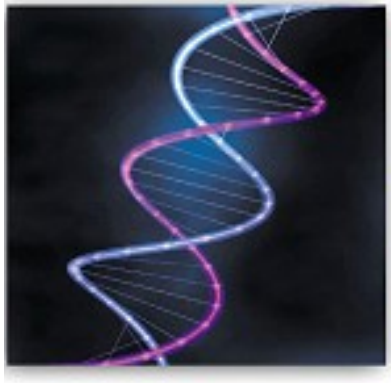
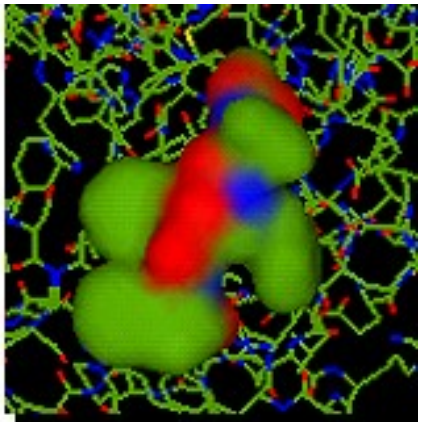
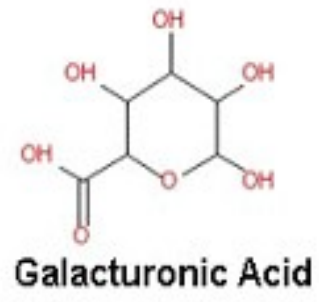
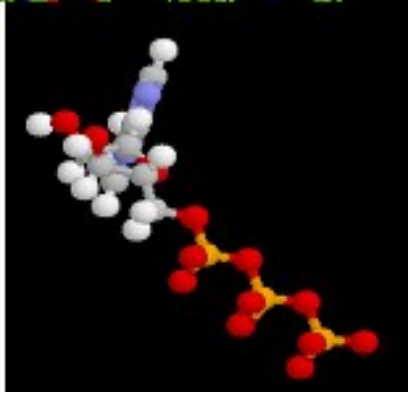
BIOINFORMATICS…….*AT A GLANCE*







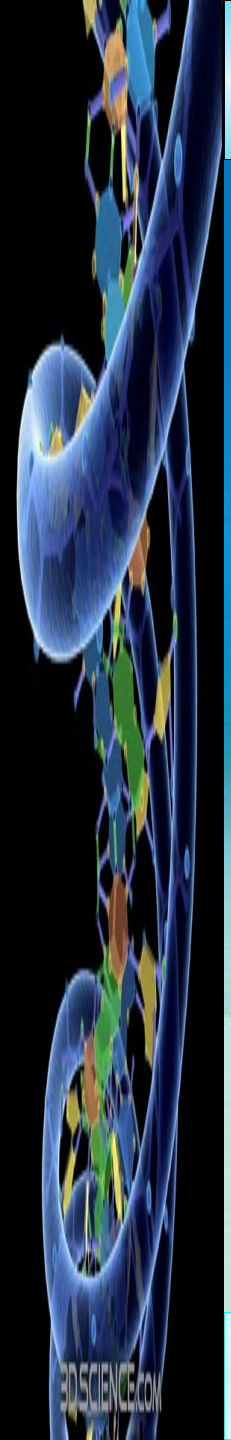
Galacturonic Acid



By :-Mr. Arvind Singh

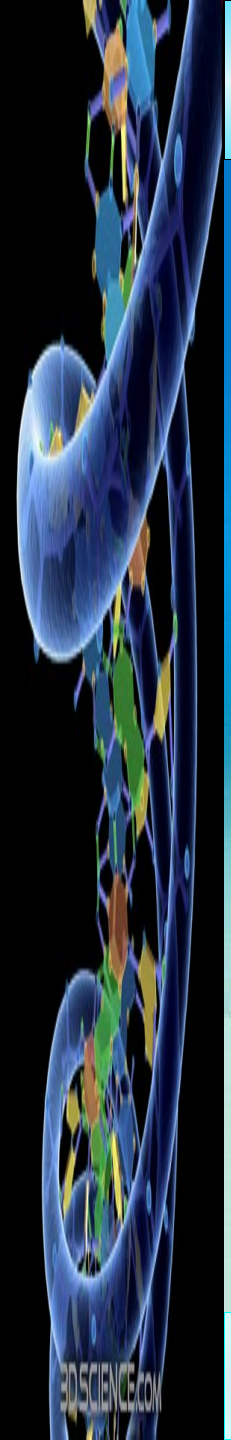
M. Sc Bioinformatics

Faculty BII



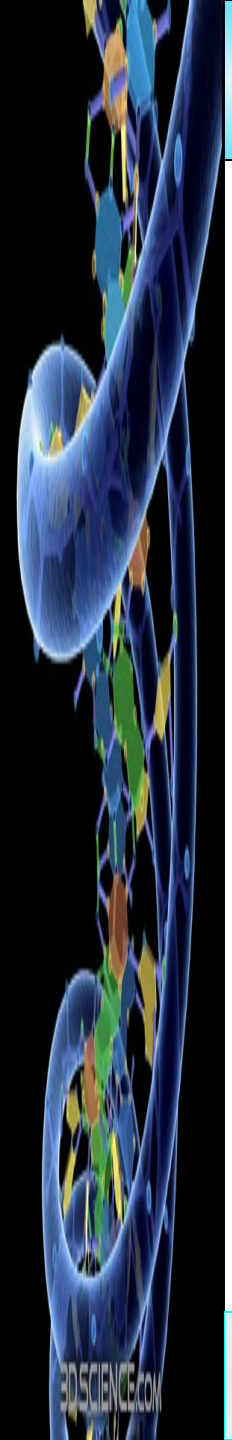
- Part I-Introduction to Bioinformatics
- Part II-Historical Overview of Bioinformatics
- Part III-Human Genome Project
- Part IV-Biological Databases
- Part V-Internet and Bioinformatics
- Part VI-Knowledge Discovery and Data mining
- Part VII-Career Prospect In Bioinformatics

