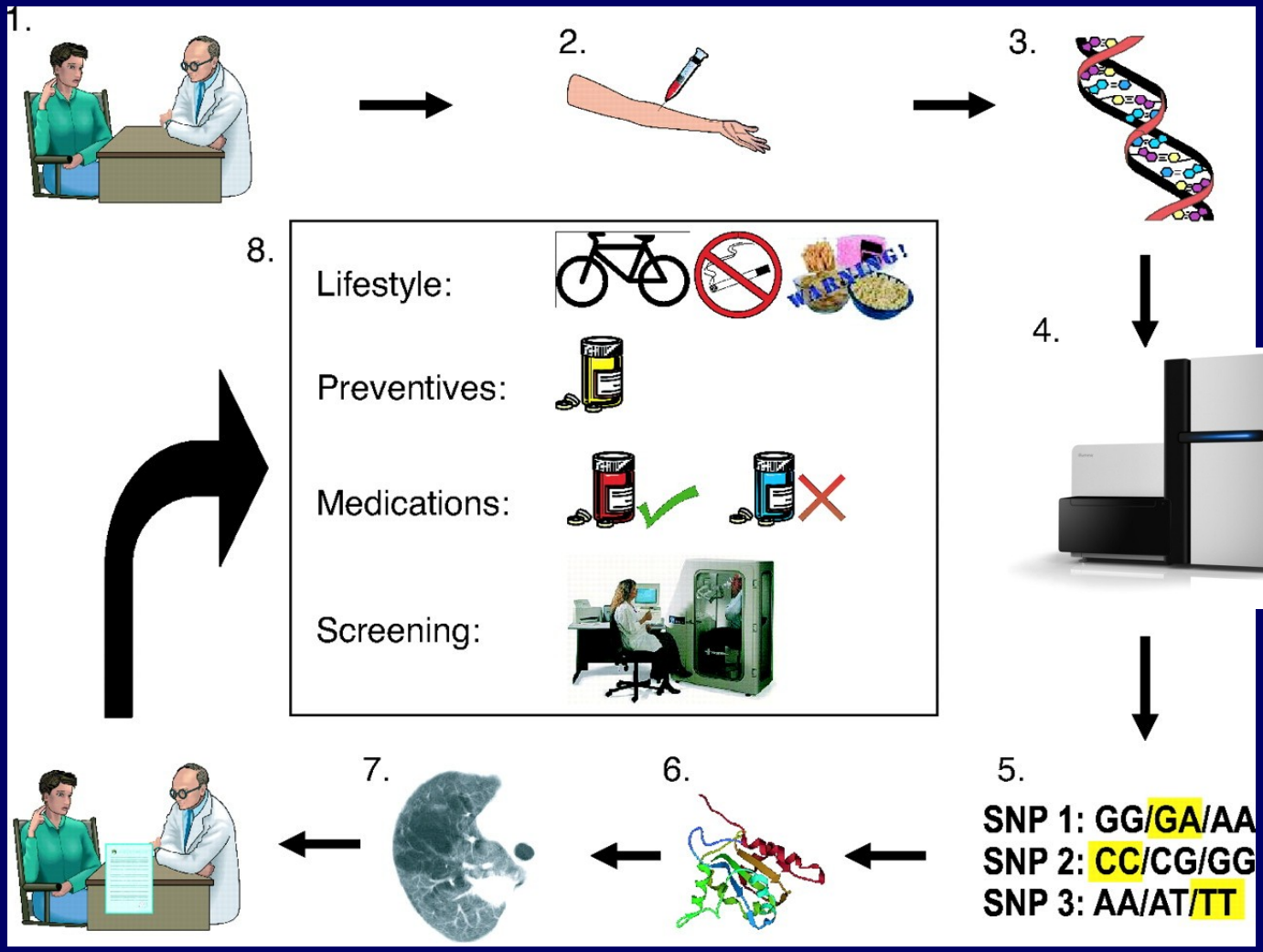


# Genome annotation tools at IMTECH

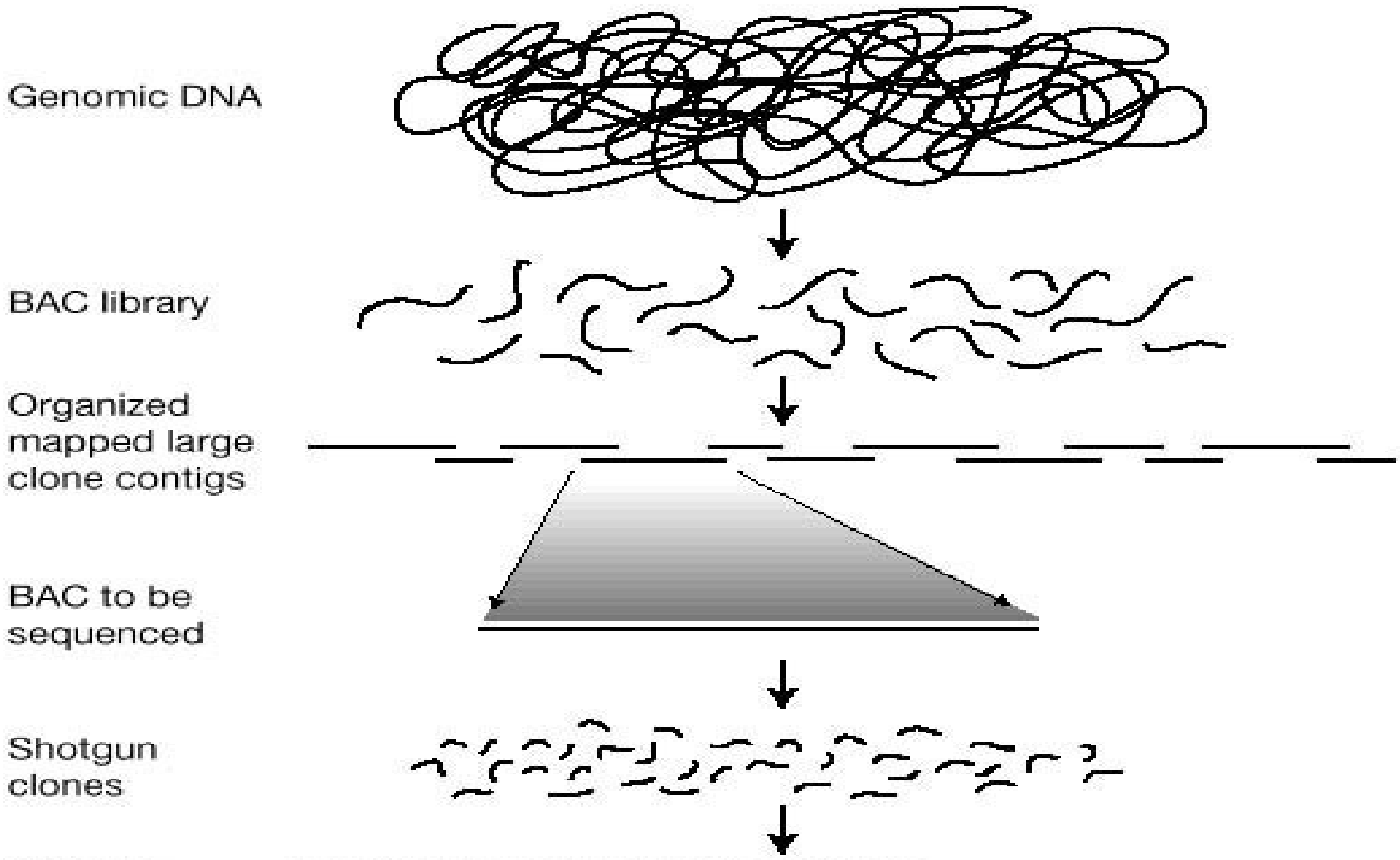
- Protein Structure prediction servers
- Servers for predicting function of proteins
- Servers for designing epitope based vaccine
- Genome annotation
- Molecular Interactions & Modifications
- Designing of Therapeutic Molecules
- Computer Aided Drug Design

