Introduction to Bioinformatics

Presented By

Dr G. P. S. Raghava

Co-ordinator, Bioinformatic Centre, IMTECH, Chandigarh, India

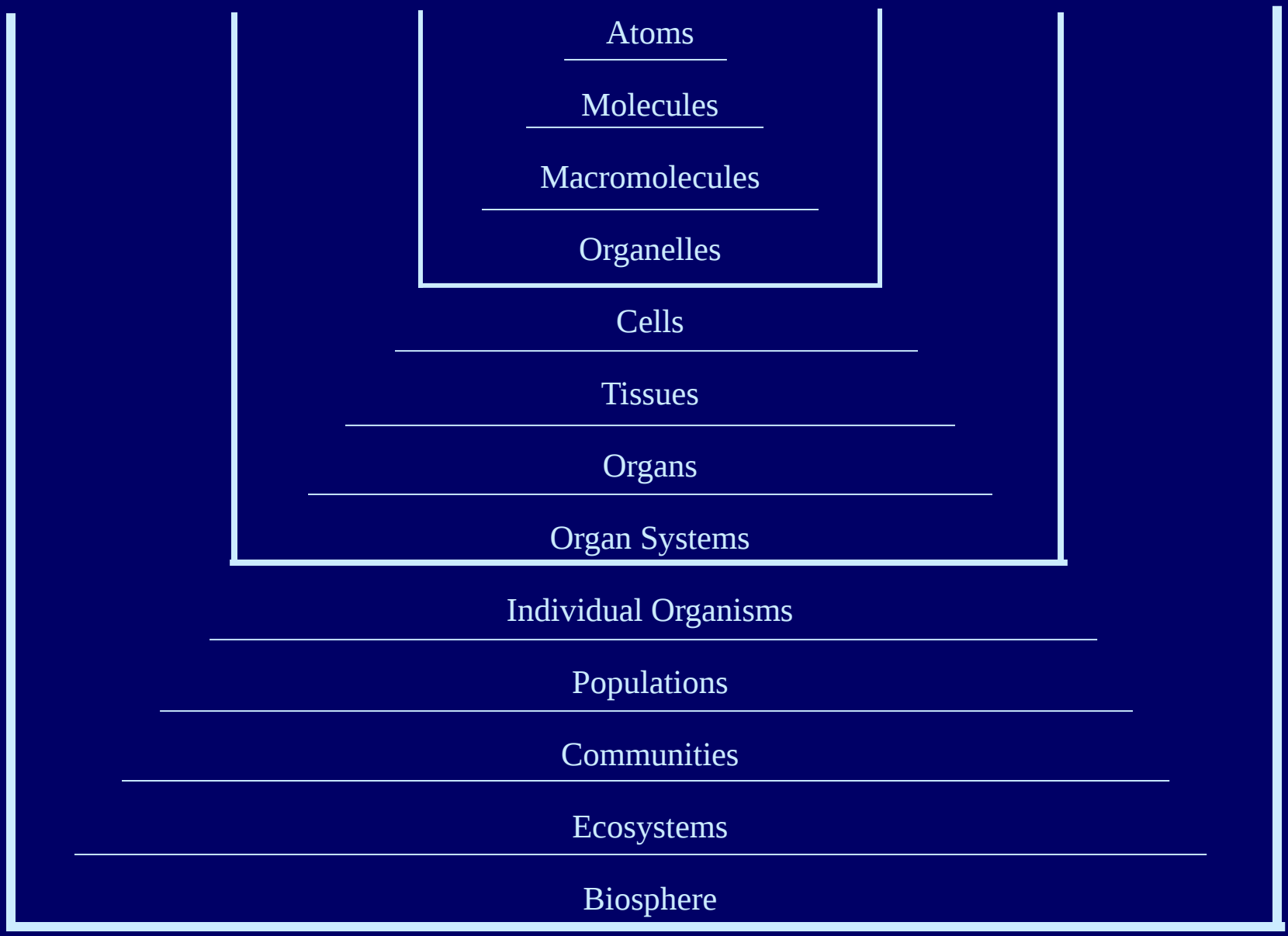
&

Visiting Professor, Pohang Univ. of Science & Technology, Republic of Korea

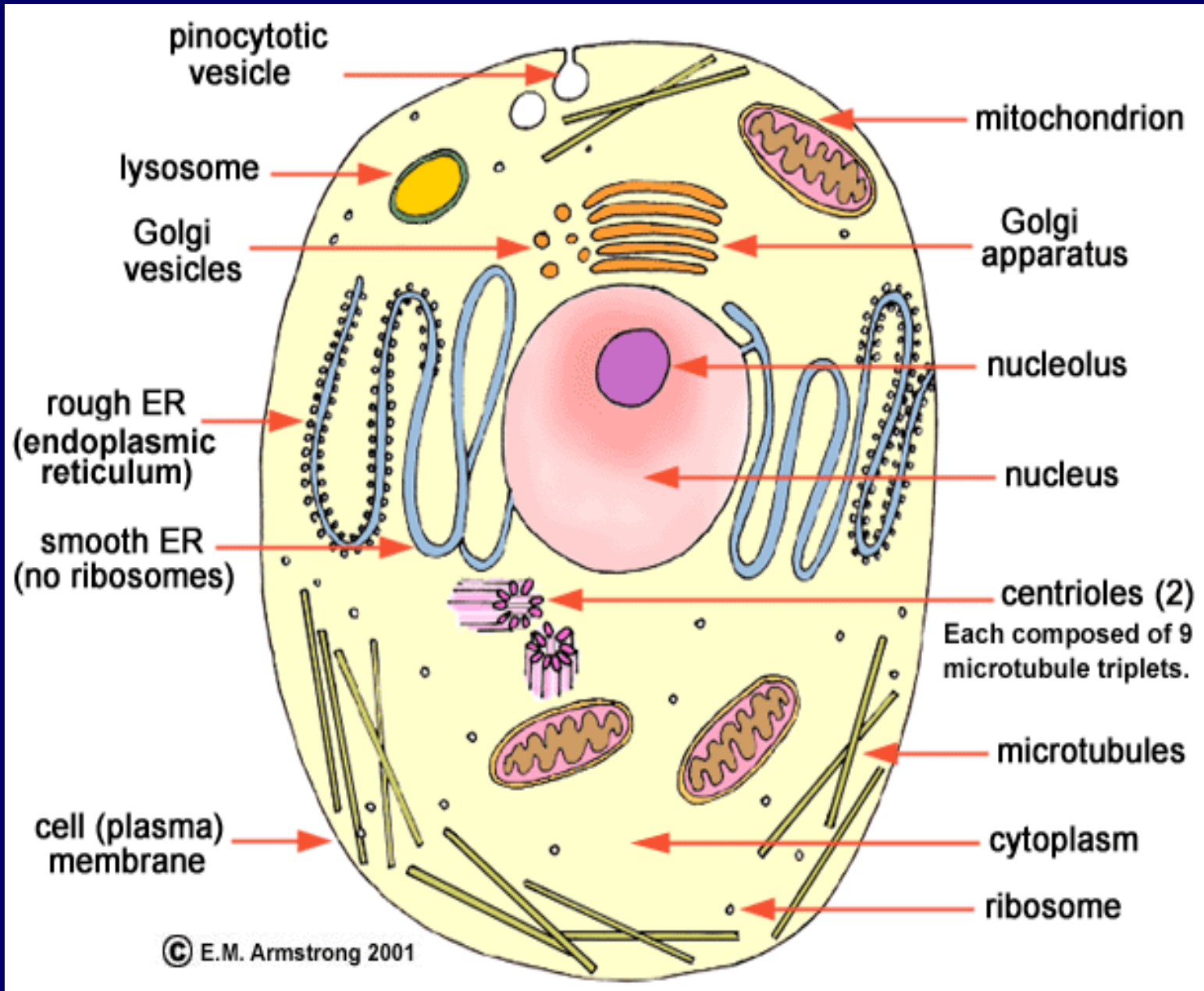
Email: raghava@imtech.res.in

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