BIOINFORMATICS INSTITUTE OF INDIA



Part I-Introduction to Bioinformatics

BIOINFORMATICS INSTITUTE OF INDIA

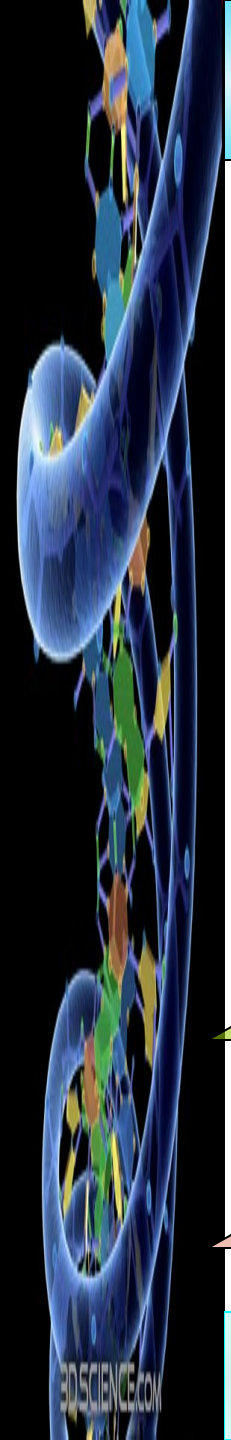


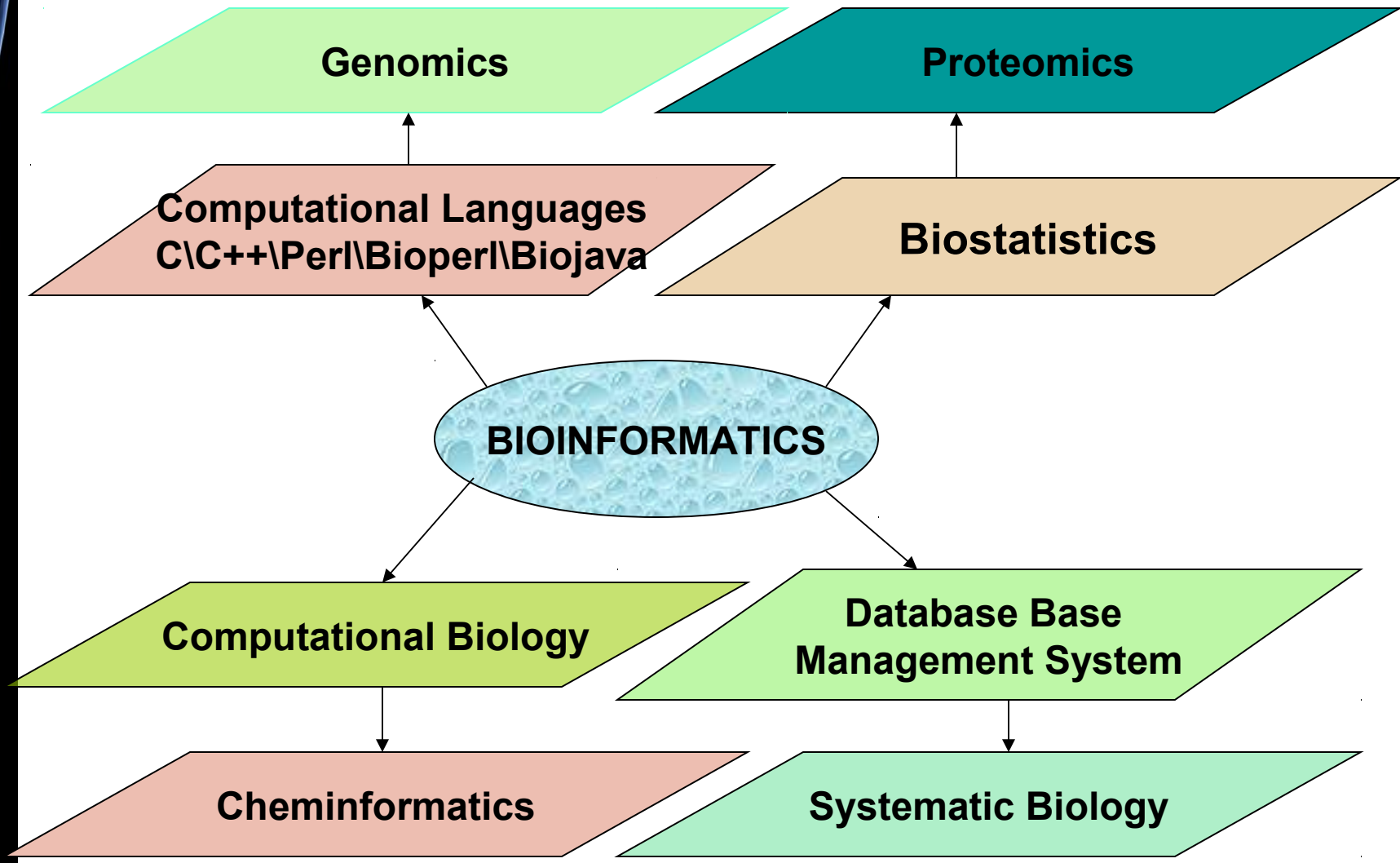
Definition of Bioinformatics

General Definition: A computational approach ,Solves the biological problem.

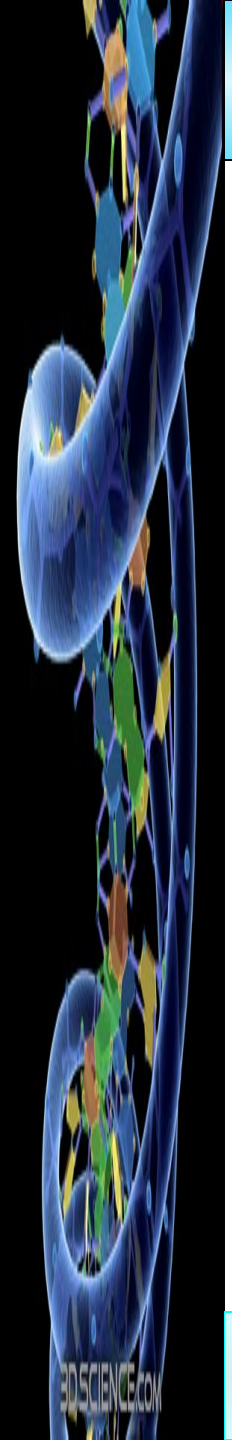
- Bioinformatics is emerging and advance branch of biological science , contain Biology mathematics and Computer Science.
- Bioinformatics developed a new thought , to maintain the concepts and store .The huge amount of Biological data.
- Bioinformatics concepts and Method are different than the Biological concepts and method.
- Bioinformatics, A logical and technical means by which not only solve the Biological problems but also can predicts the new aspects.

Bioinformatics Areas





BIOINFORMATICS INSTITUTE OF INDIA

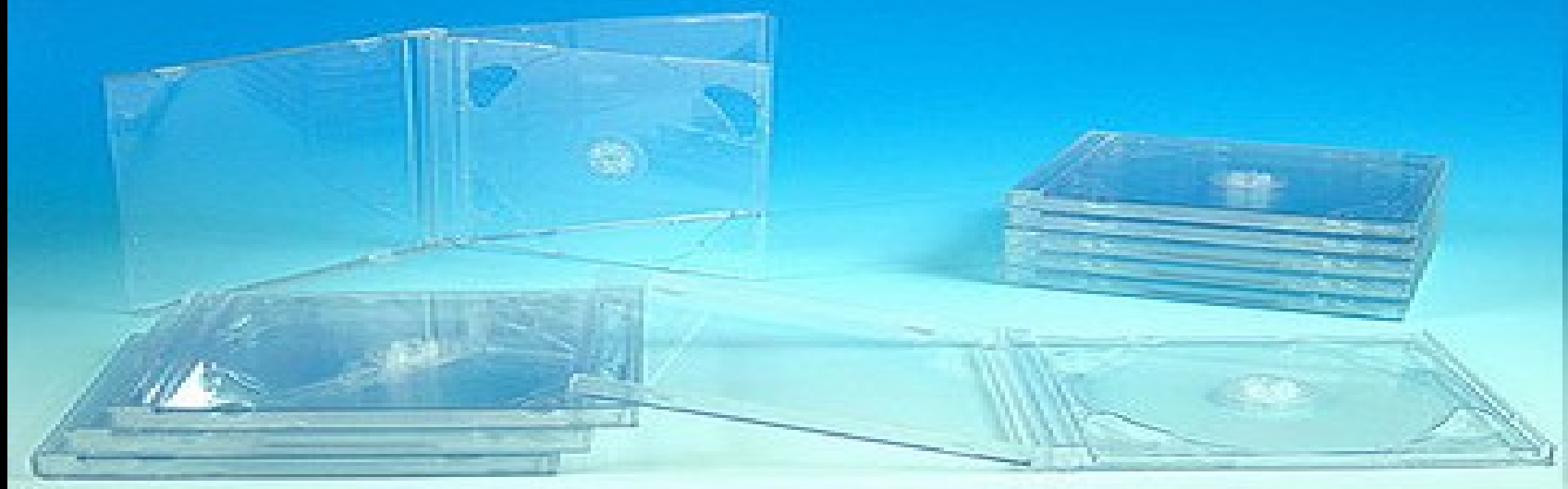


Insilico Areas of Bioinformatics

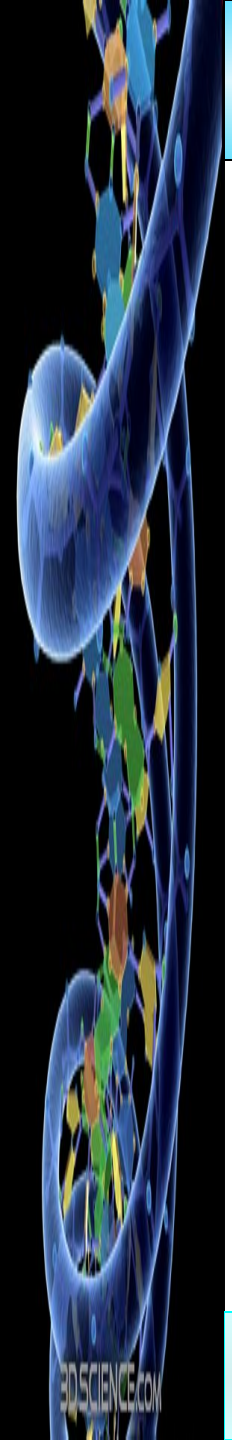
Computational Biology

- Comparative Homology Modeling
- Phylogenetic Analysis
- Protein structure prediction
- Protein Folding Problem
- Micro array analysis
- Docking Approaches& New Drug Discovery

Part II-Historical Overview of Bioinformatics

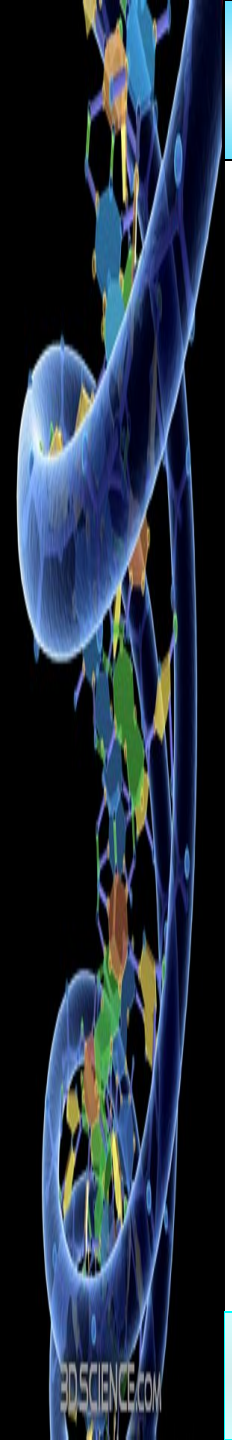


BIOINFORMATICS INSTITUTE OF INDIA



HISTORY AND SCOPE OF BIOINFORMATICS

- 1859 – The “On the Origin of Species”, published by Charles Darwin that introduced theory of genetic evolution – allows adaptation over time to produce organisms best suited to the environment.
- 1869 - The DNA from nuclei of white blood cells was first isolated by Friedrich Meischer.
- 1951 – Linus Pauling and Corey propose the structure for the alpha-helix and beta-sheet.
- 1953 - Watson and Crick propose the double helix model for DNA based on x-ray data obtained by Franklin and Wilkins.
- 1955 - The sequence of the first protein to be analyzed, bovine insulin, is announced by F. Sanger.
- 1958 - The Advanced Research Projects Agency (ARPA) is formed in the US.
-



HISTORY AND SCOPE OF BIOINFORMATICS

• 1973 - The Brookhaven Protein Data Bank(PDB) is announced.