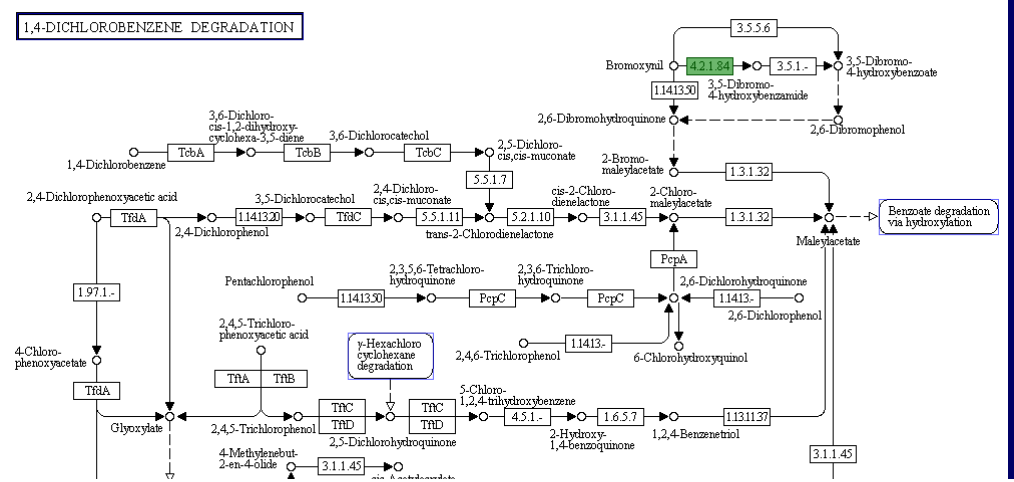
http://www.imtech.res.in/raghava/

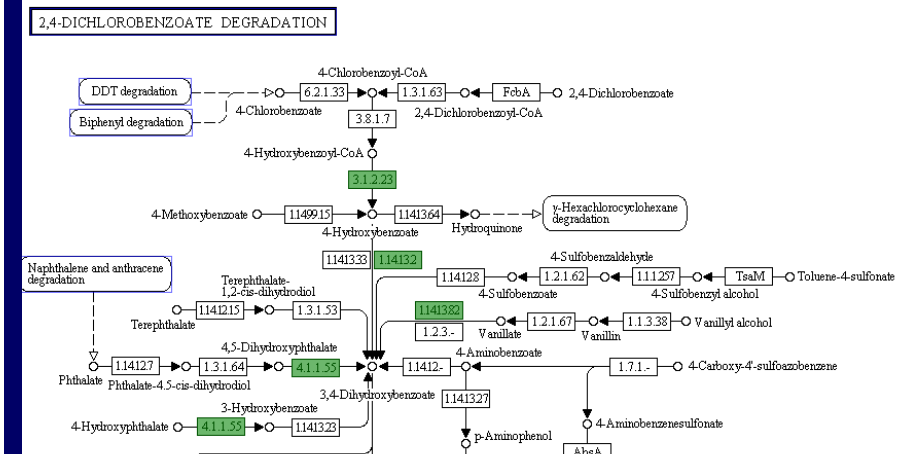
# The Future of Genomics in Medicine



# Hierarchical shotgun sequencing





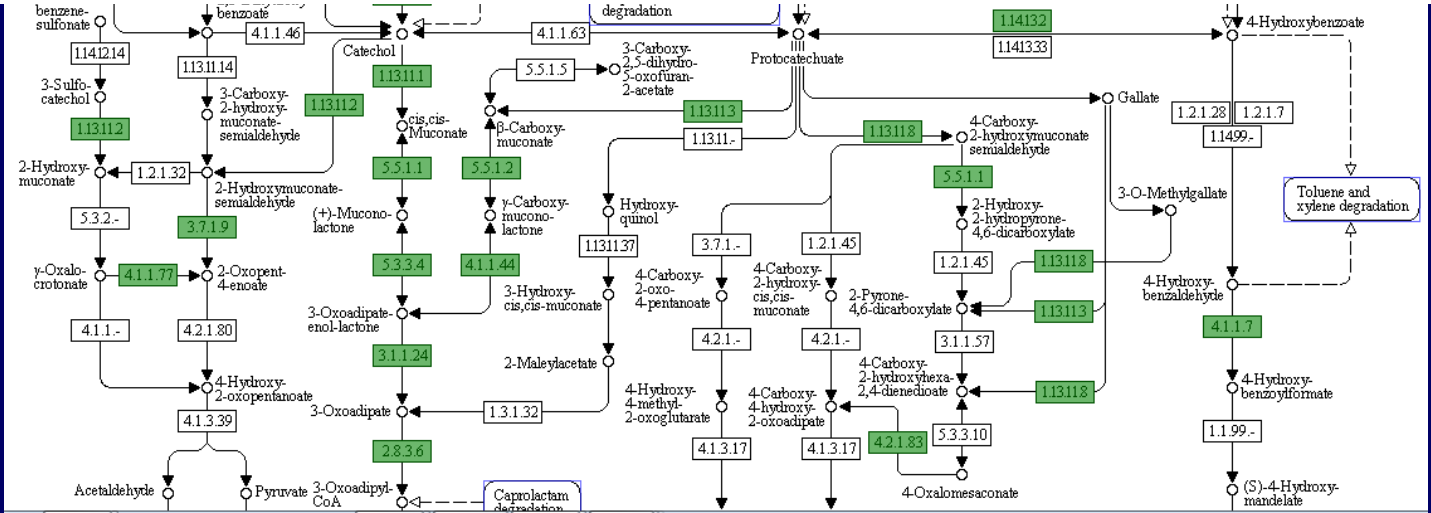




# Journal of Bacteriology

## Genome Sequence of the Nitroaromatic Compound-Degrading Bacterium Burkholderia sp. Strain SJ98

**Shailesh Kumar, Surendra Vikram and Gajendra Pal Singh Raghava**

*J. Bacteriol.* 2012, 194(12):3286. DOI: 10.1128/JB.00497-12.