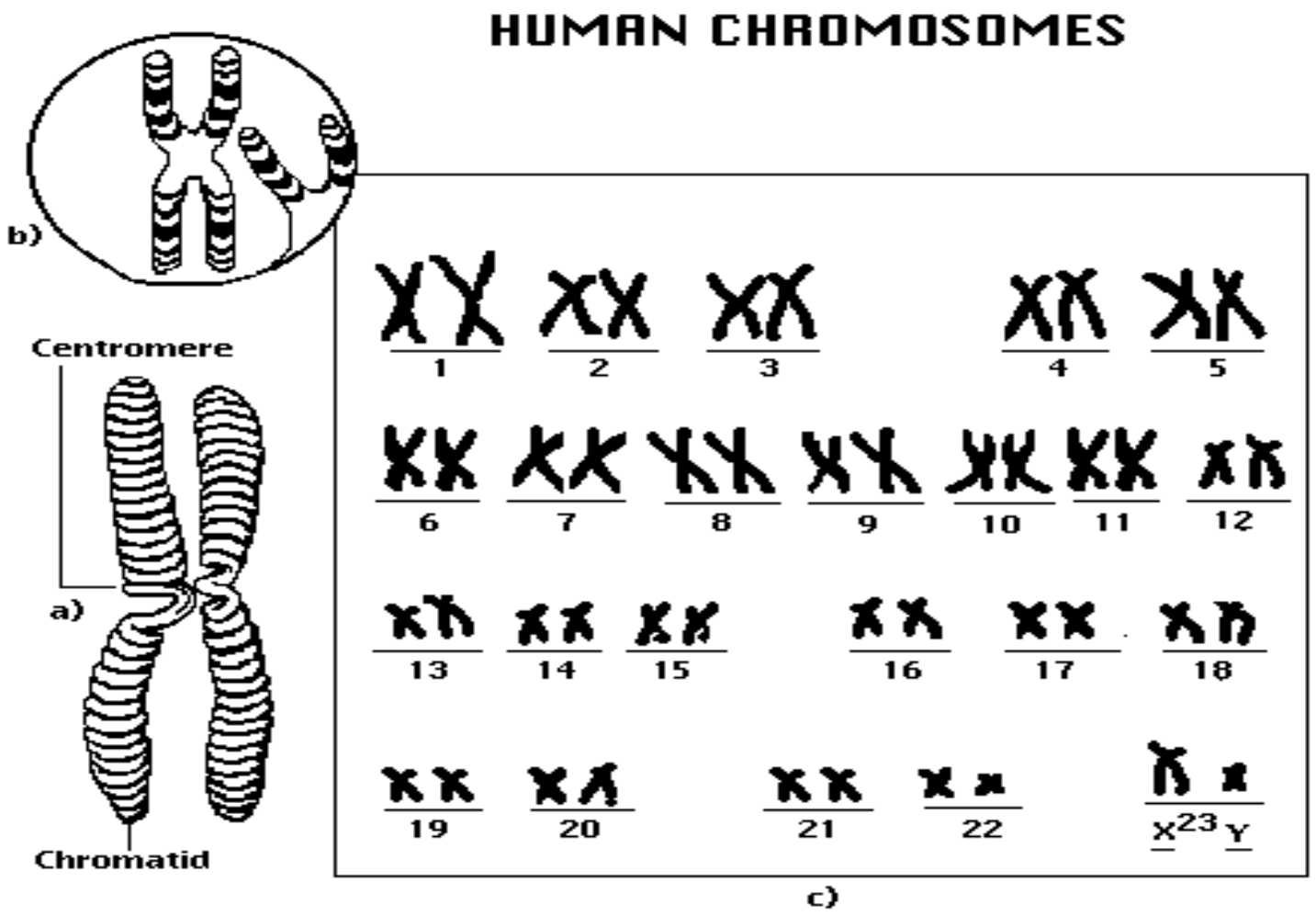
Hierarchy in Biology



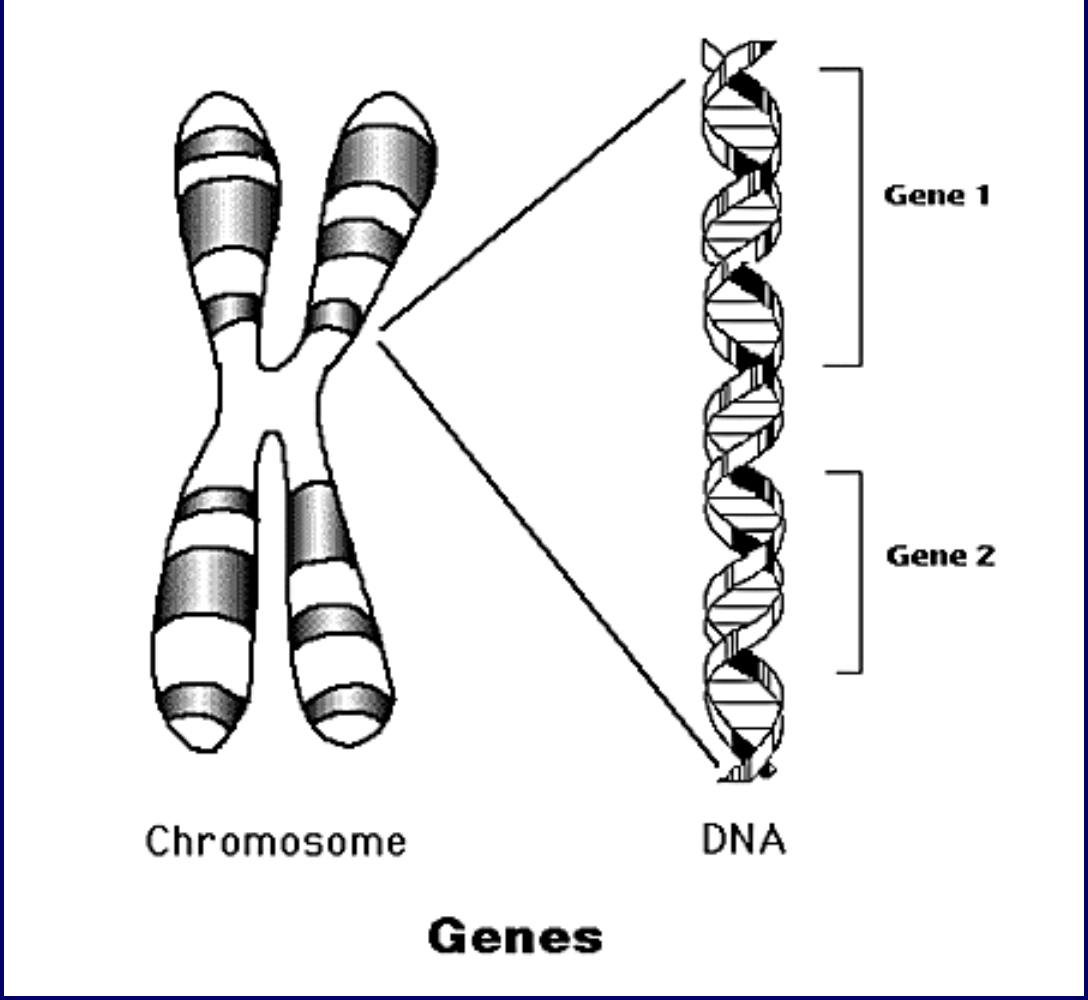
Animal cell


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

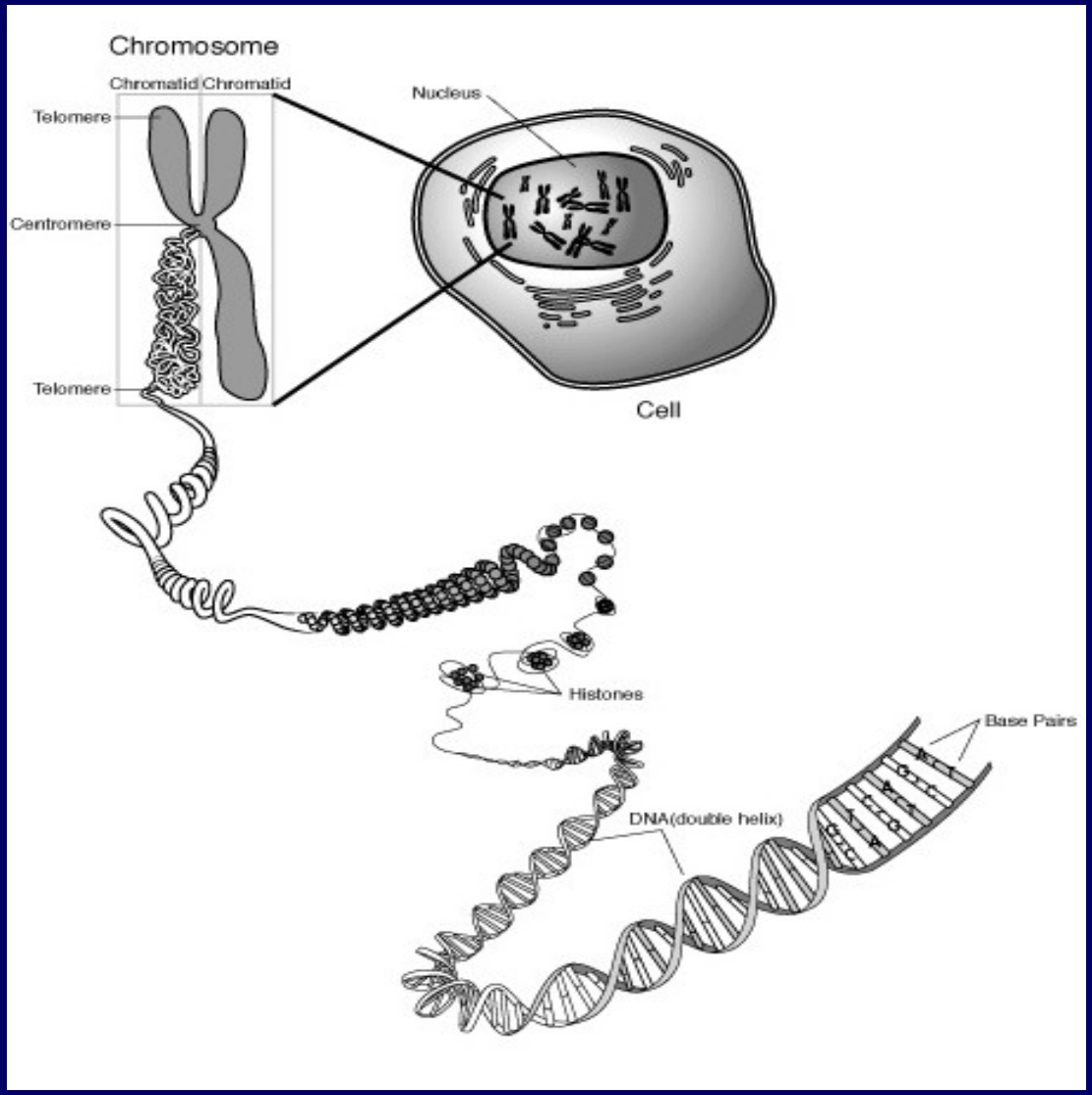
Human Chromosomes



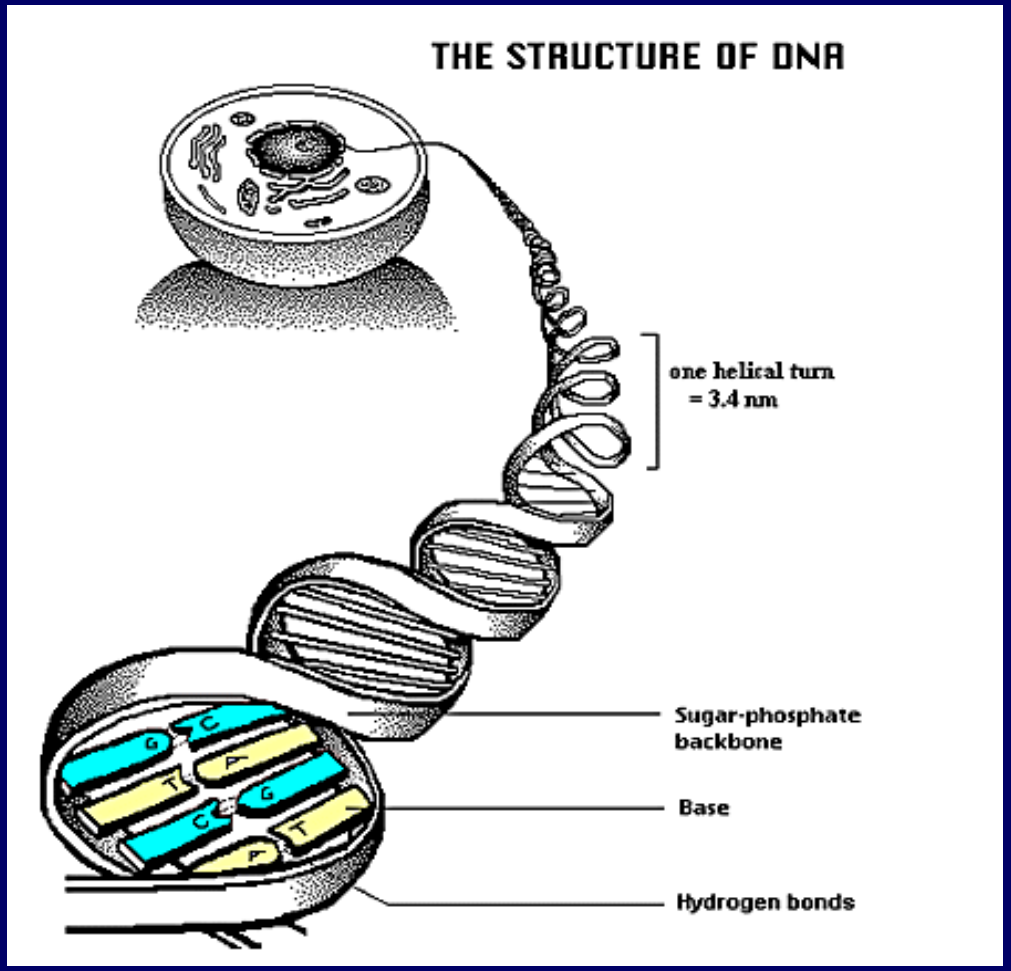
Genes are linearly arranged along chromosomes