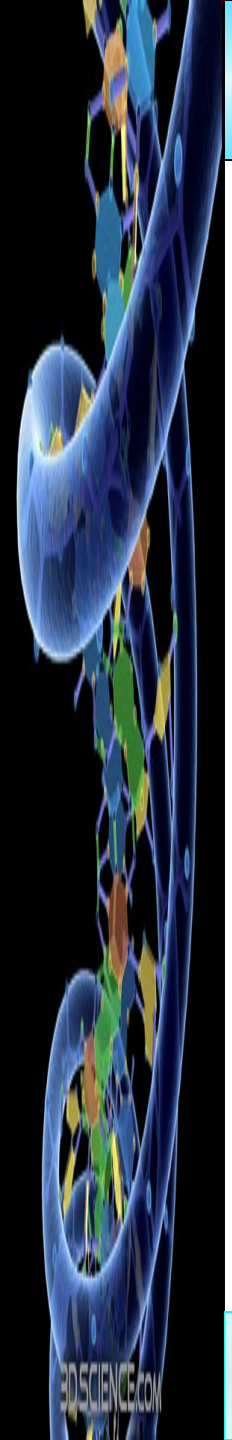
• 1987 - Perl (Practical Extraction Report Language) is released by Larry Wall.

•10. 1988 - National Centre for Biotechnology Information (NCBI) founded at NIH/NLM.

•11. 1990 - Human Genome Project launched
•BLAST program introduced by S. Karlin and S.F. Altshul.
• Tim Berners-Lee, a British scientist invented the World Wide Web in 1990.

•12. 1992 - The Institute for Genome Research (TIGR), associated with plans to exploit sequencing
•commercially through gene identification and drug discovery, was formed.

•13. 2001 - The human genome (3,000 Mbp) is published.

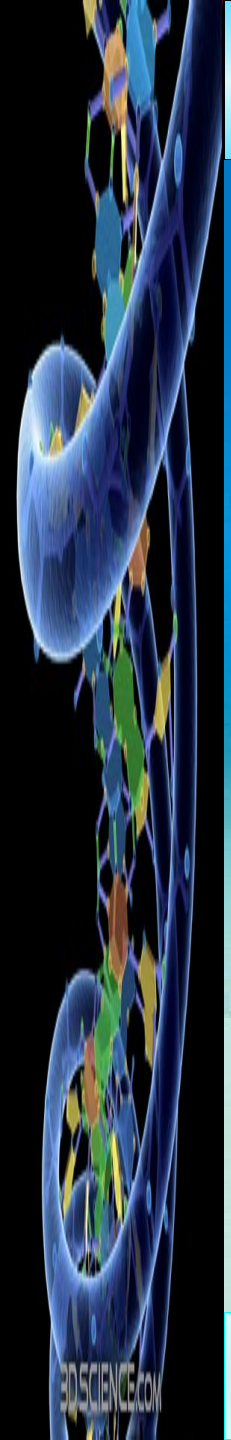


Future Goals Of Molecular Biology and Bioinformatics Research

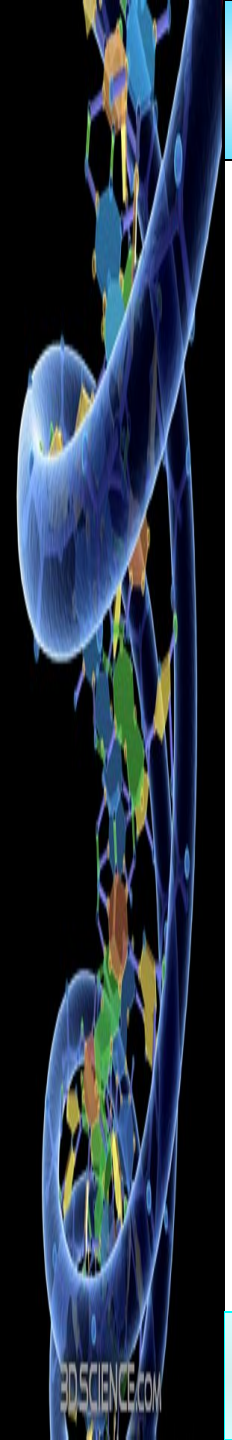
2010 :Completion of the 2010 Project: to understand the function of all genes within their cellular, organism and evolutionary context of Arabidopsis thaliana.

2050: To complete of the first computational model of a complete cell, or maybe even already of a complete organism.

Part III-Human Genome Project



BIOINFORMATICS INSTITUTE OF INDIA



Human Genome Project

U.S. govt. project coordinated by the Department of Energy and the National Institutes of Health, launched in 1986 by Charles DeLisi.

Definition: GENOME – the whole hereditary information of an organism that is encoded in the DNA.

Aims of the project:

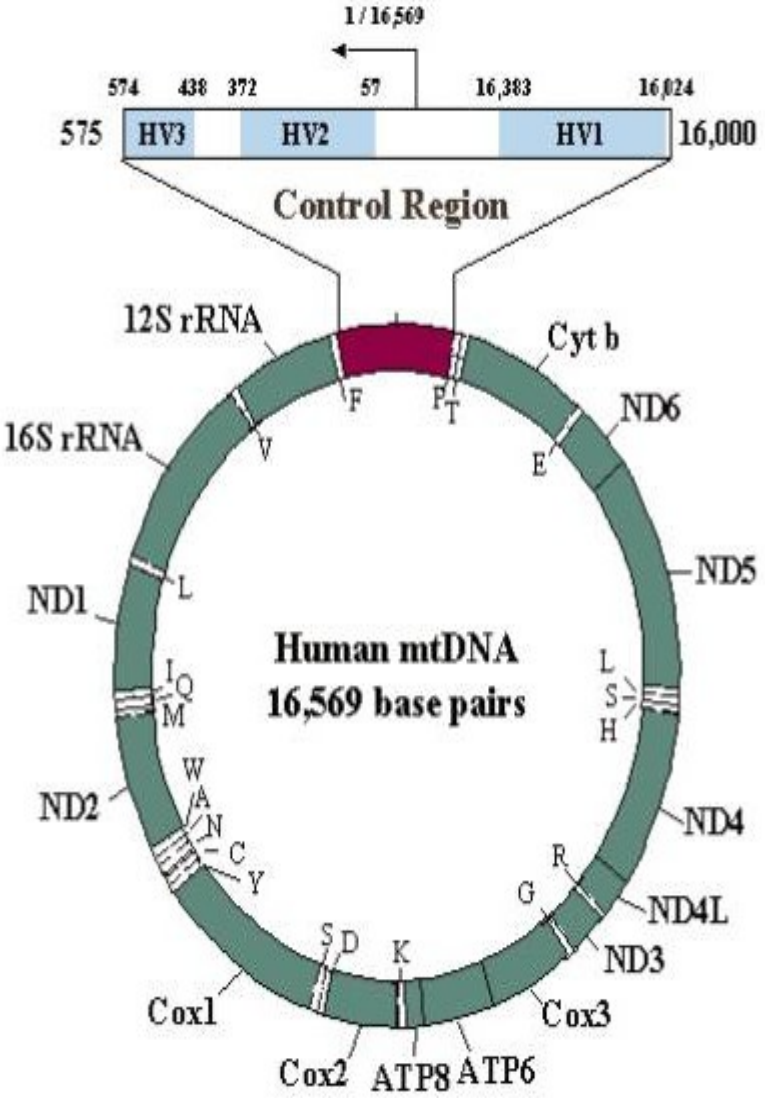
- To identify the approximate 100,000 genes in the human DNA.
- Determine the sequences of the 3 billion bases that make up human DNA.
- Store this information in databases.
- Develop tools for data analysis.
- Address the ethical, legal, and social issues that arise from genome research.

Whose genome is being sequenced?

- A group of researchers have managed to complete a genetic map of the bacterium *Haemophilus influenzae*

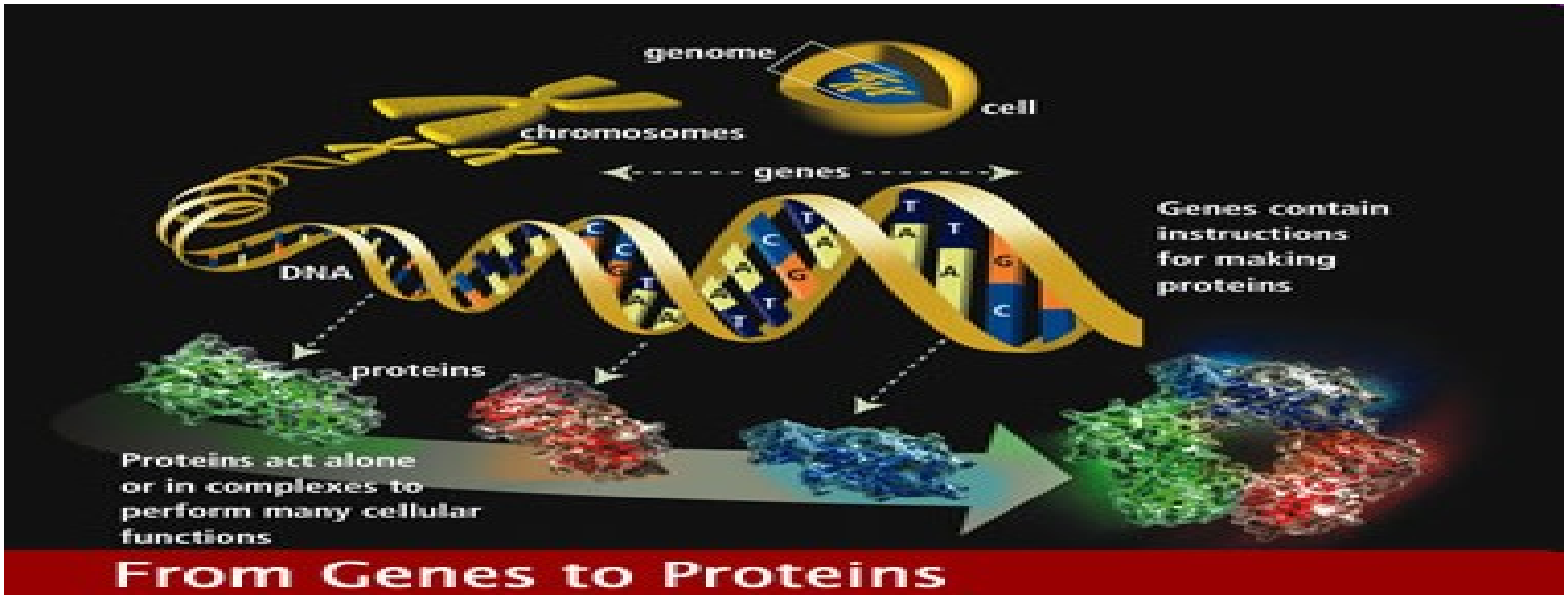
- The approach called whole Genome Shotgun Sequencing to sequence the 1,749 genes of the bacterium in minimum time period.