# Genome assembly and annotation done at IMTECH

- ***Burkholderia sp.*** **SJ98 (Kumar *et al.* 2012).**
- ***Debaryomyces hansenii*** **MTCC 234 (Kumar *et al.* 2012).**
- ***Imtechella halotolerans*** **K1$^T$ (Kumar *et al.* 2012).**
- ***Marinilabilia salmonicolor*** **JCM 21150$^T$ (Kumar *et al.* 2012).**
- ***Rhodococcus imtechensis sp.*** **RKJ300 (Vikram *et al.* 2012).**
- ***Rhodosporidium toruloides*** **MTCC 457 (Kumar *et al.* 2012).**

# Concept of Drug and Vaccine

- Concept of Drug
  - Kill invaders of foreign pathogens
  - Inhibit the growth of pathogens
- Concept of Vaccine
  - Generate memory cells
  - Trained immune system to face various existing disease agents

# Drug Discovery and Design

## History of Drug/Vaccine development

- **Plants or Natural Product**
  - Plant and Natural products were source for medical substance
  - Example: foxglove used to treat congestive heart failure
  - Foxglove contain digitalis and cardiotonic glycoside
  - Identification of active component
- **Accidental Observations**
  - Penicillin is one good example
  - Alexander Fleming observed the effect of mold
  - Mold(Penicillium) produce substance penicillin
  - Discovery of penicillin lead to large scale screening
  - Soil micoorganism were grown and tested
  - Streptomycin, neomycin, gentamicin, tetracyclines etc.

# Drug Discovery and Design

- **Chemical Modification of Known Drugs**
  - Drug improvement by chemical modification
  - Pencillin G -> Methicillin; morphine->nalorphine
- **Receptor Based drug design**
  - Receptor is the target (usually a protein)
  - Drug molecule binds to cause biological effects
  - It is also called lock and key system
  - Structure determination of receptor is important
- **Ligand-based drug design**
  - Search a lead ocompound or active ligand
  - Structure of ligand guide the drug design process

# process

*Identify disease*

**GENOMICS, PROTEOMICS & BIOPHARM.**

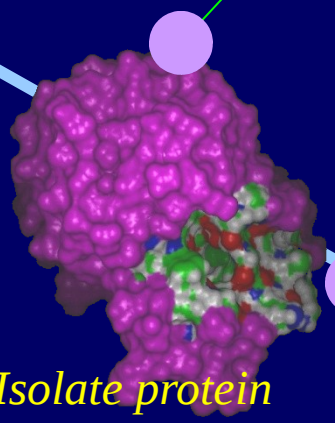
*Potentially producing many more targets and "personalized" targets*

**HIGH THROUGHPUT SCREENING**

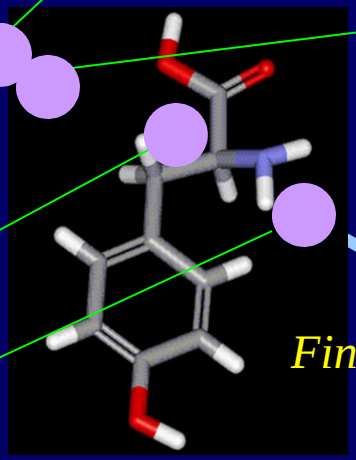
*Screening up to 100,000 compounds a day for activity against a target protein*

**VIRTUAL SCREENING**

*Using a computer to predict activity*



*Isolate protein*



*Find drug*

**COMBINATORIAL CHEMISTRY**

*Rapidly producing vast numbers of compounds*

**MOLECULAR MODELING**

*Computer graphics & models help improve activity*

*Preclinical testing*

**IN VITRO & IN SILICO ADME MODELS**

*Tissue and computer models begin to replace animal testing*

# Drug Design based on Bioinformatics Tools

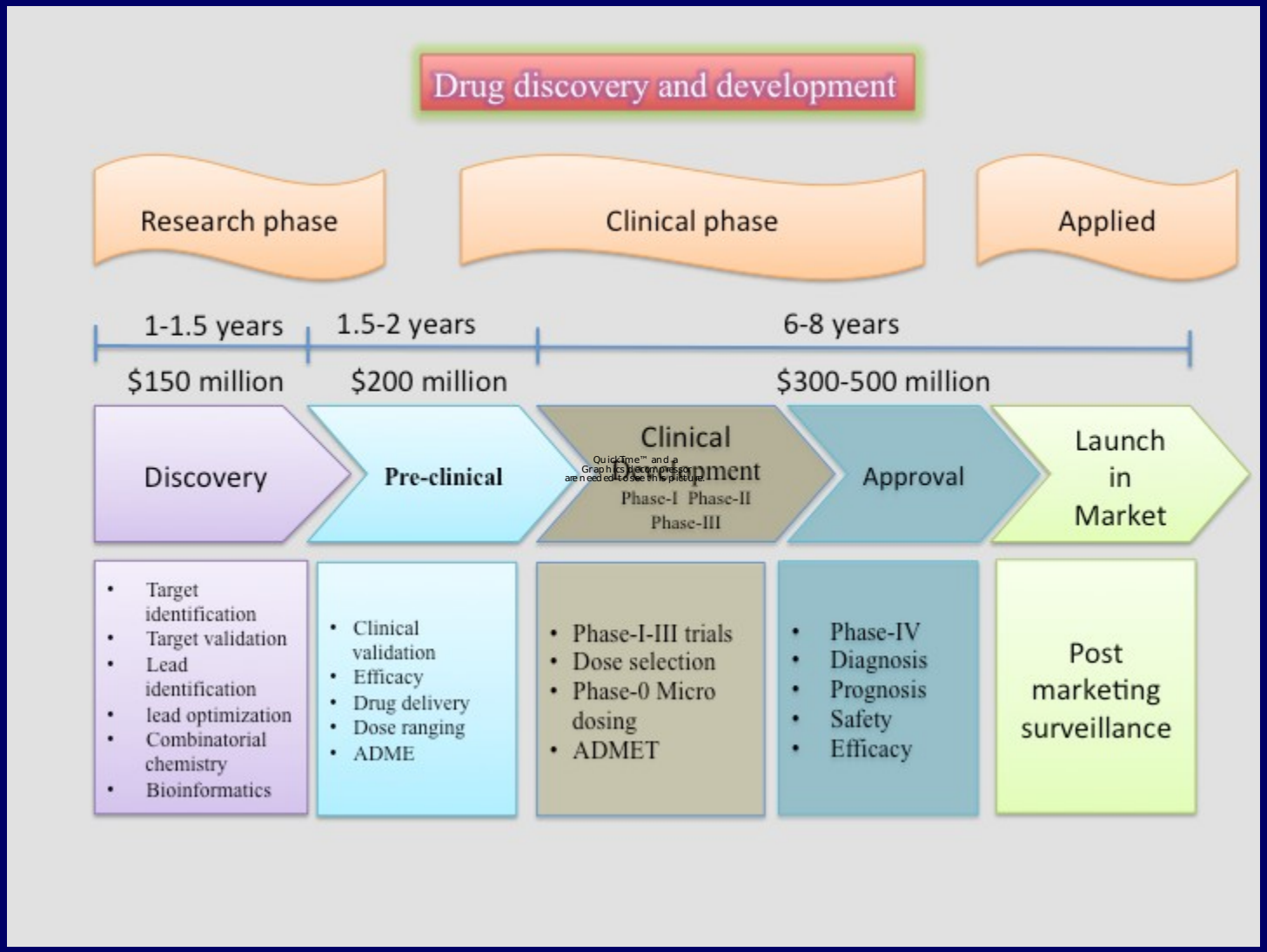
- **Detect the Molecular Bases for Disease**
  - Detection of drug binding site
  - Tailor drug to bind at that site
  - Protein modeling techniques
  - Traditional Method (brute force testing)