Chromosomes and DNA

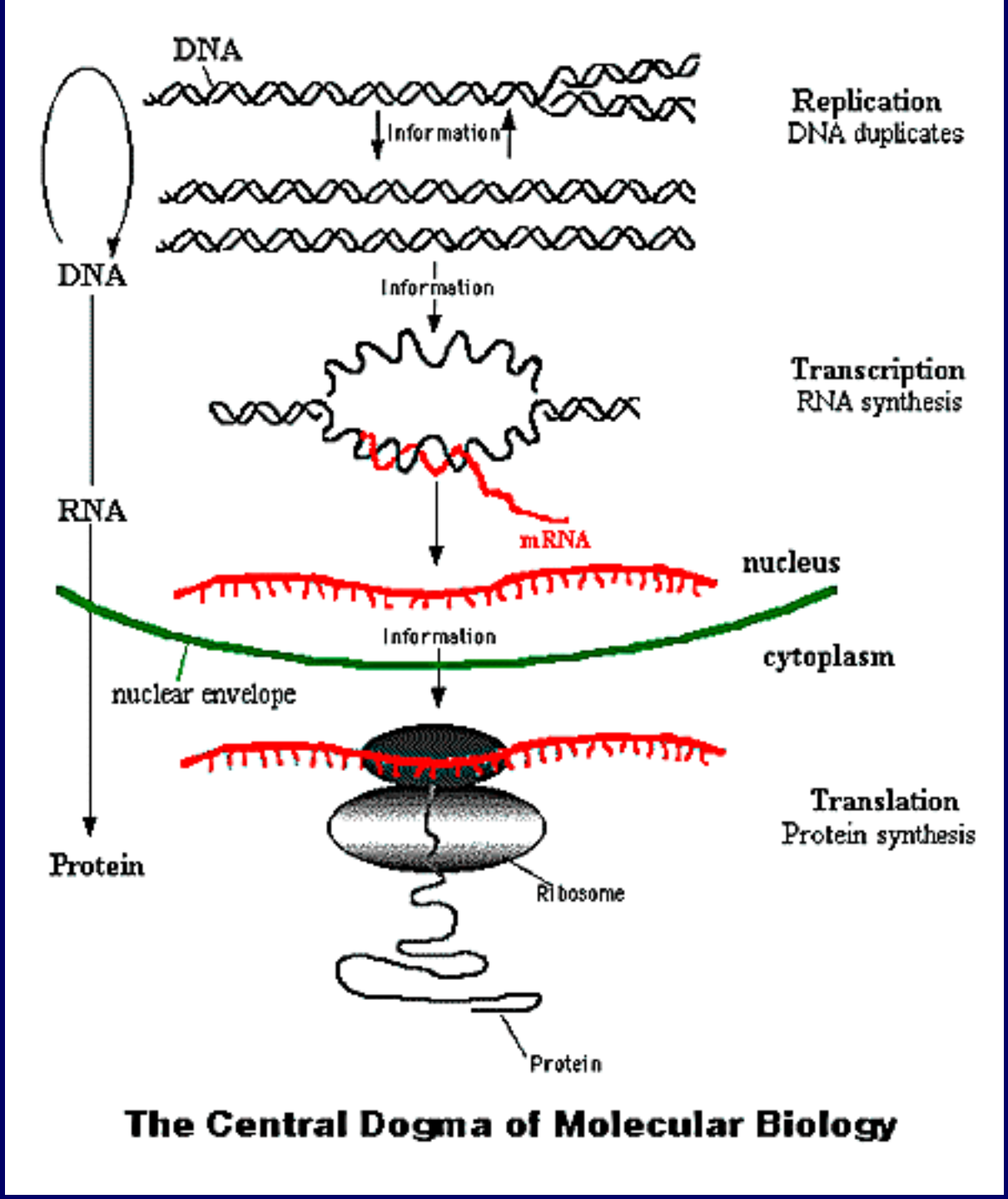


DNA can be simplified to a string of four letters



GATTACA

(RT)


DNA
Information
Replication
DNA duplicates
DNA
Information
Transcription
RNA synthesis
RNA
mRNA
nucleus
Information
cytoplasm
nuclear envelope
Translation
Protein synthesis
Protein
Ribosome
Protein
The Central Dogma of Molecular Biology

Sequence to Structure: It's a matter of dimensions!

- 1D Nucleic acid sequence

AGT-TTC-CCA-GGG…

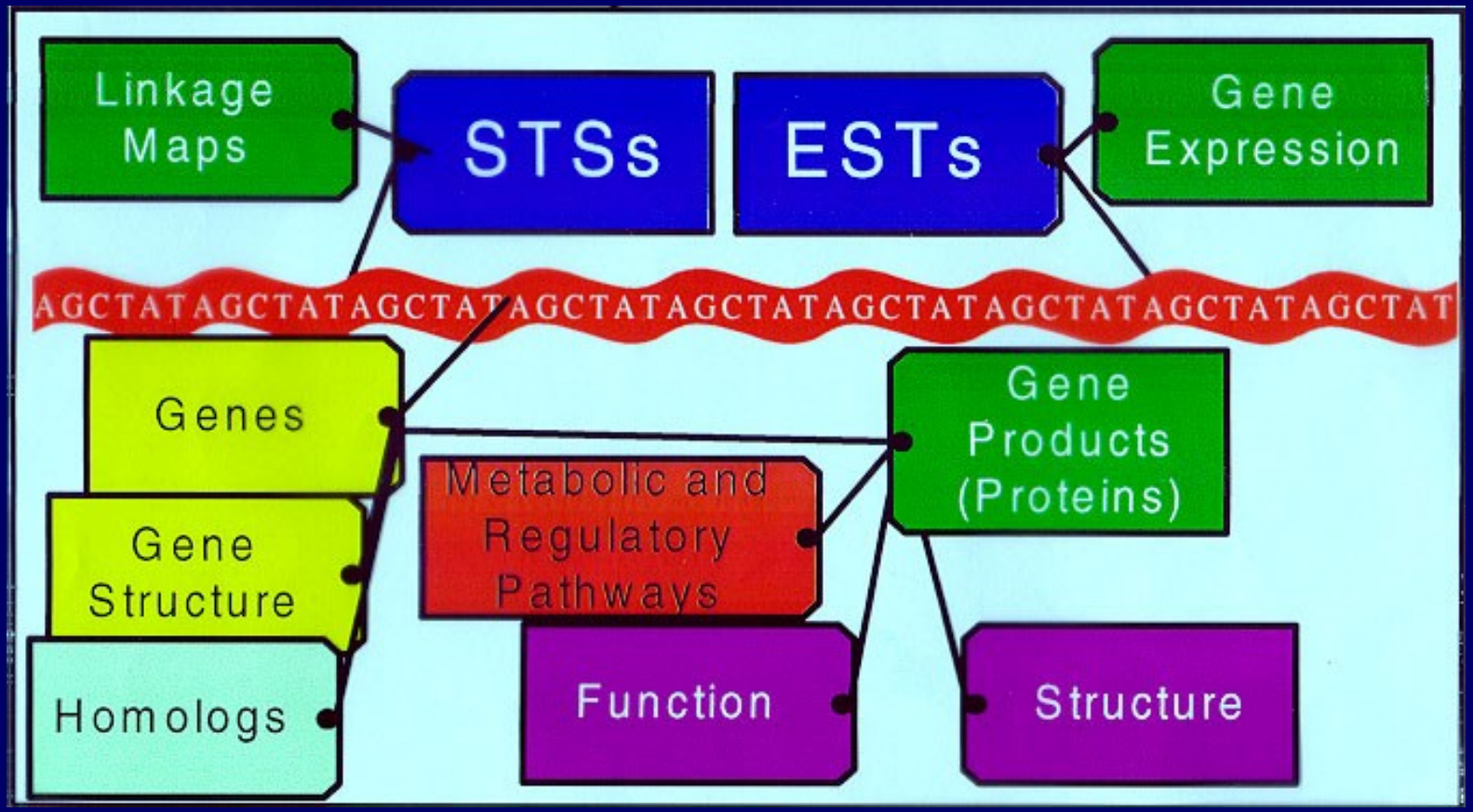
- 1D Protein sequence

Met-Ala-Gly-Lys-His…
M – A – G – K – H…

- 3D Spatial arrangement of atoms

Genome Annotation

The Process of Adding Biology Information and Predictions to a Sequenced Genome Framework



What we are doing?

- FTG: A web server for locating probable protein coding region in nucleotide sequence using fourier tranform approach (Issac, B., Singh, H., Kaur, H. and Raghava, G.P.S. (2002) Bioinformatics 18:196).

- EGPred:Similarity Aided Ab Initio Method of Gene Prediction This server allows to predict gene (protein coding regions) in eukaryote genomes that includes introns and exons, using similarity aided (double) and consensus Ab Intion methods (Issac B and Raghava GPS (2004) Genome Research (In press)).

- SVMgene: It is a support vector based approach to identify the protein coding regions in human genomic DNA.

- SRF: Spectral Repeat Finder (SRF) is a program to find repeats through an analysis of the power spectrum of a given DNA sequence. By repeat we mean the repeated occurrence of a segment of N nucleotides within a DNA sequence. SRF is an ab initio technique as no prior assumptions need to be made regarding either the repeat length, its fidelity, or whether the repeats are in tandem or not (Sharma et al. (2004) Bioinformatics, In Press)..

Protein Sequence Alignment and Database Searching

- **Alignment of Two Sequences (Pair-wise Alignment)**
  - The Scoring Schemes or Weight Matrices
  - Techniques of Alignments
  - DOTPLOT
- **Multiple Sequence Alignment (Alignment of > 2 Sequences)**
  - Extending Dynamic Programming to more sequences
  - Progressive Alignment (Tree or Hierarchical Methods)
  - Iterative Techniques
    - Stochastic Algorithms (SA, GA, HMM)
    - Non Stochastic Algorithms
- **Database Scanning**
  - FASTA, BLAST, PSIBLAST, ISS
- **Alignment of Whole Genomes**
  - MUMmer (Maximal Unique Match)

What we are doing?