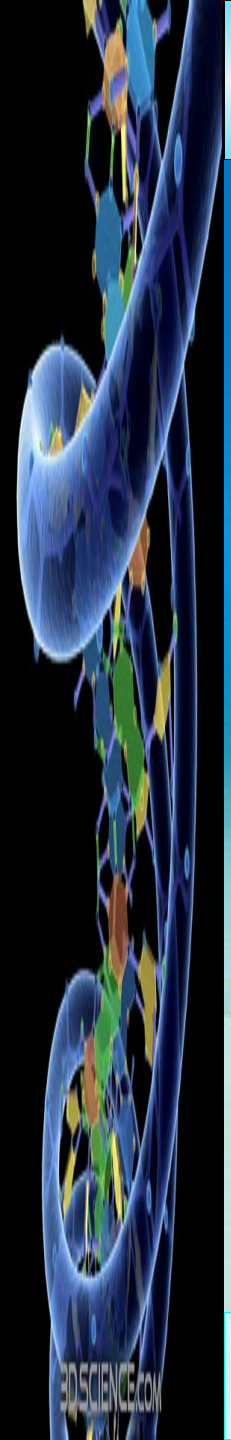
- The H. influenzae project was based on an approach to genomic analysis using sequencing and assembly of unselected pieces of DNA from the whole chromosome



Benefits of Human Genome Project research

- Improvements in medicine and Drugs used in genetic or metabolic disorder.
- Microbial genome research for Bio-fuel and environmental cleanup.
- DNA finger printing & forensics.
- Improved agriculture by improving the wild gene of high yielding variety of grain and livestock.
- Better understanding of species evolution and Human genome .
- More accurate risk assessment by gene mapping.





Part IV-Internet and Bioinformatics

BIOINFORMATICS INSTITUTE OF INDIA

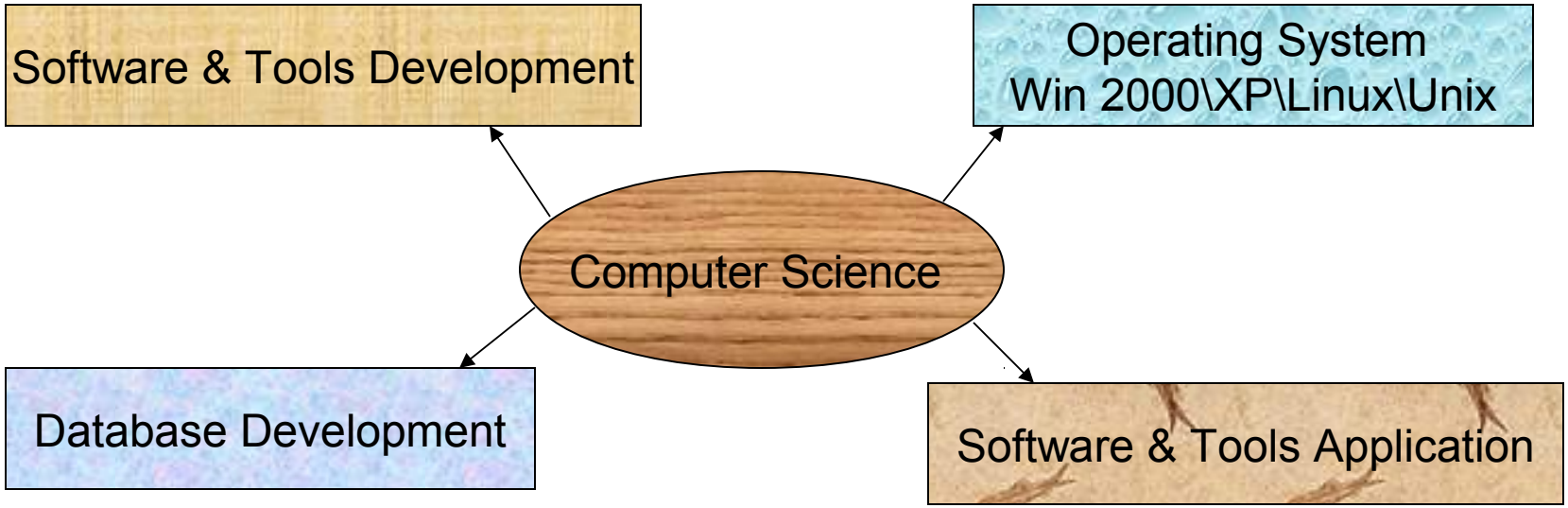
Internet and Bioinformatics

✓Internet plays an important role to retrieve the biological information.

✓Bioinformatics emerging new dimension of Biological science, include The computer science ,mathematics and life science.

✓The Computational part of bioinformatics use to optimize the biological problems like (metabolic disorder, genetic disorders).

✓Computational part contains:



Internet and Bioinformatics

The Mathematical portion helps to understand the algorithms used in Bioinformatics software and tools.

The mathematical portion which, used in Bioinformatics are :

Matrices
(Sequence alignment, Blast Fast, MSA & Phylogenetic Prediction)

Biostatistics
(HMM, ANN in secondary structure prediction)

Mathematics

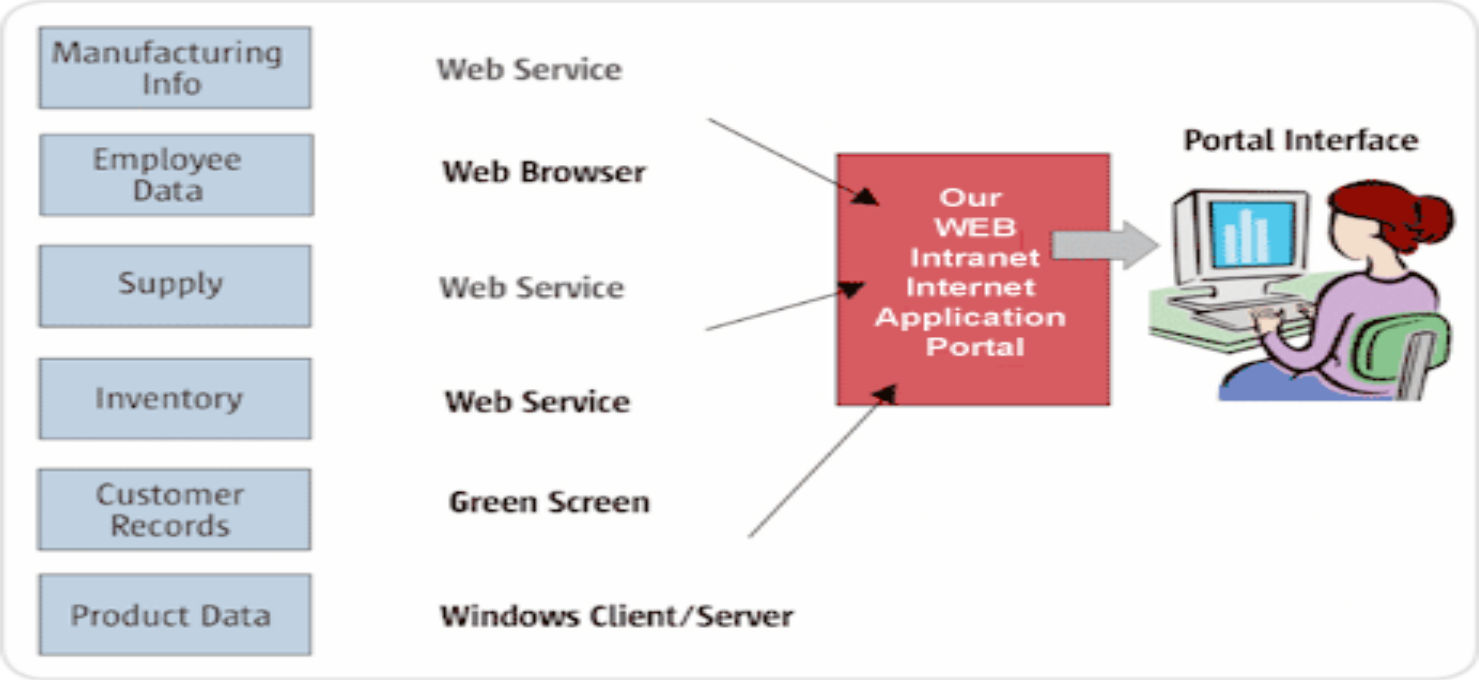
Complex Mathematics Functions
(Fourier Transformation)

Diffrentiation\Intigration
(Time and space complexity E-value ,p-values in Blast)