- **Rational drug design techniques**
  - Screen likely compounds built
  - Modeling large number of compounds (automated)
  - Application of Artificial intelligence
  - Limitation of known structures

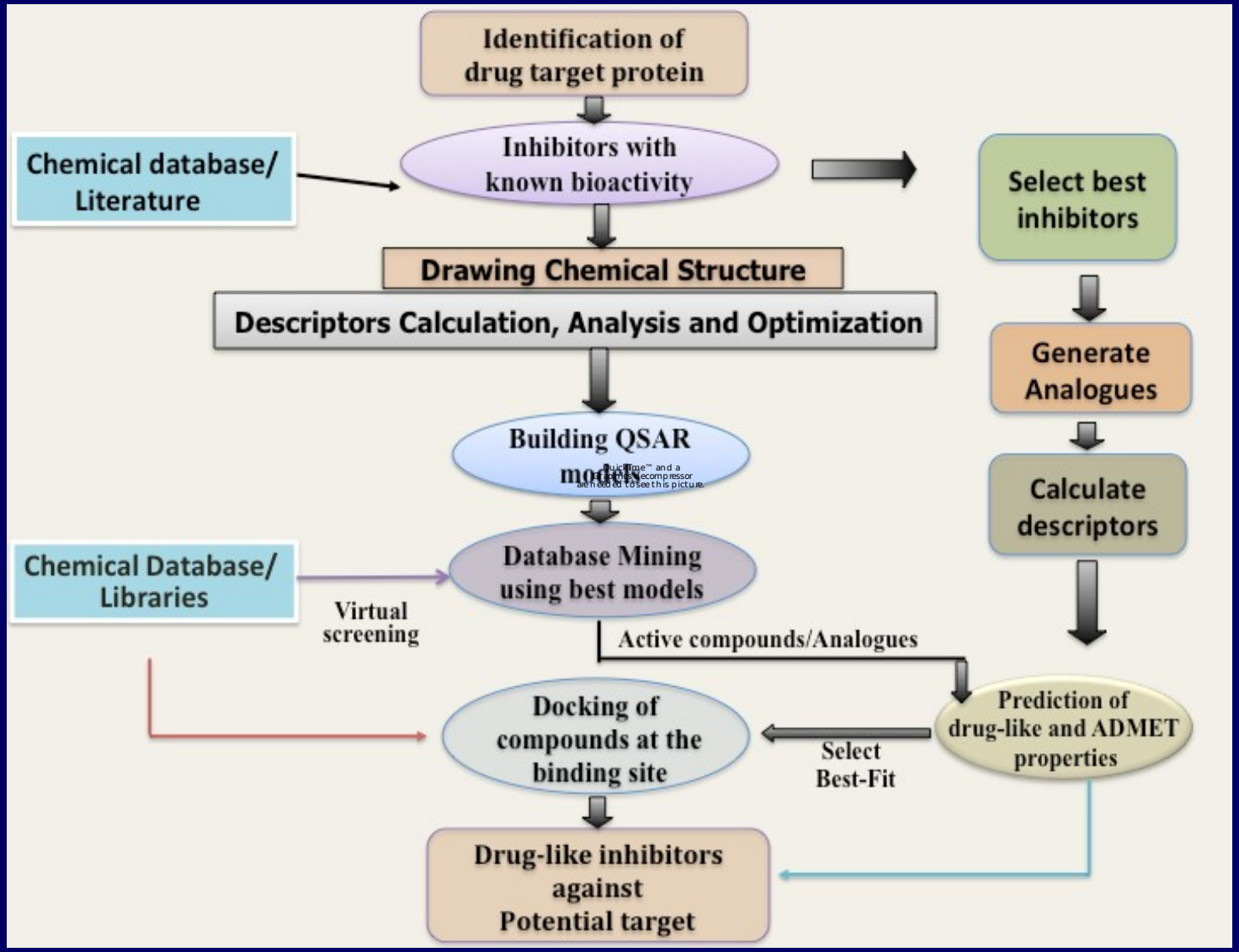
# Important Points in Drug Design based on Bioinformatics Tools

- Application of Genome
  - 3 billion bases pair
  - 30,000 unique genes
  - Any gene may be a potential drug target
  - ~500 unique target
  - Their may be 10 to 100 variants at each target gene
  - 1.4 million SNP
  - $10^{200}$ potential small molecules

# The amount of fund required depends on the success rate at the clinical trial stage



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



# 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)

# 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.**

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

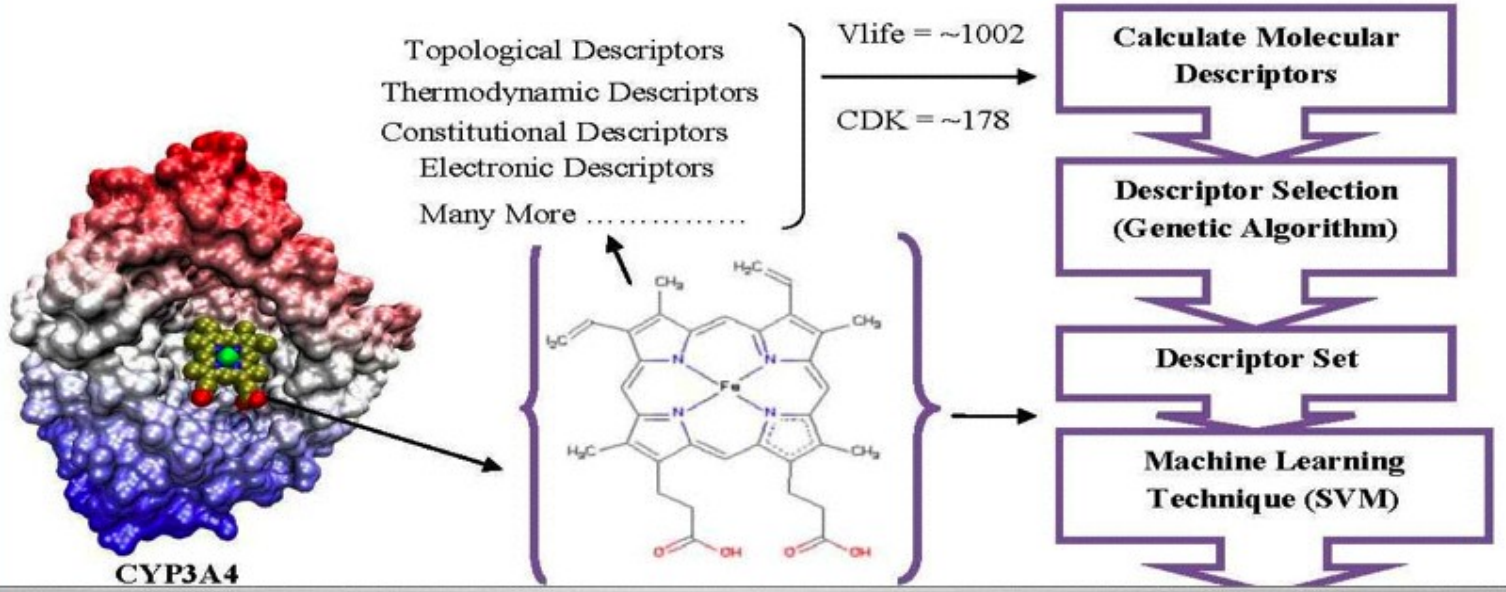


please cite:: **Prediction of cytochrome P450 isoform responsible for metabolizing a drug molecule** BMC Pharmacology