- GWFASTA: Genome Wise Sequence Similarity Search using FASTA. It allow user to search their sequence against sequenced genomes and their product proteome. This integrate various tools which allows analysys of FASTA search (Issac, B. and Raghava, G.P.S. (2002) Biotechniques 33:548-56)
- GWBLAST: A genome wide blast server. It allow user to search ther sequence against sequenced genomes and annonated proteomes. This integrate various tools which allows analysys of BLAST SEARCH
- Protein Sequence Analysis -> This server allow user to analysis of protein sequence and present the analysis in Graphical and Textual format. This allows property plots of 36 parameter (like Hydrophobicity Plot, Polarity, Charge) of single aminoacid sequence and multiple sequence alignment (Raghava, G.P.S. (2001) Biotech Software and Internet Report, 2:255).
- RPFOLD: Recognition of Protein Fold -> RPFOLD server allows to predict top 5 similar fold in PDB (Protein DataBank) for a ginen protein sequence (query)
- **OXBench:** Evaluation of protein multiple sequence alignment (Raghava et al. BMC Bioinformatics 4:47) .

Traditional Proteomics

- 1D gel electrophoresis (SDS-PAGE)
- 2D gel electrophoresis
- Protein Chips
  - Chips coated with proteins/Antibodies
  - large scale version of ELISA
- Mass Spectrometry
  - MALDI: Mass fingerprinting
  - Electrospray and tandem mass spectrometry
    - Sequencing of Peptides (N->C)
    - Matching in Genome/Proteome Databases

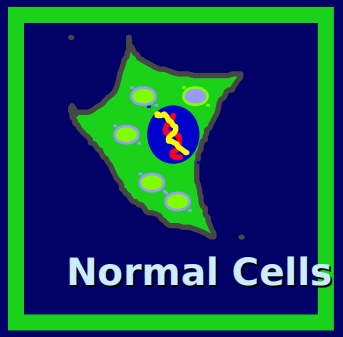
Overview of 2D Gel

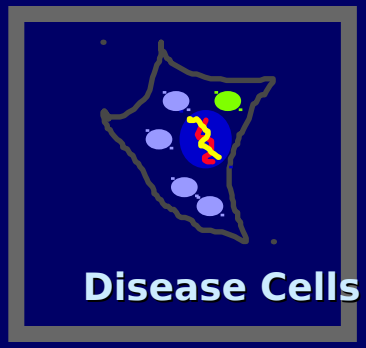
- SDS-PAGE + Isoelectric focusing (IEF)
  - Gene Expression Studies
  - Medical Applications
  - Sample Experiments
- Capturing and Analyzing Data
  - Image Acquistion
  - Image Sizing & Orientation
  - Spot Identification
  - Matching and Analysis

Comparision/Matcing of Gel Images

- Compare 2 gel images
  - Set X and y axis
  - Overlap matching spots
  - Compare intensity of spots
- Scan against database
  - Compare query gel with all gels
  - Calculate similarity score
  - Sort based on score

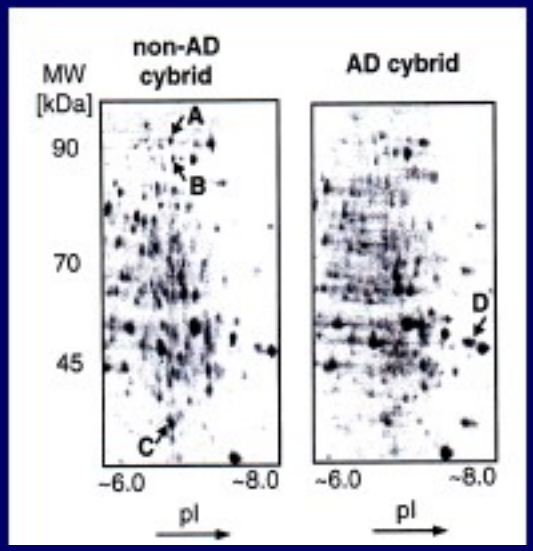
Proteomics: Fingerprints of Disease





Phenotypic Changes

- Differential protein expressi
- Protein nitration patterns
- Altered phosporylation
- Altered glycosylation profile



Utility

- Target discovery
- Disease pathways
- Disease biomarkers

Fingerprinting Technique

- What is fingerprinting
  - It is technique to create specific pattern for a given organism/person
  - To compare pattern of query and target object
  - To create Phylogenetic tree/classification based on pattern

- Type of Fingerprinting
  - DNA Fingerprinting
  - Mass/peptide fingerprinting
  - Properties based (Toxicity, classification)
  - Domain/conserved pattern fingerprinting
- Common Applications
  - Paternity and Maternity
  - Criminal Identification and Forensics
  - Personal Identification
  - Classification/Identification of organisms
  - Classification of cells

Fingerprinting Techniques
What we are doing?

- AC2DGel: is a web server for analysis and comparison of two-dimensional electrophoresis (2-DE) Gel images. It helps in annotating the virual 2-D gel image proteins on the basis of known molecular weight andpH scales of the markers.

- DNASIZE: Computation of DNA/Protein size -> This web-server allow to compute the length of DNA or protein fragments from its electropheric mobility using a graphical method (Raghava, G. P. S. (2001) Biotech Software and Internet Report, 2:198)
- GMAP: a multipurpose computer program to aid synthetic gene design, cassette mutagenesis and introduction of potential restriction sites into DNA sequences (Raghava GPS (1994) Biotechniques 16: 1116-1123).

- DNAOPT : A computer program to aid optimization of gel conditions of DNA gel electrophoresis and SDS-PAGE. (Raghava GPS (1994) Biotechniques 18: 274-81).

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

VACCINES

A. SUCCESS STORY:

- COMPLETE ERADICATION OF SMALLPOX
- WHO PREDICTION : ERADICATION OF PARALYTIC POLIO THROUGHOUT THE WORLD BY YEAR 2004
- SIGNIFICANT REDUCTION OF INCIDENCE OF DISEASES: DIPTHERIA, MEASLES, MUMPS, PERTUSSIS, RUBELLA, POLIOMYELITIS, TETANUS

B.NEED OF AN HOUR

1) SEARCH FOR NONAVAILABILE EFFECTIVE VACCINES FOR DISEASES LIKE:
MALARIA, TUBERCULOSIS AND AIDS

2) IMPROVEMENT IN SAFETY AND EFFICACY OF PRESENT VACCINES

3) LOW COST

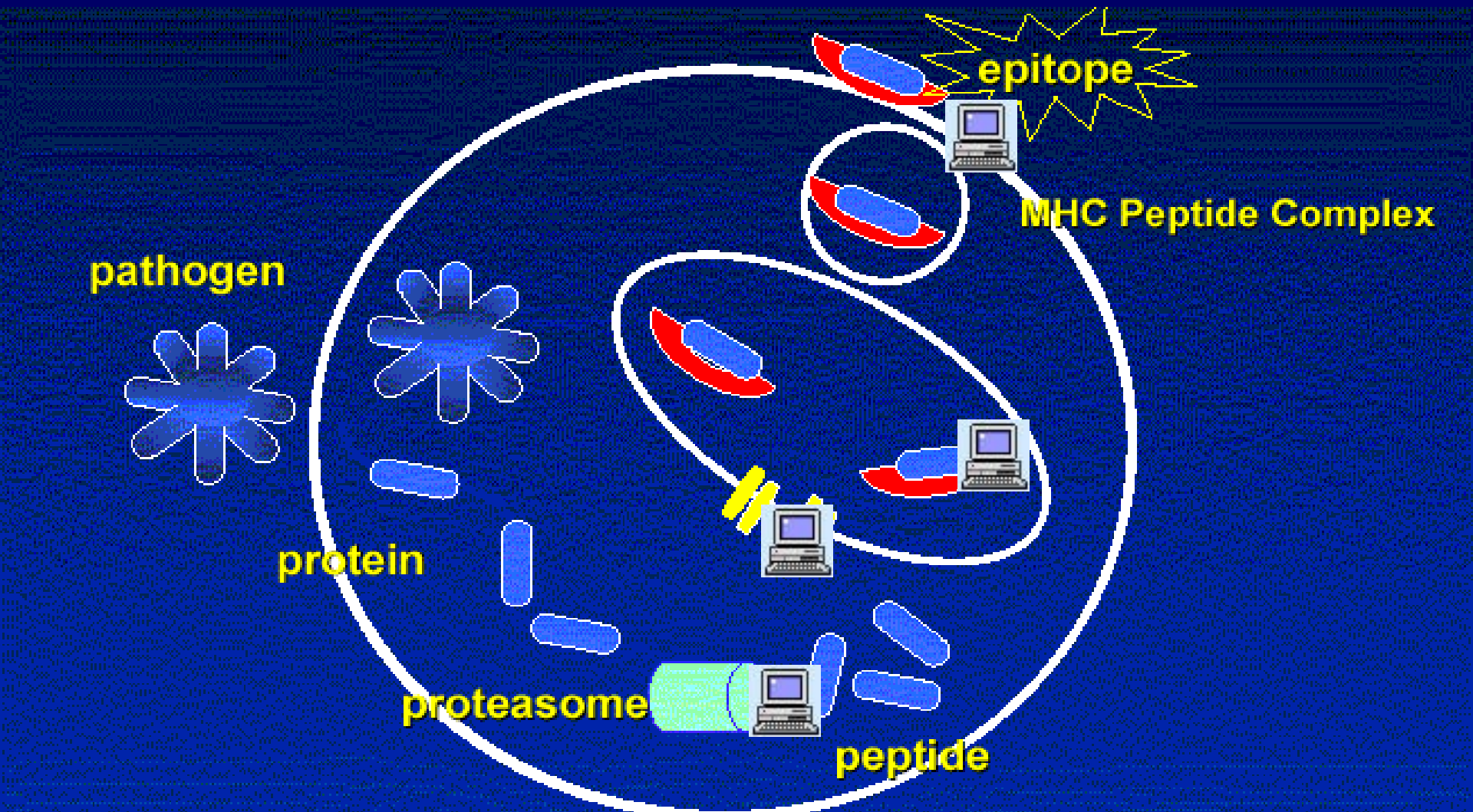
4) EFFICIENT DELIVERY TO NEEDY

5) REDUCTION OF ADVERSE SIDE EFFECTS

Computer Aided Vaccine Design

- Whole Organism of Pathogen
  - Consists more than 4000 genes and proteins
  - Genomes have millions base pair
- Target antigen to recognise pathogen
  - Search vaccine target (essential and non-self)
  - Consists of amino acid sequence (e.g. A-V-L-G-Y-R-G-C-T ……)
- Search antigenic region (peptide of length 9 amino acids)

Major steps of endogenous antigen processing



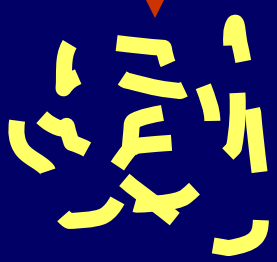
Computer Aided Vaccine Design

- Problem of Pattern Recognition
  - ATGGTRDAR — Epitope
  - LMRGTCAAY — Non-epitope
  - RTTGTRAWR — Epitope
  - EMGGTCAAY — Non-epitope
  - ATGGTRKAR — Epitope
  - GTCVGYATT — Epitope
- Commonly used techniques
  - Statistical (Motif and Matrix)
  - AI Techniques

Why computational tools are required for prediction.