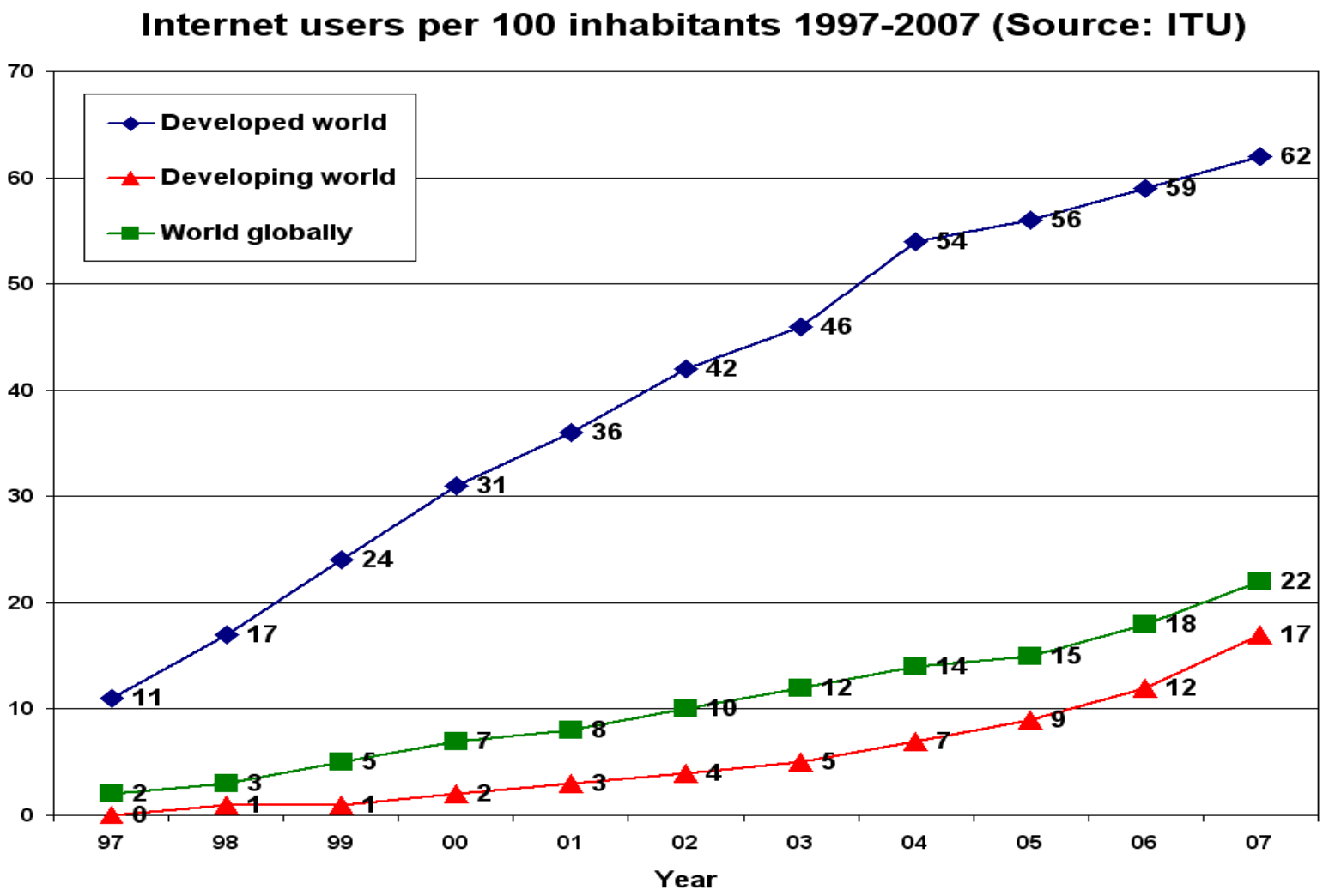
Internet and Bioinformatics

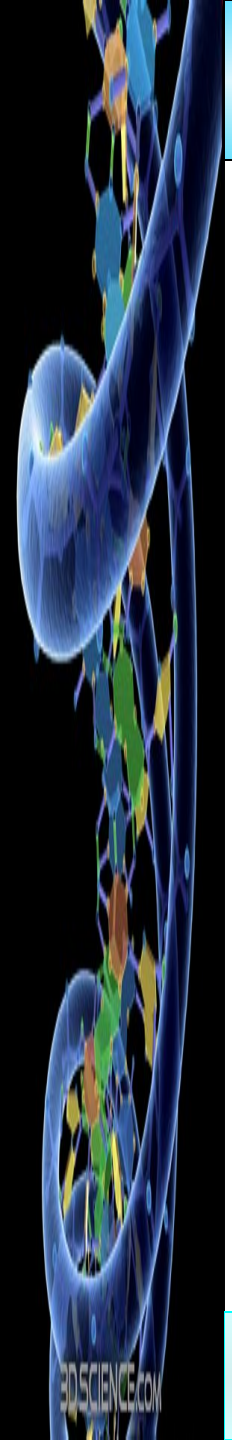
The Internet is a global data communications system. It is a hardware and software infrastructure that provides connectivity between computers.

The Web is one of the services communicated via the Internet. It is a collection of interconnected documents and other resources, linked by hyperlink and URLs.