» 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.



# Home Page of HIVcoPRED

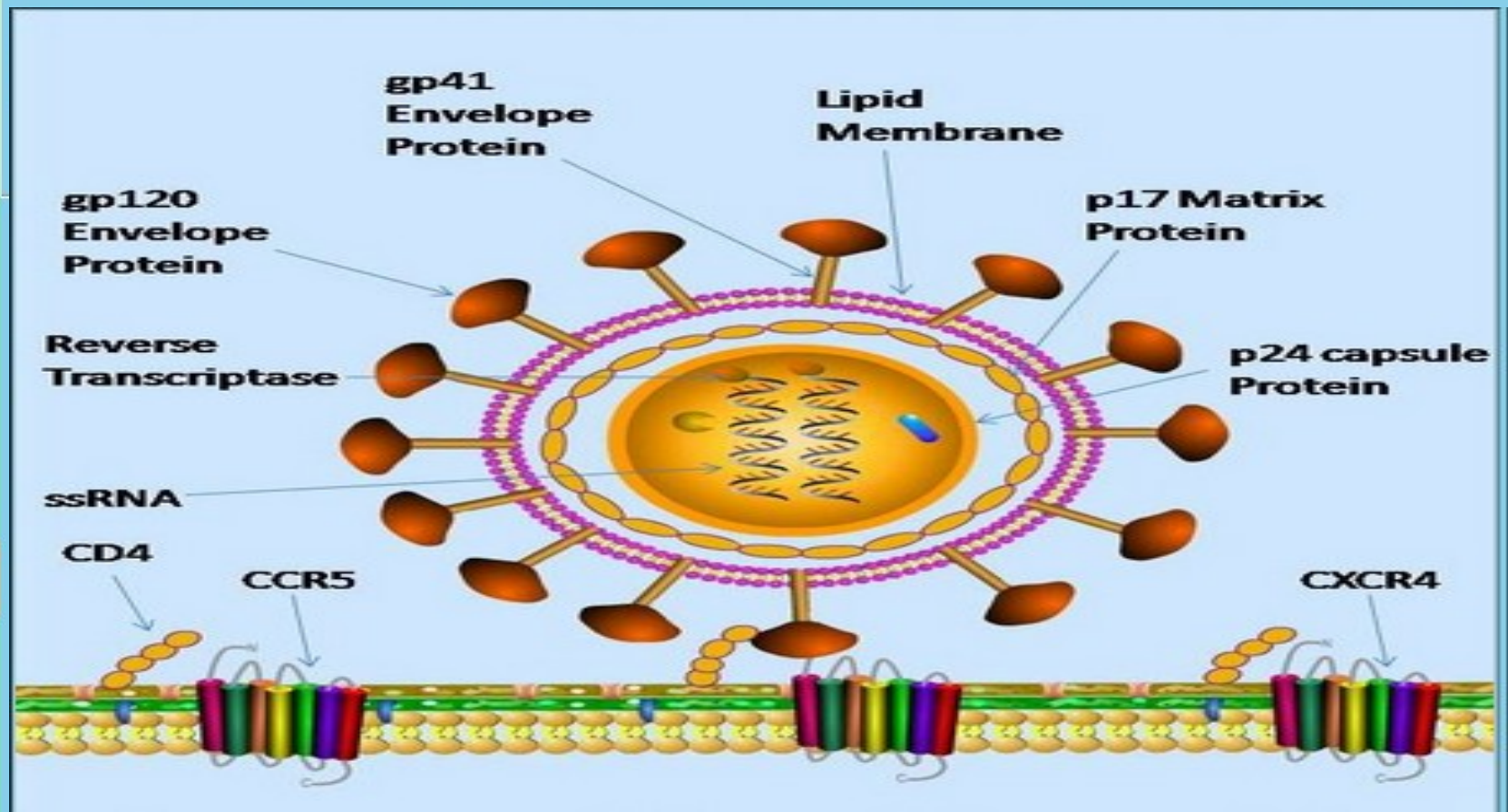
Server for prediction of HIV coreceptor usage