200 aa proteins

Chopped to overlapping peptides of 9 amino acids



Bioinformatics Tools

192 peptides

10-20

invitro or *invivo* experiments for detecting which snippets of protein will spark an immune response.

Immunounformatics: Computer Aided Vaccine Design

What we are doing?

- MHC Class II binding peptide -> Matrix Optimization Technique for Predicting MHC binding Core (Singh, H. and Raghava, G. P. S. (2002) Biotech Software and Internet Report, 3:146)
- MMBPred Prediction of of MHC class I binders which can bind to wide range of MHC alleles with high affinity. This server has potential to develop sub-unit vaccine for large population (Bhasin, M., and Raghava, G.P.S. (2003) Hybridoma and Hybridomics 22: 229)
- nHLAPred: Prediction of MHC Class I Restricted T Cell Epitopes -> This server allow to predict binding peptide for 67 MHC Class I alleles. This also allow to predict the proteasome cleavage site and binding peptide that have cleavage site at C terminus (potential T cell epitopes). This uses the hybrid approach for prediction (Neural Network + Quantitative Matrix)
- ProPred1: Prediction of MHC Class I binding peptide -> The aim of this server is to predict MHC Class-I binding regions in an antigen sequence (Singh, H. and Raghava, G.P.S. (2003) Bioinformatics, 19: 1009)
- ProPred: Prediction of MHC Class II binding peptide -> The aim of this server is to predict MHC Class-II binding regions in an antigen sequence (Singh, H. and Raghava, G. P. S. (2001) Bioinformatics 17: 1236)
- CTLPred: Direct method of prediction of CTL Epitopes in an antigen sequence. This server utlize the machine learning techniques Support Vector Machine(SVM) and Aritificial Neural Network (ANN) for prediction (Bhasin, M. and Raghava, G. P. S. (2004) Vaccine (In Press))

Immunounformatics: Computer Aided Vaccine Design

What we are doing?

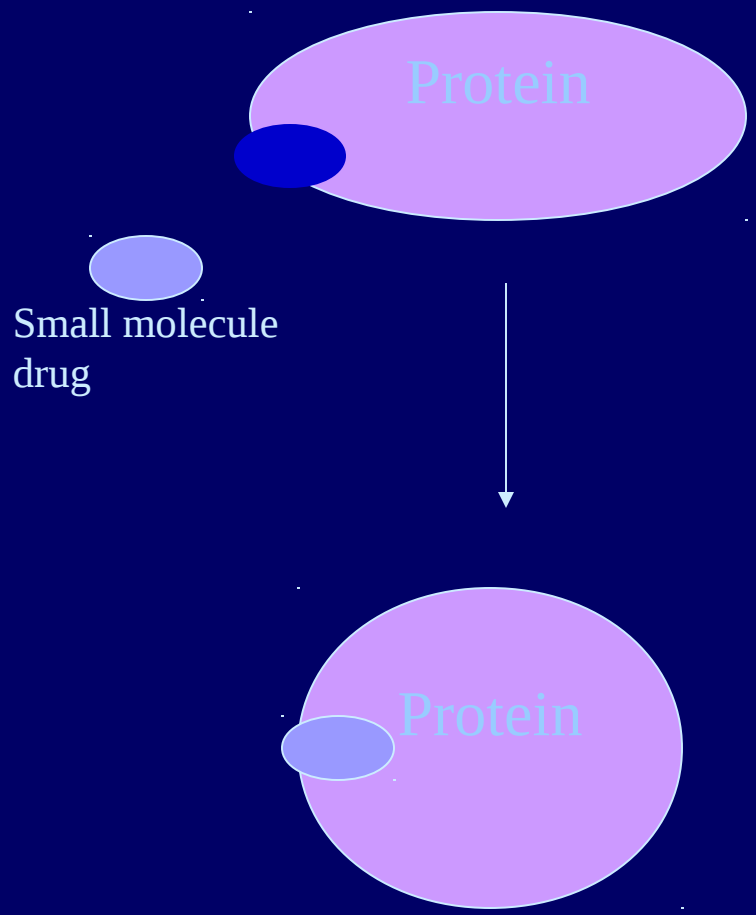
- HLADR4Pred: SVM and ANN based methods for predicting HLA-DRB1*0401 binding peptides in an Antigen Sequence (Bhasin, M. and Raghava, G.P.S. (2003) Bioinformatics 20:421).
- TAPPred: TAPPred is an on-line service for predicting binding affinity of peptides toward the TAP transporter. The Prediction is based on cascade SVM, using sequence and properties of the the amino acids(Bhasin, M. and Raghava, G. P. S. (2004) Protein Science 13:596-607).
- ABCpred: server is to predict linear B cell epitope regions in an antigen sequence, using artificial neural network. This server will assist in locating epitope regions that are useful in selecting synthetic vaccine candidates, disease diagonosis and also in allergy research.
- MHCBN: The MHCBN is a curated database consisting of detailed information about Major Histocompatibility Complex (MHC) Binding,Non-binding peptides and T-cell epitopes.The version 3.1 of database provides information about peptides interacting with TAP and MHC linked autoimmune diseases (Bhasin, M., Singh, H. and Raghava, G. P. S. (2003) Bioinformatics 19: 665). This databse is also launched by European Bioinformatics Institute (EBI) Hinxton, Cambridge, UK.
- BCIPep: is collection of the peptides having the role in Humoral immunity. The peptides in the database has varying measure of immunogenicity.This database can assist in the development of method for predicting B cell epitopes, desigining synthetic vaccines and in disease diagnosis. This databse is also launched by European Bioinformatics Institute (EBI) Hinxton, Cambridge, UK.

Drug 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.
- **Chemical Modification of Known Drugs**
  - Drug improvement by chemical modification
  - Pencillin G -> Methicillin; morphine->nalorphine

A simple example



Protein disabled … disease cured

Chemoinformatics

Small molecule drug

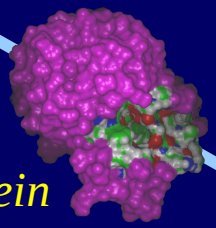
- Large databases
- Not all can be drugs
- Opportunity for data mining techniques

Bioinformatics

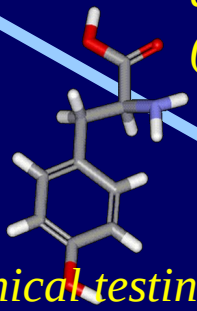
- Large databases
- Not all can be drug targets
- Opportunity for data mining techniques

Drug Discovery & Development

Identify disease



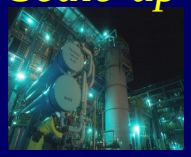
Isolate protein involved in disease (2-5 years)



Find a drug effective against disease protein (2-5 years)

Preclinical testing (1-3 years)

Scale-up



File IND

Human clinical trials (2-10 years)

Formulation

File NDA

FDA approval (2-3 years)

Technology is impacting this process

Identify disease

GENOMICS, PROTEOMICS & BIOPHARM.

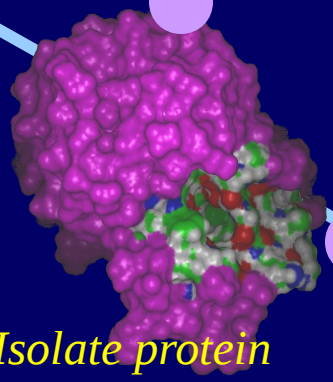
Potentially producing many more targets and "personalized" targets

HIGH THROUGHPUT SCREENING

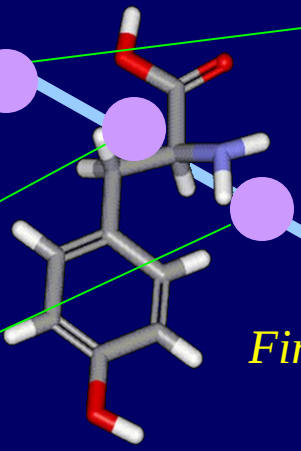
Screening up to 100,000 compounds a day for activity against a target protein