Graph of internet users per 100 inhabitants between 1997 and 2007





Internet Resources for Bioinformatics

Database Search Engine
NCBI\Swiss-Prot\ Uni-Prot-K

Online Databases
KEGG\ ModBase\PDB\Zinc Database\MolSoft

Online serve tools
Clustal W\Swiss-Model Serve\

Biological Databases
Primary\Secondary\Composite

Relational Database Management System

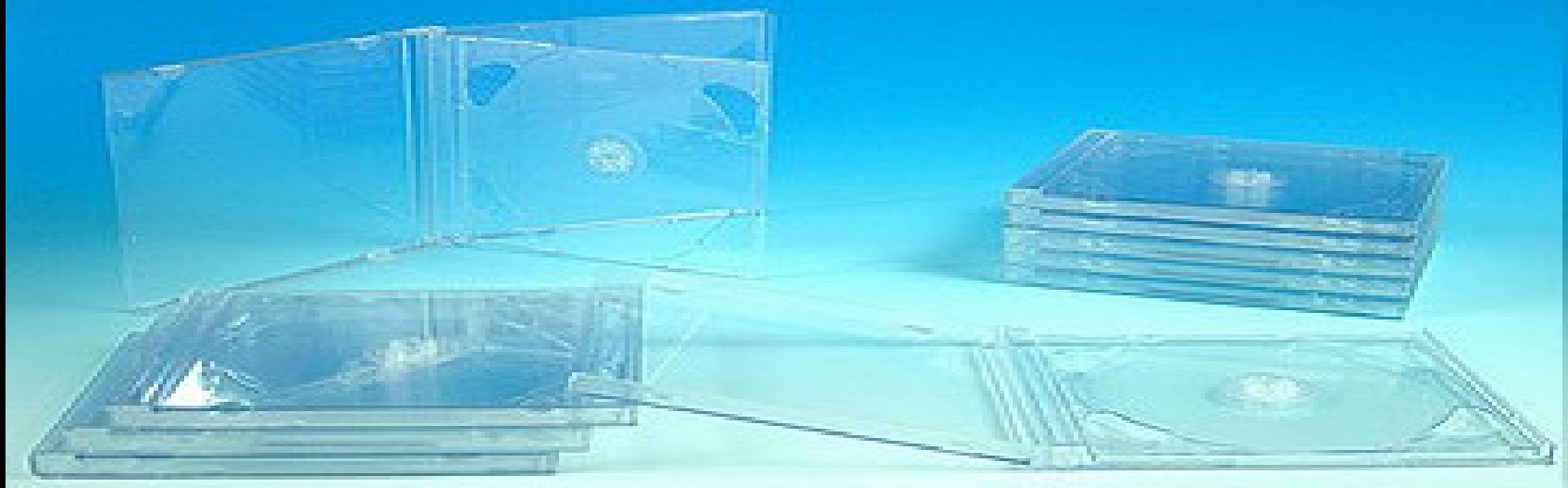
Bioinformatics Software's Application

Other Software & Information Sources

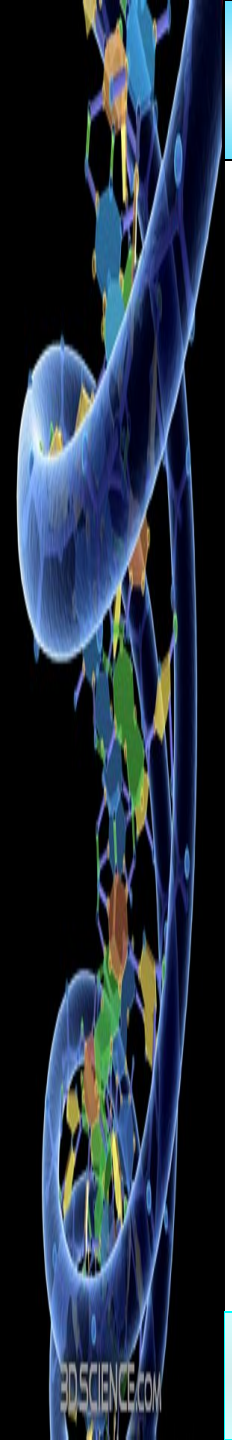
Bioinformatics Software's Development

BIOINFORMATICS INSTITUTE OF INDIA

Part V- Biological Databases



BIOINFORMATICS INSTITUTE OF INDIA



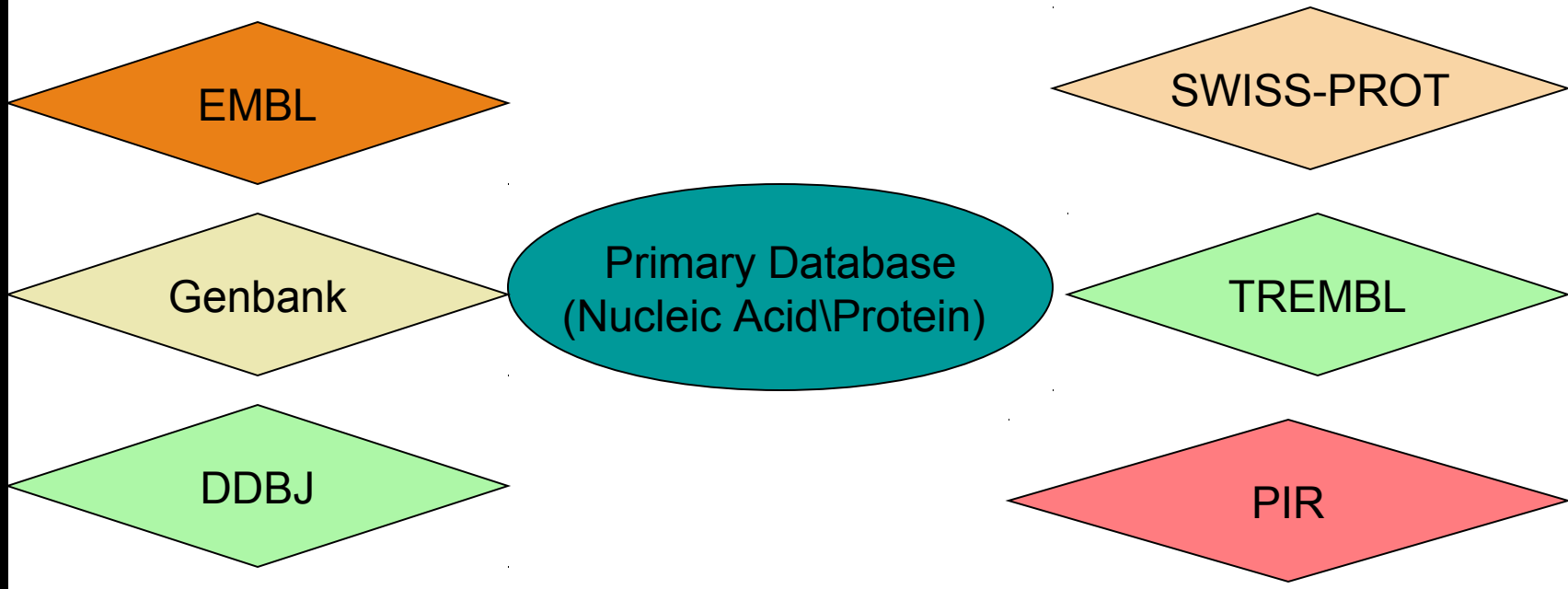
Biological Databases

Type of databases	Information Contain
Bibliographic databases	Literature
Taxonomic databases	Classification
Nucleic acid databases	DNA information
Genomic databases	Gene level information
Protein databases	Protein information
Protein families, domains and functional sites	Classification of proteins and identifying domains
Enzymes/ metabolic pathways	Metabolic pathways

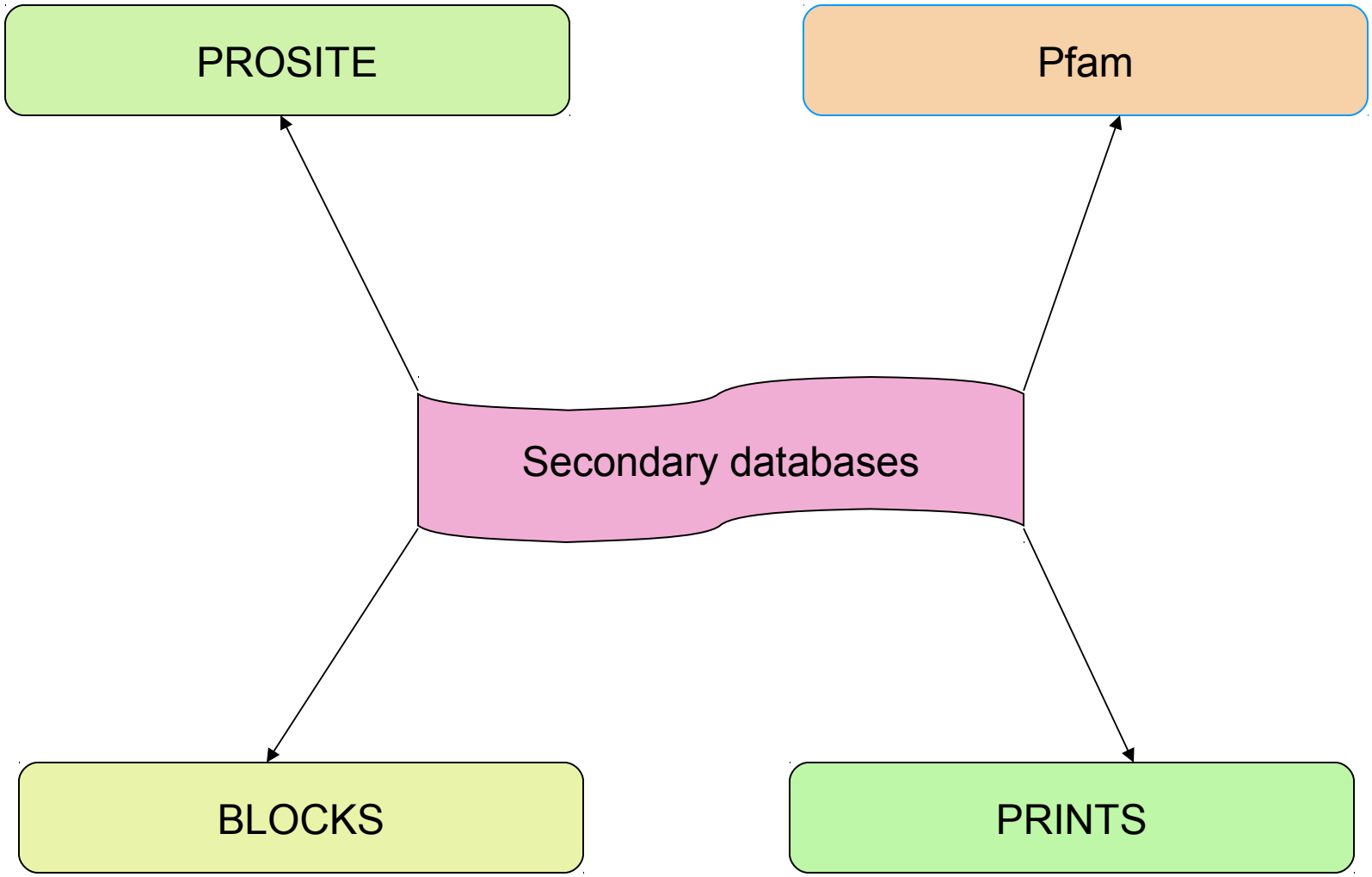
Types Of Biological Databases Accessible

There are many different types of database but for routine sequence analysis, the following are initially the most important.