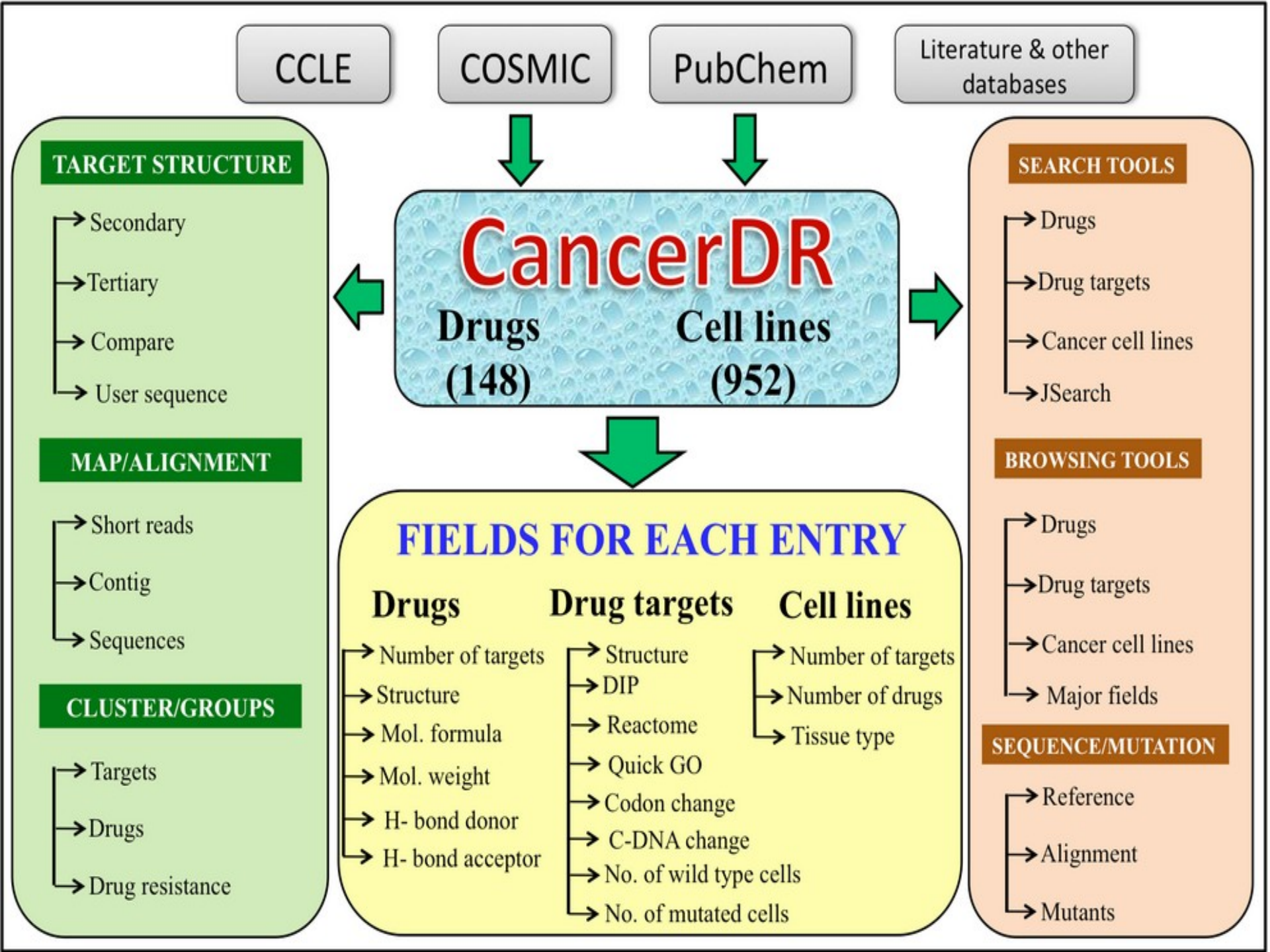


Home Page of HIVcoPRED

***** **Reference:** Kumar, R. and Raghava, GPS (2013) Hybrid Approach for Predicting Coreceptor Used by HIV-1 from Its V3 Loop Amino Acid Sequence. PLoS ONE 8(4): e61437 ******

Home
Submit
Help
Algorithm
Contact
Team
Supplement
Other Links
Raghava
HIVBio
CRDD

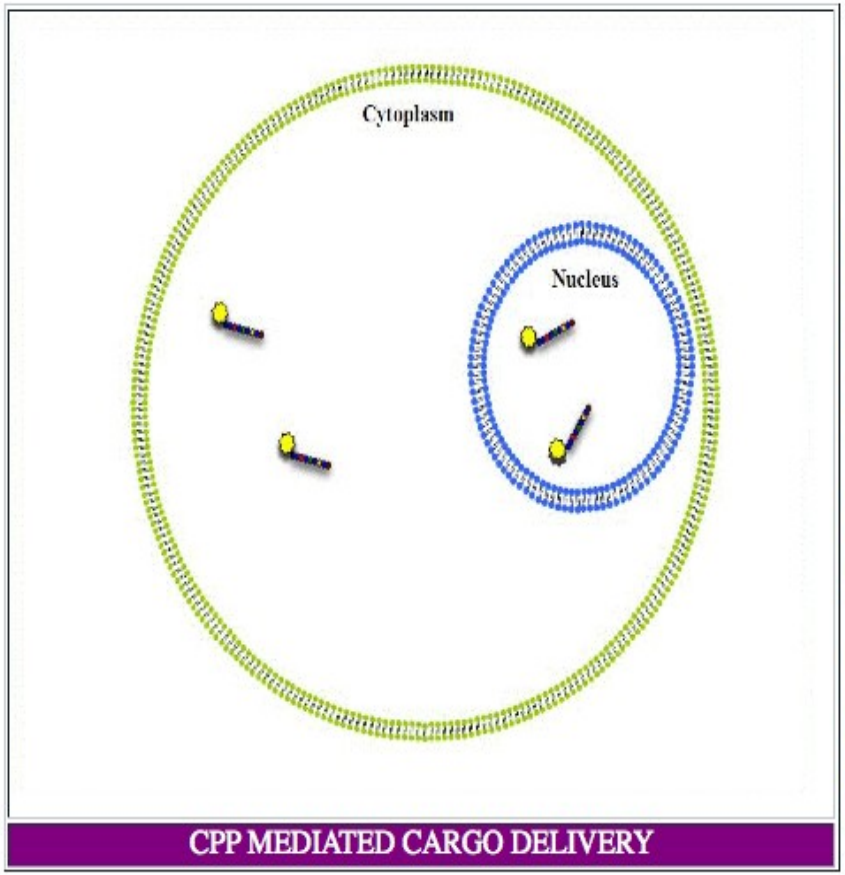



CCLE
COSMIC
PubChem
Literature & other databases
CancerDR
Drugs (148)
Cell lines (952)
TARGET STRUCTURE
Secondary
Tertiary
Compare
User sequence
MAP/ALIGNMENT
Short reads
Contig
Sequences
CLUSTER/GROUPS
Targets
Drugs
Drug resistance
FIELDS FOR EACH ENTRY
Drugs
Number of targets
Structure
Mol. formula
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

# 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

# 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

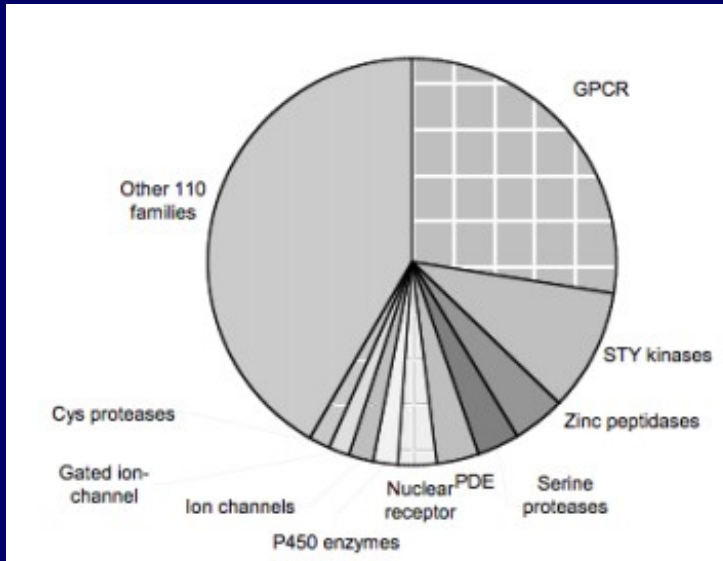
**PSLpred: localization of prokaryotic proteins**
**ESLpred: localization of Eukaryotic proteins**
**HSLpred: localization of Human proteins**
**MITpred: Prediction of Mitochndrial proteins**
**TBpred: Localization of mycobacterial proteins**