VIRTUAL SCREENING

Using a computer to predict activity



Isolate protein



COMBINATORIAL CHEMISTRY

Rapidly producing vast numbers of compounds

Find drug

MOLECULAR MODELING

Computer graphics & models help improve activity

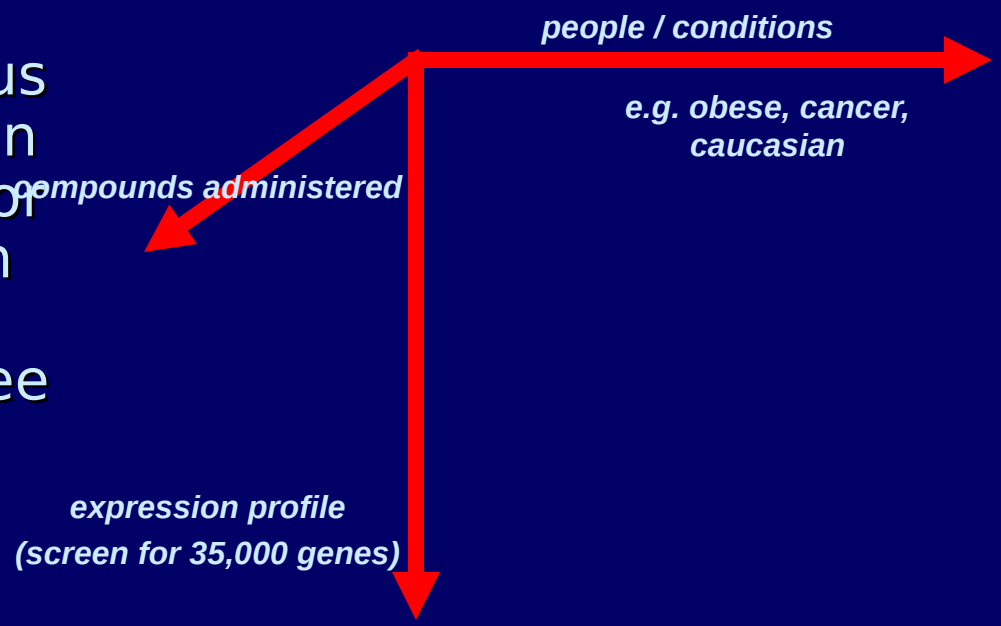
Preclinical testing

IN VITRO & IN SILICO ADME MODELS

Tissue and computer models begin to replace animal testing

1. Gene Chips

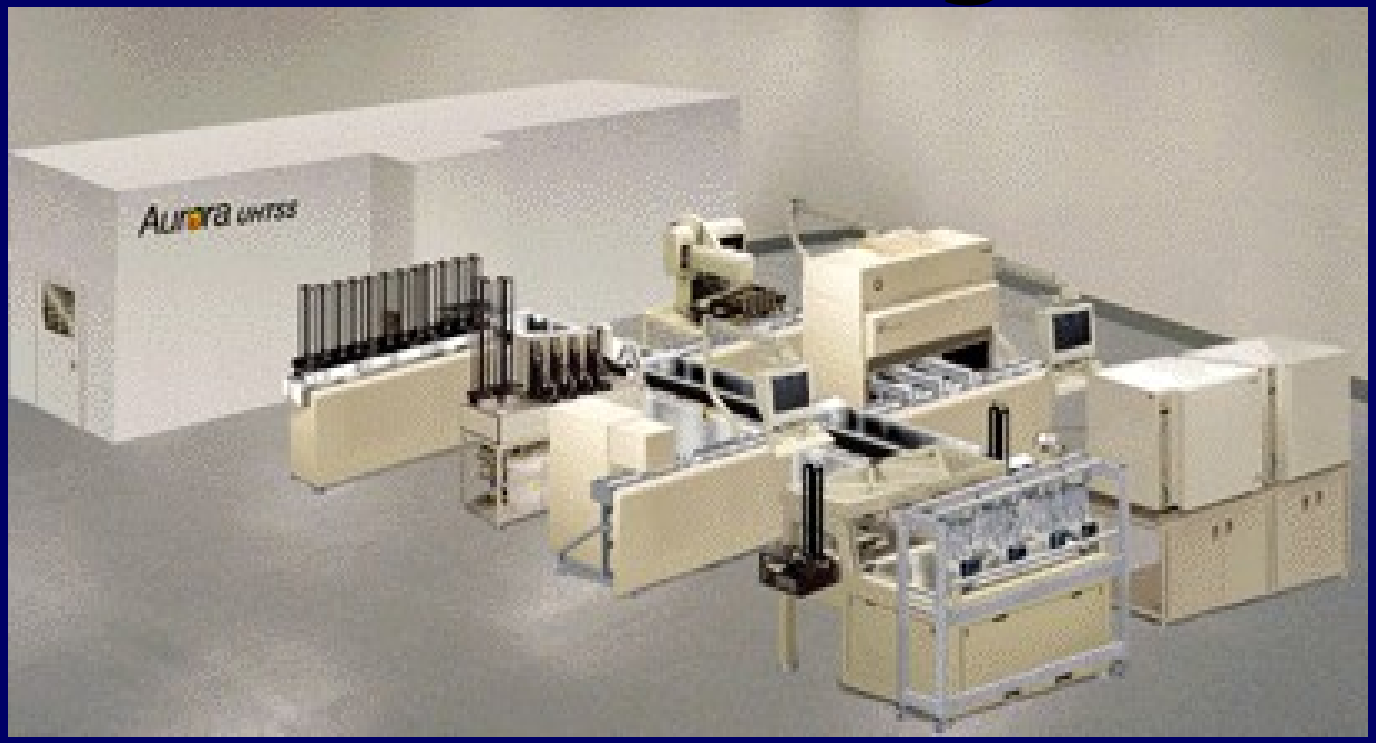
- “Gene chips” allow us to look for changes in protein expression for different people with a variety of conditions, and to see if the presence of drugs changes that expression
- Makes possible the design of drugs to target different phenotypes



Biopharmaceuticals

- Drugs based on proteins, peptides or natural products instead of small molecules (chemistry)
- Pioneered by biotechnology companies
- Biopharmaceuticals can be quicker to discover than traditional small-molecule therapies
- Biotechs now paring up with major pharmaceutical companies

2. High-Throughput Screening



Screening perhaps millions of compounds in a corporate collection to see if any show activity against a certain disease protein

High-Throughput Screening

- Drug companies now have millions of samples of chemical compounds
- High-throughput screening can test 100,000 compounds a day for activity against a protein target
- Maybe tens of thousands of these compounds will show some activity for the protein
- The chemist needs to intelligently select the 2 - 3 classes of compounds that show the most promise for being drugs to follow-up

Informatics Implications

- Need to be able to store chemical structure and biological data for millions of datapoints
  - *Computational representation of 2D structure*
- Need to be able to organize thousands of active compounds into meaningful groups
  - *Group similar structures together and relate to activity*
- Need to learn as much information as possible from the data (data mining)
  - *Apply statistical methods to the structures and related information*

3. Computational Models of Activity

- Machine Learning Methods
  - E.g. Neural nets, Bayesian nets, SVMs, Kahonen nets
  - Train with compounds of known activity
  - Predict activity of “unknown” compounds
- Scoring methods
  - Profile compounds based on properties related to target
- Fast Docking
  - Rapidly “dock” 3D representations of molecules into 3D representations of proteins, and score according to how well they bind

4. Combinatorial Chemistry

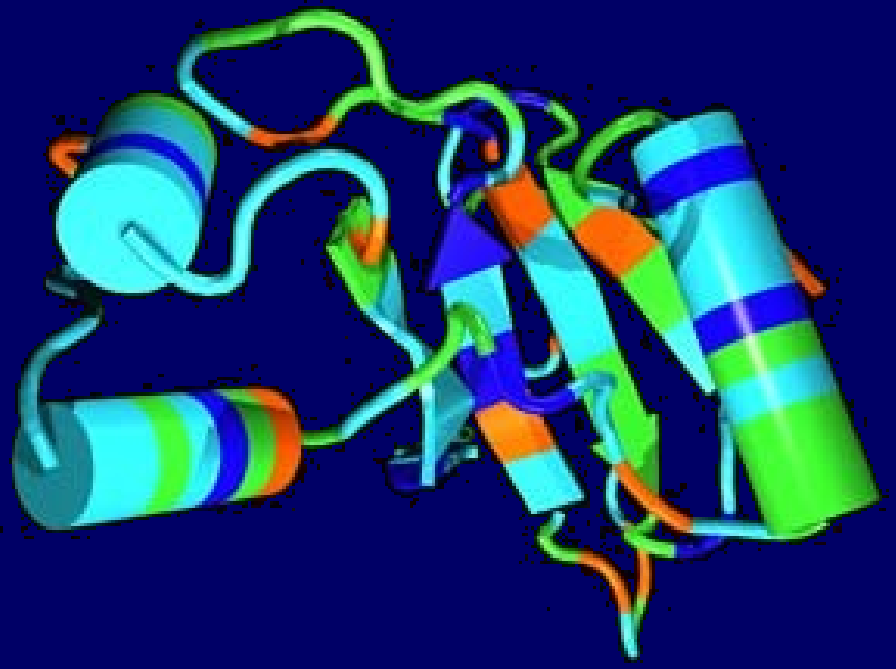
- By combining molecular “building blocks”, we can create very large numbers of different molecules very quickly.

- Usually involves a “scaffold” molecule, and sets of compounds which can be reacted with the scaffold to place different structures on “attachment points”.

Combinatorial Chemistry Issues

- Which R-groups to choose
- Which libraries to make
  - “Fill out” existing compound collection?
  - Targeted to a particular protein?
  - As many compounds as possible?
- Computational profiling of libraries can help
  - “Virtual libraries” can be assessed on computer

5. Molecular Modeling

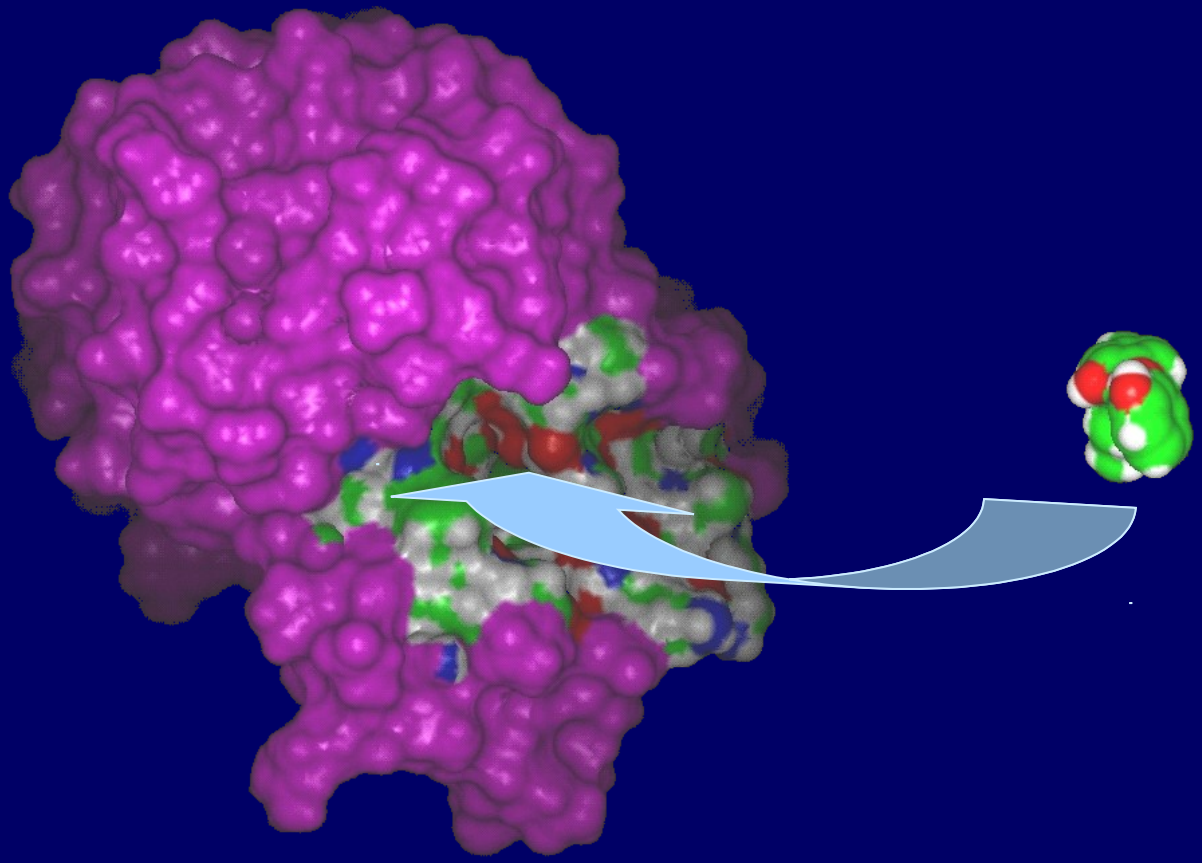


- *3D Visualization of interactions between compounds and proteins*
- *“Docking” compounds into proteins computationally*