- Primary databases
- Secondary databases
- Composite databases

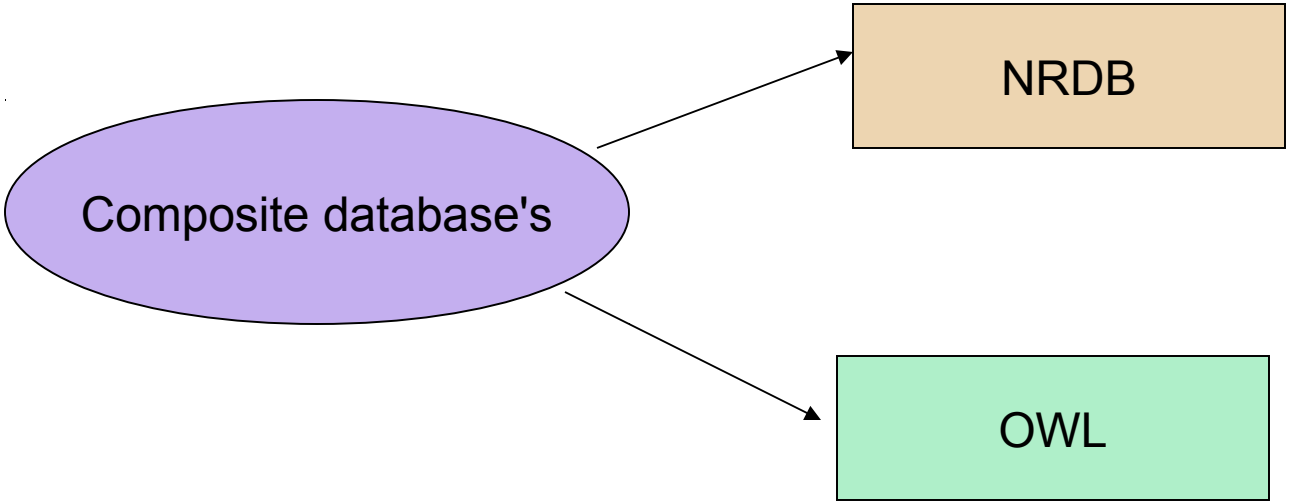


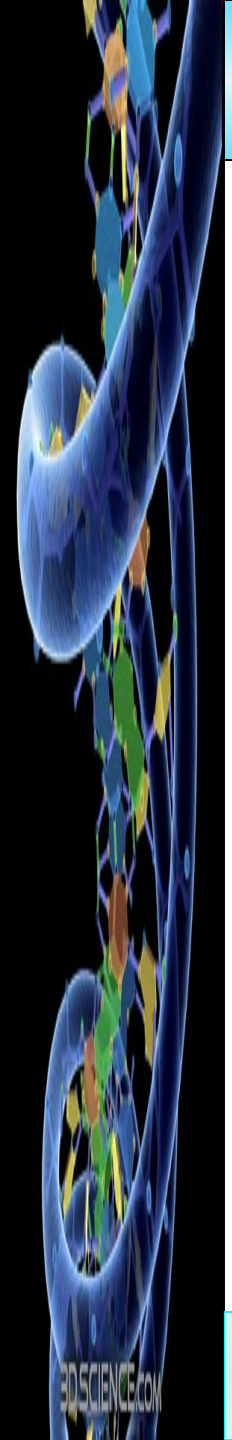
Secondary databases



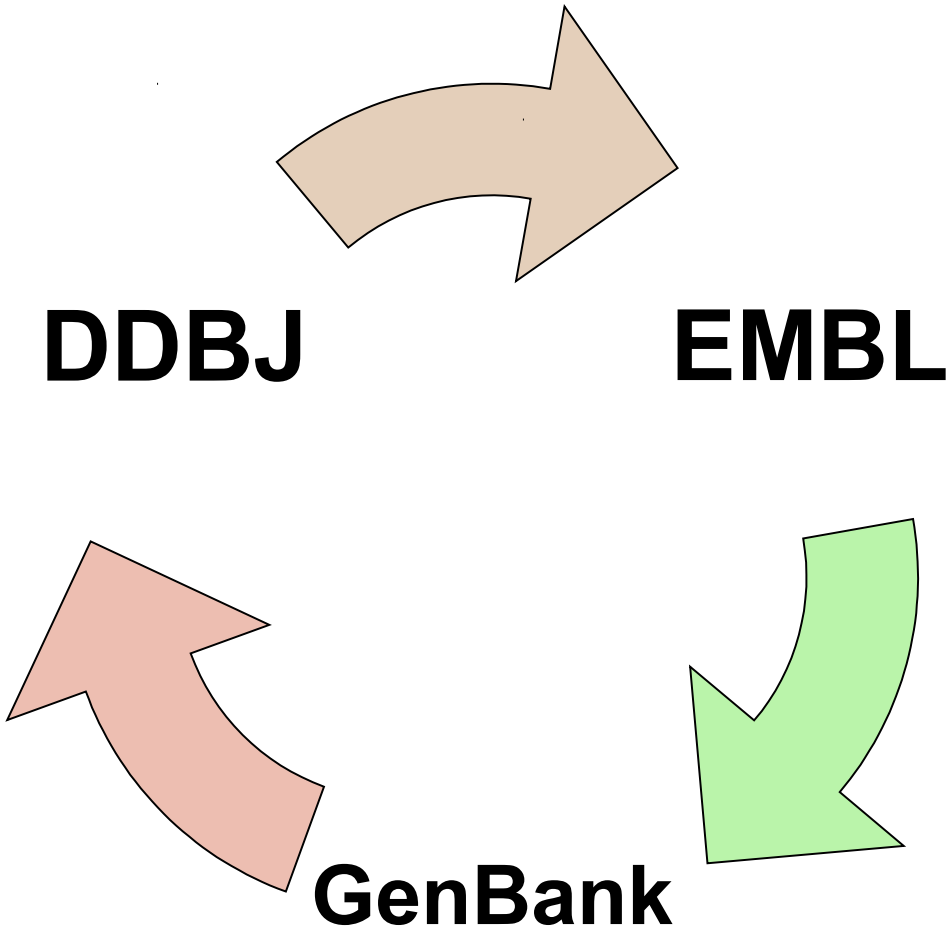
Composite databases

- ✓Combine different sources of primary databases.

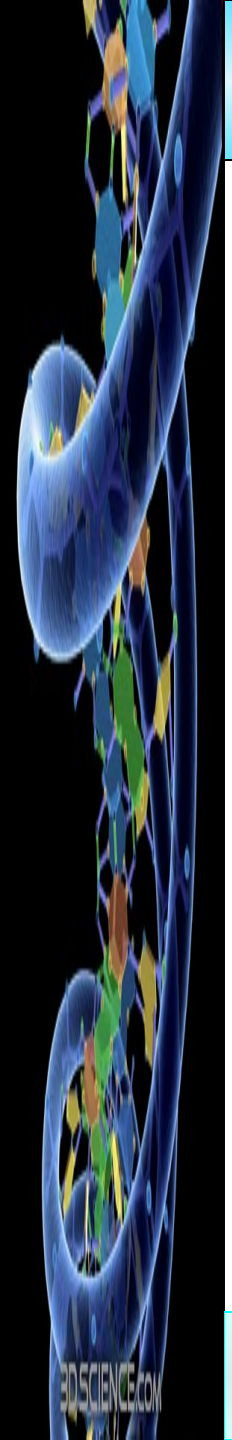




The International Sequence Database Collaboration



BIOINFORMATICS INSTITUTE OF INDIA



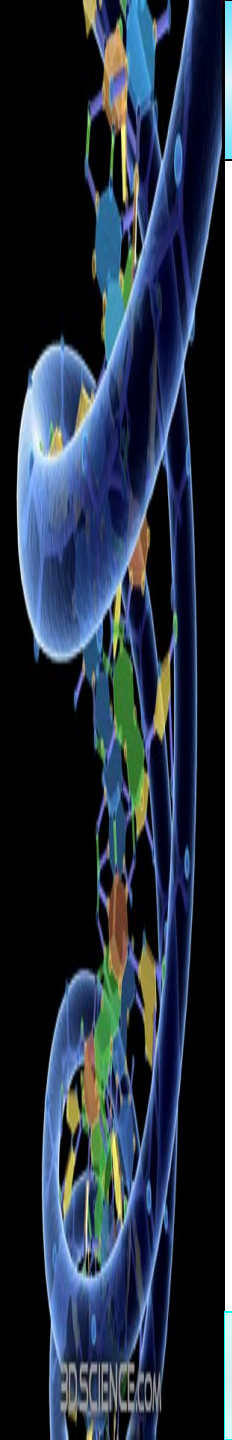
GenBank

GenBank® is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences

DDBJ

DDBJ (DNA Data Bank of Japan) began DNA data bank activities in earnest in 1986 at the National Institute of Genetics (NIG) with the endorsement of the Ministry of Education, Science, Sport and Culture.

The Center for Information Biology at NIG was reorganized as the Center for Information Biology and DNA Data Bank of Japan (CIB-DDBJ) in 2001. The new center is to play a major role in carrying out research in information biology and to run DDBJ operation in the world.

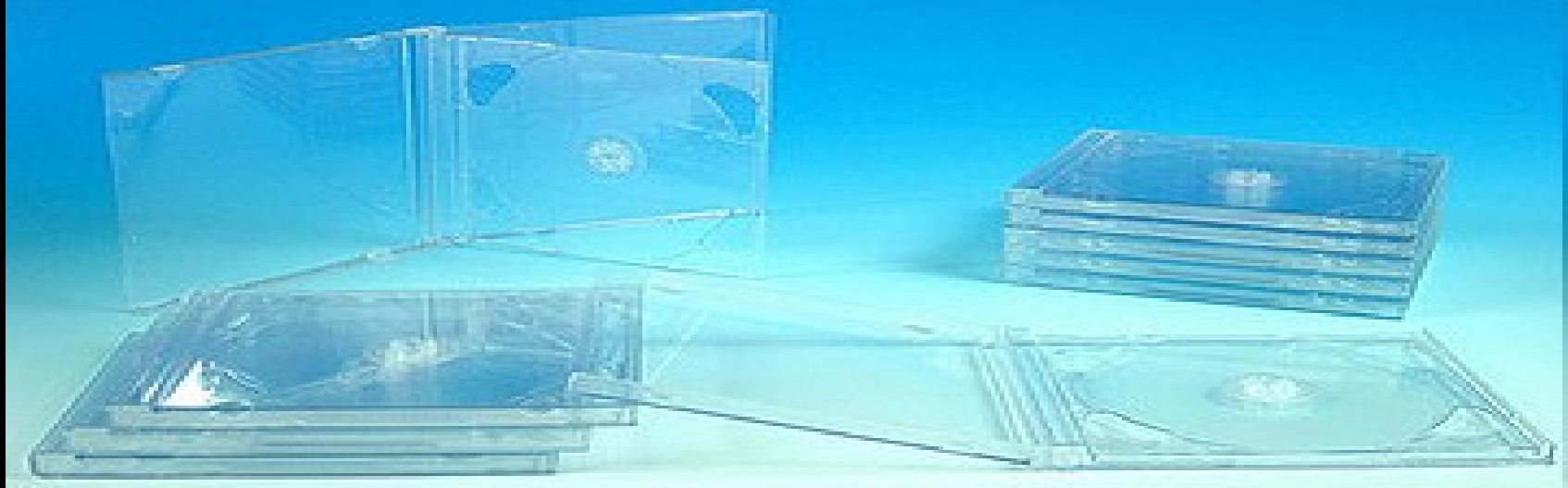


EMBL Nucleotide Sequence Database

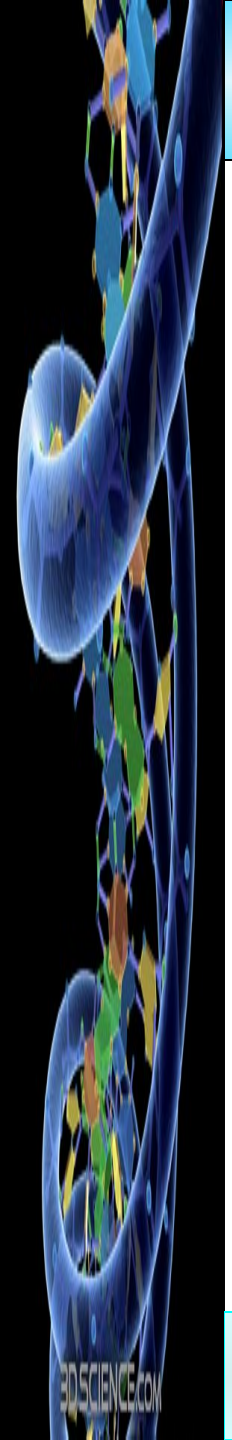
The EMBL Nucleotide Sequence Database (also known as EMBL-Bank) constitutes Europe's primary nucleotide sequence resource. Main sources for DNA and RNA sequences are direct submissions from individual researchers, genome sequencing projects and patent applications.