### Prediction of drugable proteins

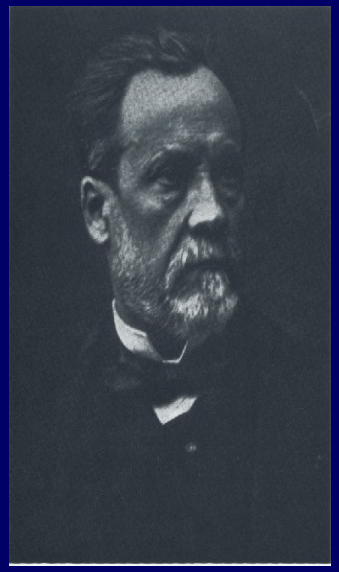
**Nrpred:** Classification of nuclear receptors
**GPCRpred:** Prediction of G-protein-coupled receptors
**GPCRsclass:** Amine type of GPCR
**VGIchan:** Voltage gated ion channel
**Pprint:** RNA interacting residues in proteins
**GSTpred:** Glutathione S-transferases proteins

### 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



# Major impact on public health and incidence of infectious diseases



- E. Jenner, the pioneer of vaccination in the Western world,
- Lady M. Montagu, an early advocate of smallpox inoculation
- L. Pasteur, who discovered attenutation

# Human Vaccines against pathogens

| Organism | Type | Vaccine Type | Year |
|---|---|---|---|
| Variola virus | Virus | Live | 1798 |
| Rabies virus | Virus | Inactivated | 1885 |
| *Salmonella typhi* | Bacteria | Live | 1896 |
| *Vibrio cholerae* | Bacteria | Inactivated | 1896 |
| *Yersinia pestis* | Bacteria | Inactivated | 1897 |
| *Corynebacterium diphtheriae* | Bacteria | Toxoid | 1923 |
| *Bordetella pertussis* | Bacteria | Acellular | 1926 |
| *Clostridium tetani* | Bacteria | Toxoid | 1927 |
| *Mycobacterium tuberculosis* | Bacteria | Live | 1927 |
| Yellow fever virus | Virus | Live | 1935 |
| Influenza virus type A | Virus | Inactivated | 1936 |
| Influenza virus type B | Virus | Inactivated | 1936 |
| *Coxiella burnetii* | Bacteria | Inactivated | 1938 |
| *Rickettsia prowazekii* | Bacteria | Inactivated | 1938 |
| *Rickettsia rickettsii* | Bacteria | Inactivated | 1938 |
| Central European encephalitis virus | Virus | Inactivated | 1939 |
| Poliovirus types 1, 2, and 3 | Virus | Inactivated/Live | 1962 |
| Measles virus | Virus | Live | 1963 |
| Mumps virus | Virus | Live | 1967 |
| Rubivirus | Virus | Live | 1969 |
| *Staphylococcus aureus* | Bacteria | Staphage lysate | 1976 |
| *Streptococcus pneumoniae* | Bacteria | Polysaccharide | 1977 |
| Human adenovirus types 4 and 7 | Virus | Live | 1980 |
| *Neisseria meningitidis* | Bacteria | Polysaccharide | 1981 |
| Hepatitis B | Virus | Recombinant | 1986 |
| *Haemophilus influenzae* | Bacteria | Conjugate | 1987 |
| Hantaan virus | Virus | Inactivated | 1989 |
| Japanese encephalitis virus | Virus | Inactivated | 1992 |
| Varicella-zoster virus | Virus | Live | 1994 |
| Hepatitis A | Virus | Inactivated | 1995 |
| *Escherichia coli* | Bacteria | Inactivated | 1995 |
| Junin virus | Virus | Live | 1996 |
| *Bacillus anthracis* | Bacteria | Adsorbed | 1998 |
| *Borrelia burgdorferi* | Bacteria | Recombinant | 1998 |

# History of Immunization

- **Children protected who recovered from smallpox**
- **Immunity induce, a process known as variolation**
- **Variolation spread to England and America**
- **Stopped due to the risk of death**
- **Edward Jenner found that protection against smallpox**
- **Inoculation with material from an individual infected with cowpox**
- **This process was called vaccination (cowpox is vaccina)**
- **Inoculum was termed a vaccine**
- **Protective antibodies was developed**

# Traditional Vaccine to Epitope Based Vaccines in Genomics Era

- **Concept of vaccine and Drug**
  - **Drug:** Kill invaders/pathogens and/or Inhibit the growth of pathogens
  - **Vaccine:** Trained immune system to face various existing disease agents

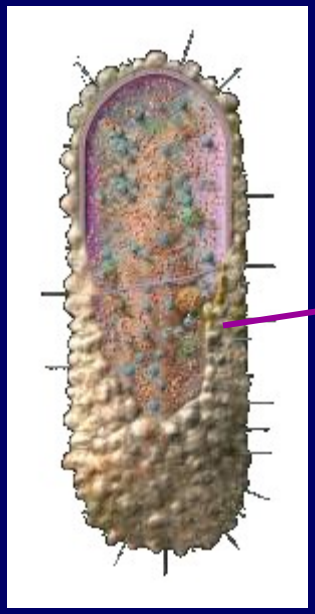
- **Type of Vaccines**
  - Whole Organism of Pathogen (MTb, 4000 proteins)
  - Target proteins/antigens which can activate immune system
  - Subunit Vaccine, Epitope Based Vaccines ( T & B cell epitopes)

- **Limitations of present methods of subunit vaccine design**
  - Developed for one or two MHC alleles (not suitable for large population)
  - Do not consider pathways of antigen processing
  - Difficult to detect experimentally known epitopes

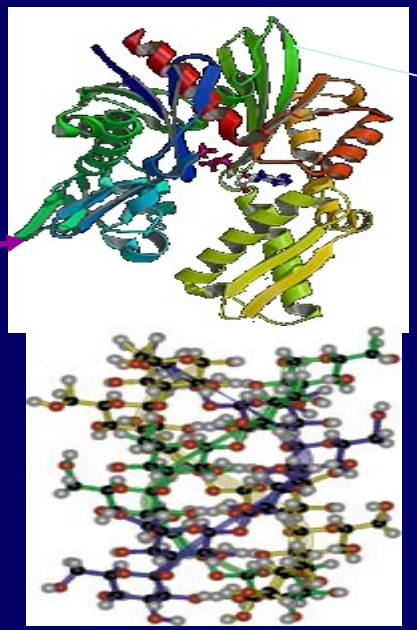
- **Initiatives taken by Bioinformatics Centre at IMT, Chandigarh India**

# Different vaccine design strategies

WHOLE ORGANISM → Purified Antigen → Epitopes (Subunit Vaccine)