3D Visualization

- X-ray crystallography and NMR Spectroscopy can reveal 3D structure of protein and bound compounds
- Visualization of these “complexes” of proteins and potential drugs can help scientists understand the mechanism of action of the drug and to improve the design of a drug
- Visualization uses computational “ball and stick” model of atoms and bonds, as well as surfaces
- Stereoscopic visualization available

“Docking” compounds into proteins computationally



6. In Vitro & In Silico ADME models

- Traditionally, animals were used for pre-human testing. However, animal tests are expensive, time consuming and ethically undesirable
- ADME (Absorbtion, Distribution, Metabolism, Excretion) techniques help model how the drug will likely act in the body
- These methods can be experemental (*in vitro*) using cellular tissue, or *in silico*, using computational models

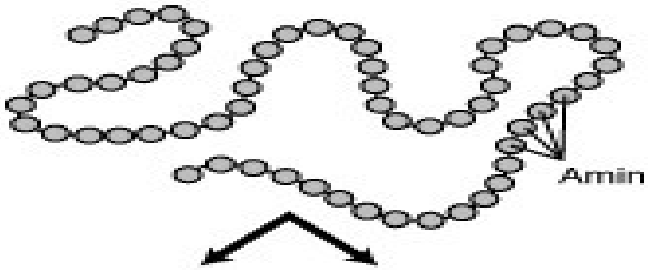
Size of databases

- Millions of entries in databases
  - CAS : 23 million
  - GeneBank : 5 million
- Total number of drugs worldwide: 60,000
- Fewer than 500 characterized molecular targets
- Potential targets : 5,000-10,000

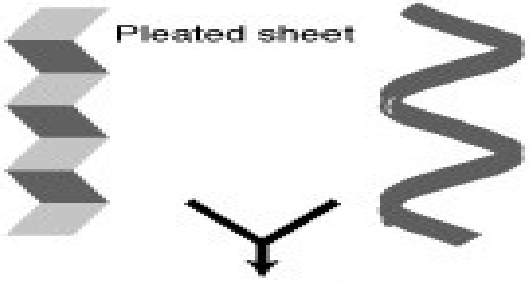
Protein Structure Prediction

- Experimental Techniques
  - X-ray Crystallography
  - NMR
- Limitations of Current Experimental Techniques
  - Protein DataBank (PDB) -> 24000 protein structures
  - SwissProt -> 100,000 proteins
  - Non-Redudant (NR) -> 1,000,000 proteins
- Importance of Structure Prediction
  - Fill gap between known sequence and structures
  - Protein Engg. To alter function of a protein
  - Rational Drug Design

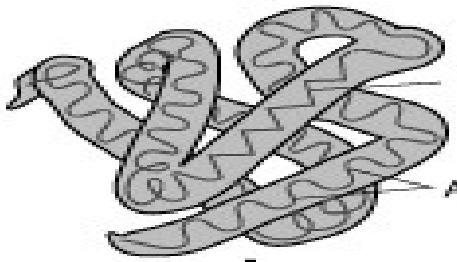
Protein Structures



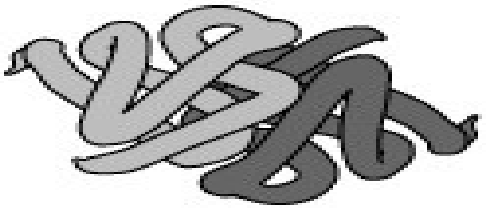
Primary protein structure
is sequence of a chain of amino acids



Secondary protein structure
occurs when the sequence of amino acids are linked by hydrogen bonds



Tertiary protein structure
occurs when certain attractions are present between alpha helices and pleated sheets.



Quaternary protein structure
is a protein consisting of more than one amino acid chain.

Techniques of Structure Prediction

- Computer simulation based on energy calculation
  - Based on physio-chemical principles
  - Thermodynamic equilibrium with a minimum free energy
  - Global minimum free energy of protein surface
- Knowledge Based approaches
  - Homology Based Approach
  - Threading Protein Sequence
  - Hierarchical Methods

Energy Minimization Techniques

Energy Minimization based methods in their pure form, make no priori assumptions and attempt to locate global minma.

- **Static Minimization Methods**
  - Classical many potential-potential can be construted
  - Assume that atoms in protein is in static form
  - Problems(large number of variables & minima and validity of potentials)
- **Dynamical Minimization Methods**
  - Motions of atoms also considered
  - Monte Carlo simulation (stochastics in nature, time is not cosider)
  - Molecular Dynamics (time, quantum mechanical, classical equ.)
- Limitations
  - large number of degree of freedom,CPU power not adequate
  - Interaction potential is not good enough to model

Knowledge Based Approaches

- Homology Modelling
  - Need homologues of known protein structure
  - Backbone modelling
  - Side chain modelling
  - Fail in absence of homology
- Threading Based Methods
  - New way of fold recognition
  - Sequence is tried to fit in known structures
  - Motif recognition
  - Loop & Side chain modelling
  - Fail in absence of known example

Hierarcial Methods

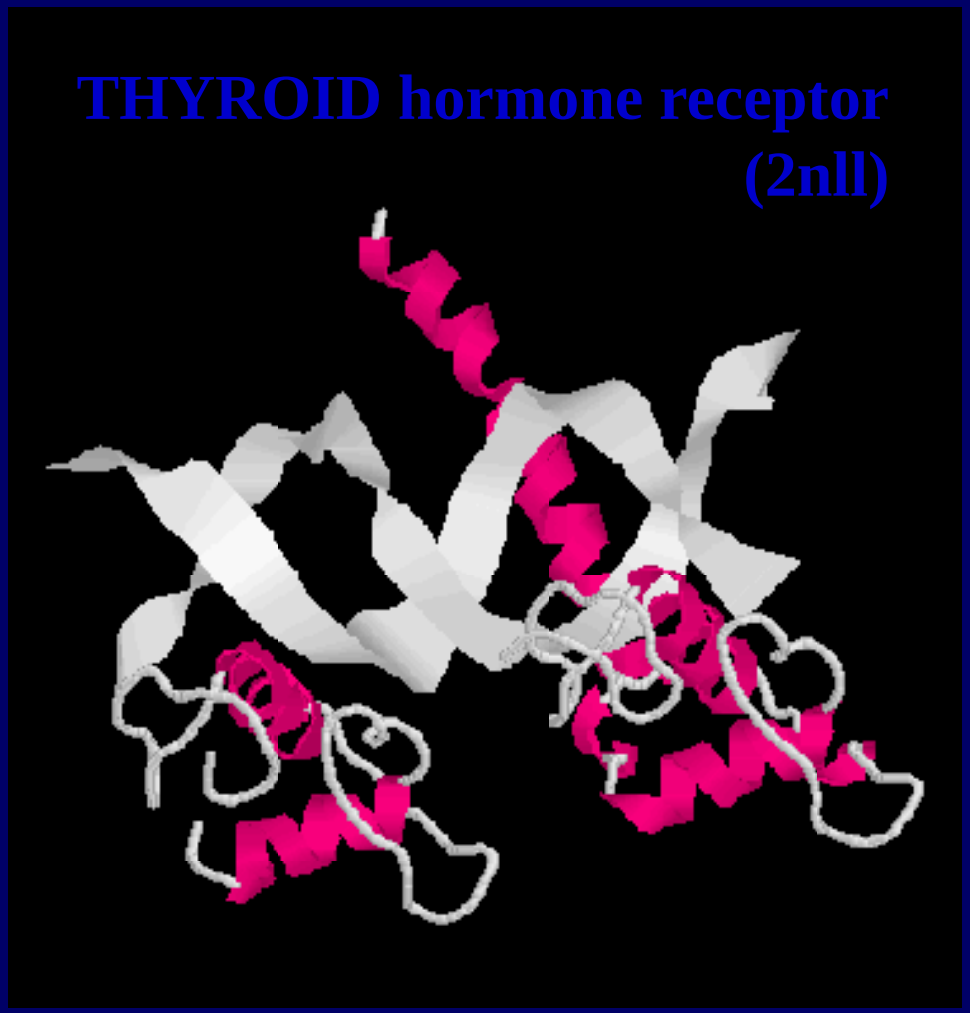
Intermidiate structures are predicted, instead of predicting tertiary structure of protein from amino acids sequence

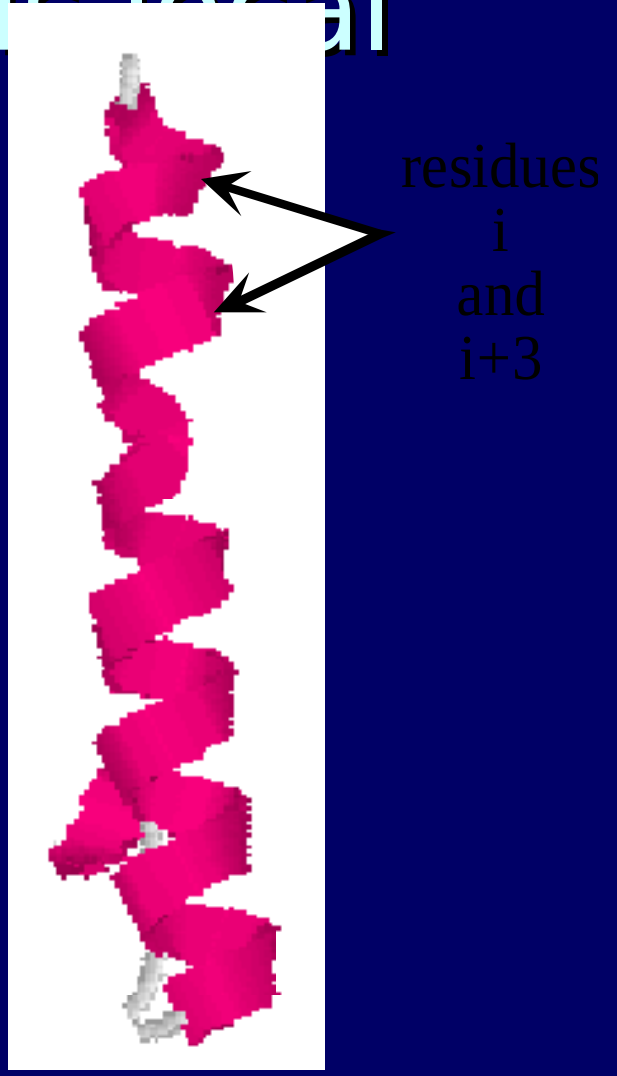
- Prediction of backbone structure
  - Secondary structure (helix, sheet,coil)
  - Beta Turn Prediction
  - Super-secondary structure
- Tertiary structure prediction
- Limitation

  Accuracy is only 75-80 %

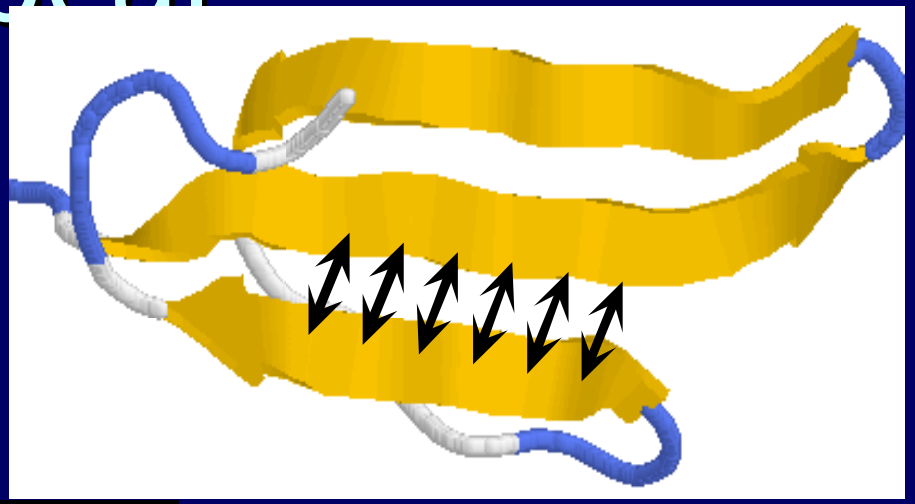
  Only three state prediction

Helix formation is local





β-sheet formation is NOT local

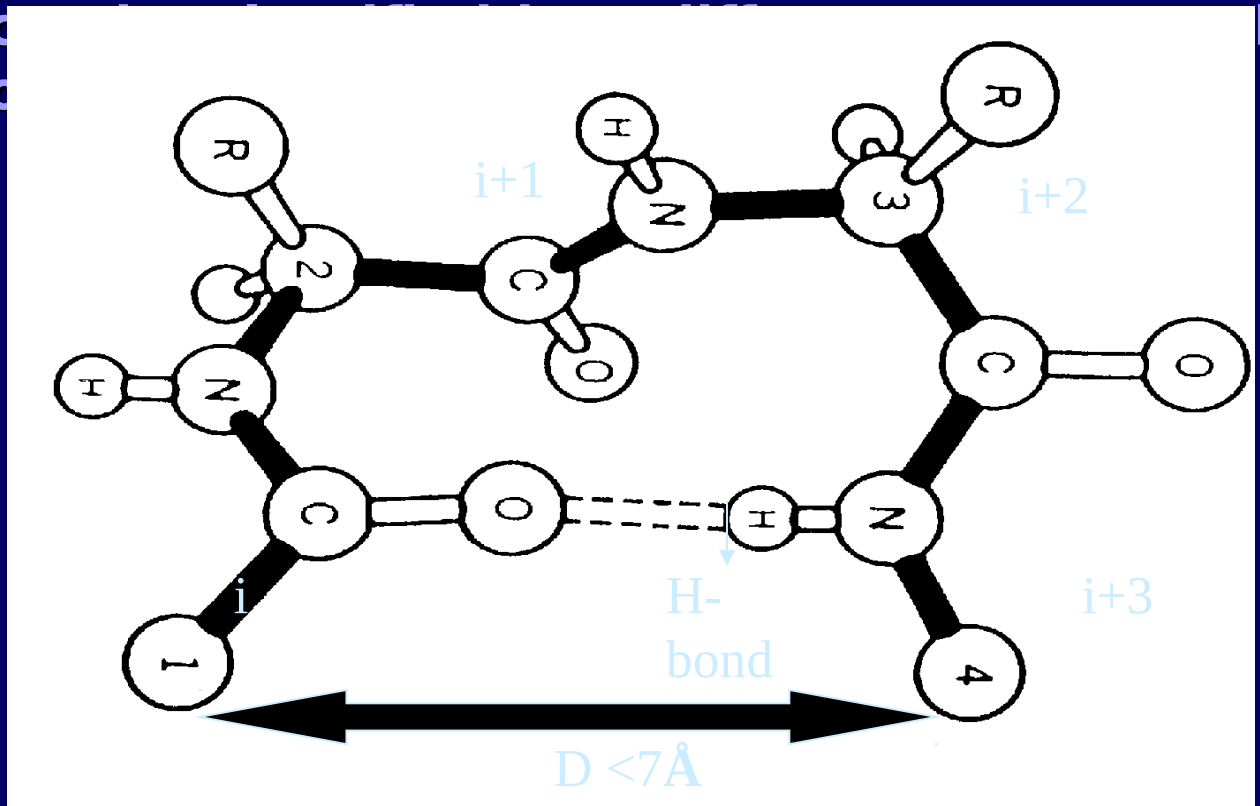


Erabutoxin β (3ebx)

Definition of β-turn

A β-turn is defined by four consecutive residues *i*, *i*+1, *i*+2 and *i*+3 that do not form a helix and have a Cα(*i*)-Cα(*i*+3) distance less than 7Å and the turn lead to reversal in the protein chain. (Richardson, 1981).

The conformation of β-turn is defined in terms of ϕ and ψ of two central residues, *i*+1 and *i*+2 and



Protein Structure Prediction
What we are doing?

- APSSP2: Advanced Protein Secondary Structure Prediction -> This server allow to predict the secondary structure of protein's from their amino acid sequence with high accuracy. It utilize the multiple alignment, neural network and MBR techniques. This server participates in number of world wide competition like CASP, CAFASP and EVA.
- Protein Structural Classes -> It predict weather protein belong to class Alpha or Beta or Alpha+Beta or Alpha/Beta (Raghava, G.P.S. (1999) J. Biosciences 24, 176)
- BTeval: Benchmarking of Beta Turn prediction methos on-line via Internet(Kaur, H. and Raghava G.P.S. Bioinformatics 18:1508-14). The user can see the performance of their method or existing methods (Kaur, H. and Raghava, G.P.S. (2003) Journal of Bioinformatics and Computational Biology 1:495-504)
- BetatTPred2: Prediction of Beta Turns in Proteins using Neural Network and multiple alignment techniques. This is highly accurate method for beta turn prediction (Kaur, H. and Raghava, G.P.S. (2003) Protein Science 12:627).
- GammaPred: Prediction of Gamma-turns in Proteins using Multiple Alignment and Secondary Structure Information (Kaur H. and Raghava, G.P.S. (2003) Protein Science; 12:923).
- AlphaPred: Prediction of Alpha-turns in Proteins using Multiple Alignment and Secondary Structure Information (Kaur & Raghava (2004) Proteins 55:83-90. (
- BetaTPred: A server for predicting Beta Turns in proteins using existing statistical methods. This allows consensus prediction from various methods (Kaur H., and Raghava G.P.S. (2002) Bioinformatics 18:498)

Protein Structure Prediction
What we are doing?

- CHpredict: The CHpredict server predict two types of interactions: C-H...O and C-H...PI interactions. For C-H...O interaction, the server predicts the residues whose backbone Calpha atoms are involved in interaction with backbone oxygen atoms and for C-H...PI interactions, it predicts the residues whose backbone Calpha atoms are involved in interaction with PI ring system of side chain aromatic moieties.

- AR_NHPred: A web server for predicting the aromatic backbone NH interaction in a given amino acid sequence where the pi ring of aromatic residues interact with the backbone NH groups. The method is based on the neural network training on PSI-BLAST generated position specific matrices and PSIPRED predicted secondary structure (Kaur,H. and Raghava G.P.S. (2004) Febs Lett. 564:47-57)

- TBBpred: Transmembrane Beta Barrel prediction server predicts the transmembrane Beta barrel regions in a given protein sequence. The server uses a forked strategy for predicting residues which are in transmembrane beta barrel regions. Prediction can be done based only on neural networks or based on statistical learning technique - SVM or combination of two methods (Natt et al. (2004) Proteins 56: 11-8).

- Betaturns: This server allows to predict the beta turns and type in a protein from their amino acid sequence (Kaur,H. and Raghava G.P.S. (2004)Bioinformatics (In press)) .

- PEPstr: The Pepstr server predicts the tertiary structure of small peptides with sequence length varying between 7 to 25 residues. The prediction strategy is based on the realization that ?-turn is an important and consistent feature of small peptides in addition to regular structures.

Selection of Target and Classification of Proteins

What we are doing?

- ESLpred: is a SVM based method for predicting subcellular localization of Eukaryotic proteins using dipeptide composition and PSIBLAST generated pfofile (Bhasin, M. and Raghava, G. P. S., 2004, Nucleic Acid Res. (In Press)). Using this server user may know the function of their protein based on its location in cell.

- NRpred: is a SVM based tool for the classification of nuclear receptors on the basis of amino acid composition or dipeptide composition. The overall prediction accuracy of amino acid composition and dipeptide composition based methods is 82.6% and 97.2% (Bhasin, M. and Raghava, G. P. S., 2004, Journal of Biological Chemistry (In Press)).

- GPCRpred: is a server for predicting G-protein-coupled receptors and for classifying them in families and sub-families. This server can play vital role in drug design, as GPCR are commonly used as drug targets (Bhasin, M. and Raghava, G. P. S., 2004, Nucleic Acid Res. (In Press))

- GPCRSclass: is a dipeptide composition based method for predicting Amine Type of G-protein-coupled receptors. In this method type amine is predicted from dipeptide composition of proteins using SVM.

Important Database of Hapten
What we are doing?

- Hapten: It is a small molecule, not immunogenic by itself, that can react with antibodies of appropriate specificity and elicit the formation of such antibodies when conjugated to a larger antigenic molecule (usually protein called carrier in this context). These hapten molecules are of great importance in the production of antibodies of desired specificity as antibody production involves activation of B lymphocytes by the hapten and helper T lymphocytes by the carrier protein.

- HaptenDB: It is a collection of haptens, information is collected and compiled from published literature and web resources. Presently database have more than 1700 entries where each entry provides comprehensive detail about a hapten molecule that include

- URL: http://www.imtech.res.in/ragahva/haptendb/

Thanks