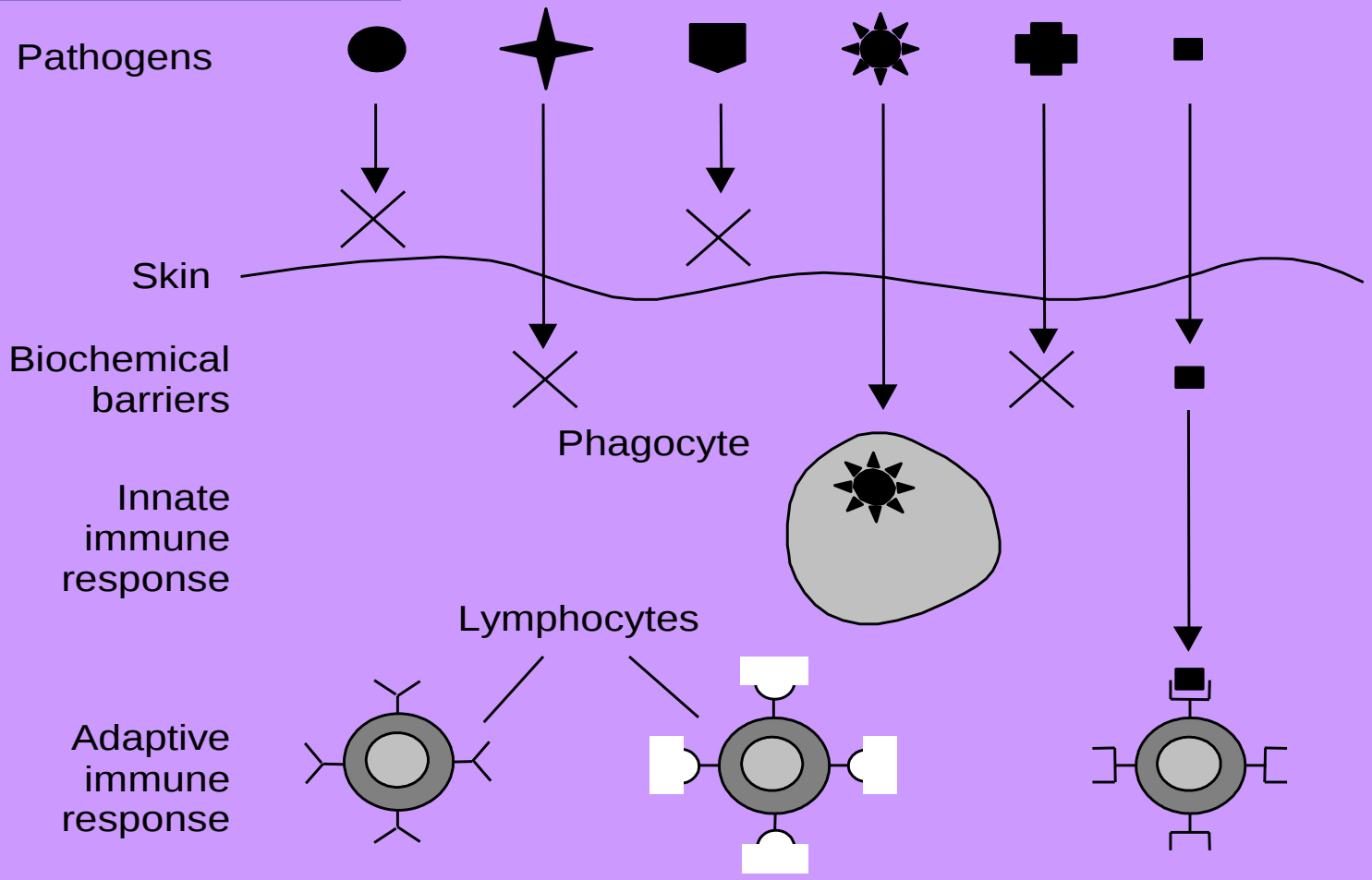


attenuated



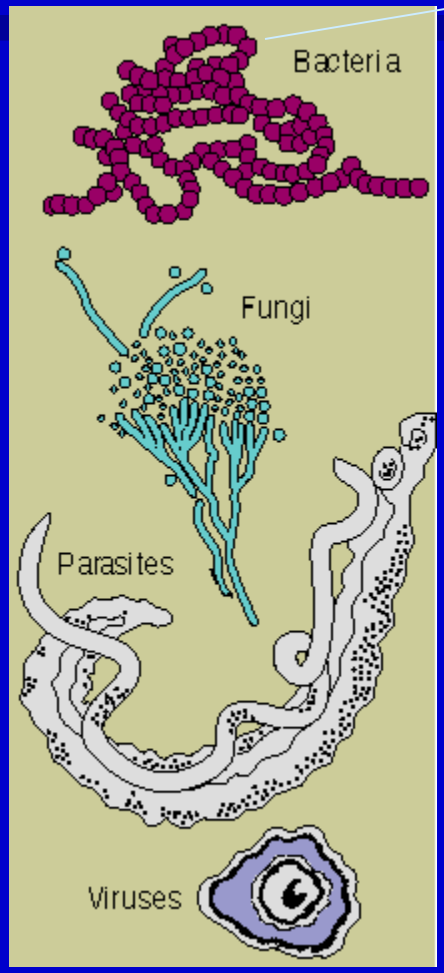

T cell epitope

# Multiple layers of the immune system



# Immune Defense and Long Term Protection

**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

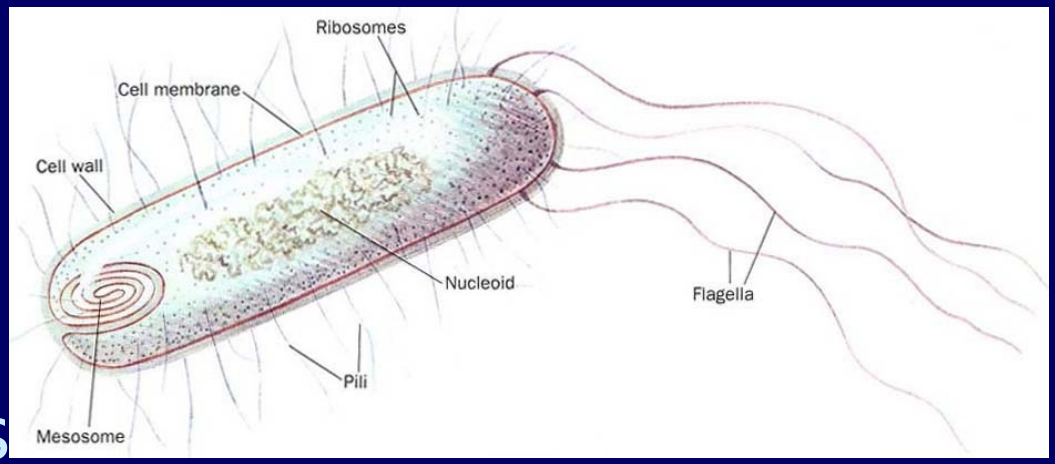
**Endogenous Antigen Processing**

**Exogenous Antigen Processing**

**CTL Epitopes**
MHC I binders
TAP Binders
P. Cleavage Sites

**T-Helper Epitopes**
MHC II binders

# Challenge in Vaccine Design




- $10^{12}$ types of B-cells
- Lack of effective vaccine against HIV, Cancer, M.Tb., Malaria
- Universal vaccines (MHC alleles)
- Bacterial genomes
  - Around 10 million bases
  - 4000 genes, 4000 proteins/antigens & cell
  - Selection of antigen/vaccine target
  - Identification of antigenic regions (epitopes)

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

# OSDD LINUX

A Custo
fo

Operating System for Drug Discovery

- 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

# Thanks