The database is produced in an international collaboration with GenBank (USA) and the DNA Database of Japan (DDBJ).

Part VI-Knowledge Discovery and Data mining



BIOINFORMATICS INSTITUTE OF INDIA



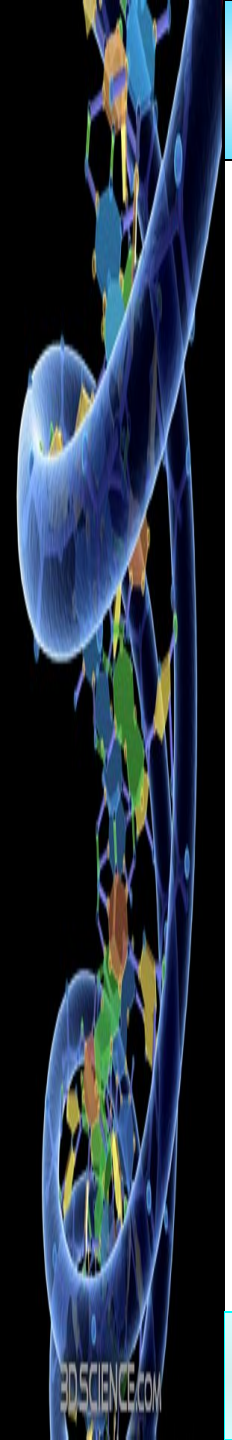
Why Data Mining ?

Biology: Language and Goals

- A gene can be defined as a region of DNA.
- A genome is one haploid set of chromosomes with the genes they contain.
- Perform competent comparison of gene sequences across species and account for inherently noisy biological sequences due to random variability amplified by evolution
- Assumption: if a gene has high similarity to another gene then they perform the same function

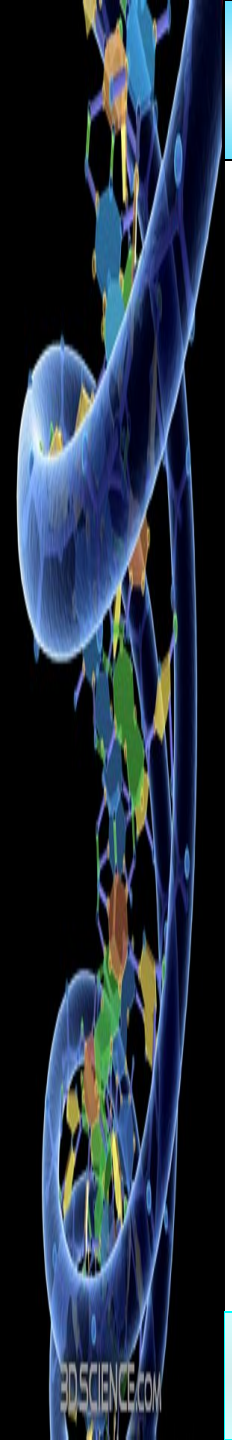
Analysis: Language and Goals

- Feature is an extractable attribute or measurement (e.g., gene expression, location)
- Pattern recognition is trying to characterize data pattern (e.g., similar gene expressions, equidistant gene locations).
- Data mining is about uncovering patterns, anomalies and statistically significant structures in data (e.g., find two similar gene expressions with confidence > x)



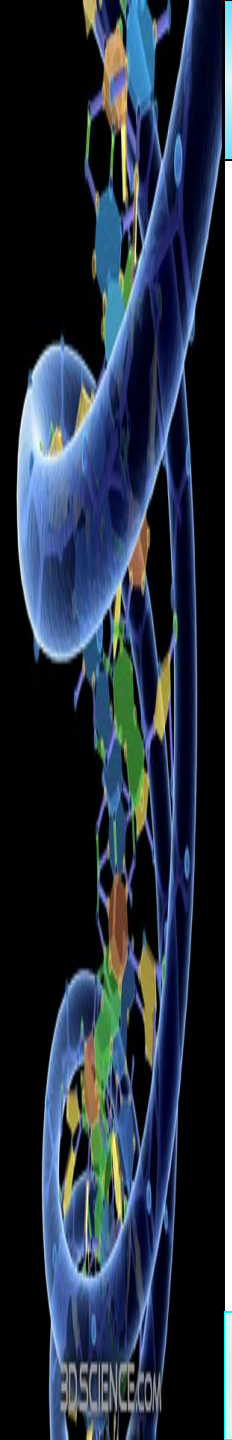
What is Data Mining

- Data mining (knowledge discovery from data)
- Extraction of interesting (non-trivial, implicit, previously unknown and potentially useful) patterns or knowledge from huge amount of data
- Data mining: a misnomer?
- Alternative names
- Knowledge discovery (mining) in databases (KDD), knowledge extraction, data/pattern analysis, data archeology, data dredging, information harvesting, business intelligence, etc.
- Watch out: Is everything "data mining"?
  - Simple search and query processing (Deductive) expert systems



Evolution of Database Technology

- 1960s:
  - Data collection, database creation, IMS and network DBMS
- 1970s:
  - Relational data model, relational DBMS implementation
- 1980s:
  - RDBMS, advanced data models (extended-relational, OO, deductive, etc.)
  - Application-oriented DBMS (spatial, scientific, engineering, etc.)
- 1990s:
  - Data mining, data warehousing, multimedia databases, and Web databases
- 2000s
  - Stream data management and mining
  - Data mining and its applications
  - Web technology (XML, data integration) and global information systems



Why Data Mining?—Potential Applications

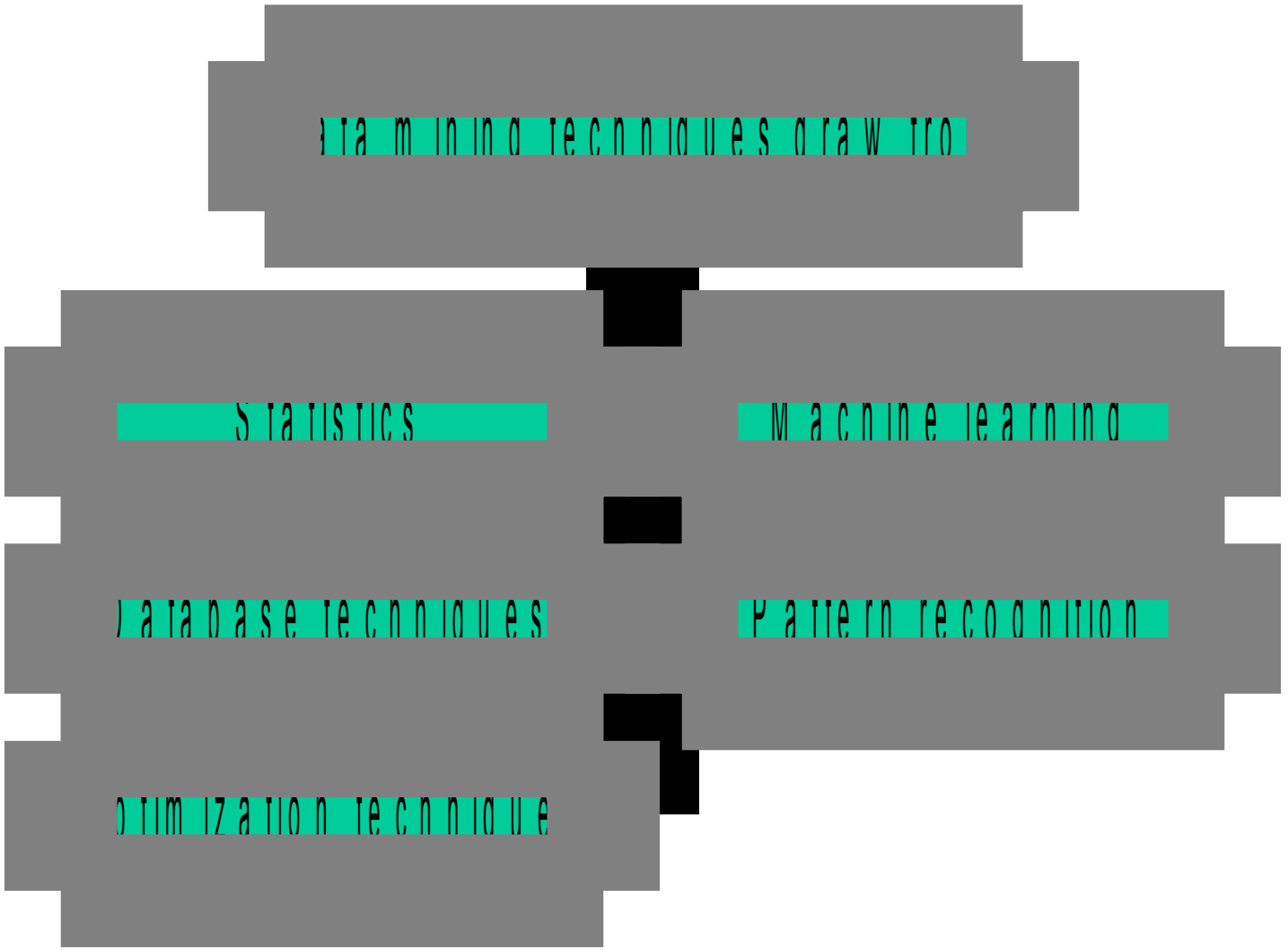
Data analysis and decision support

- Market analysis and management
  - Target marketing, customer relationship management (CRM),
  - market basket analysis, cross selling, market segmentation
- Risk analysis and management
  - Forecasting, customer retention, improved underwriting,
  - quality control, competitive analysis
- Fraud detection and detection of unusual patterns (outliers)

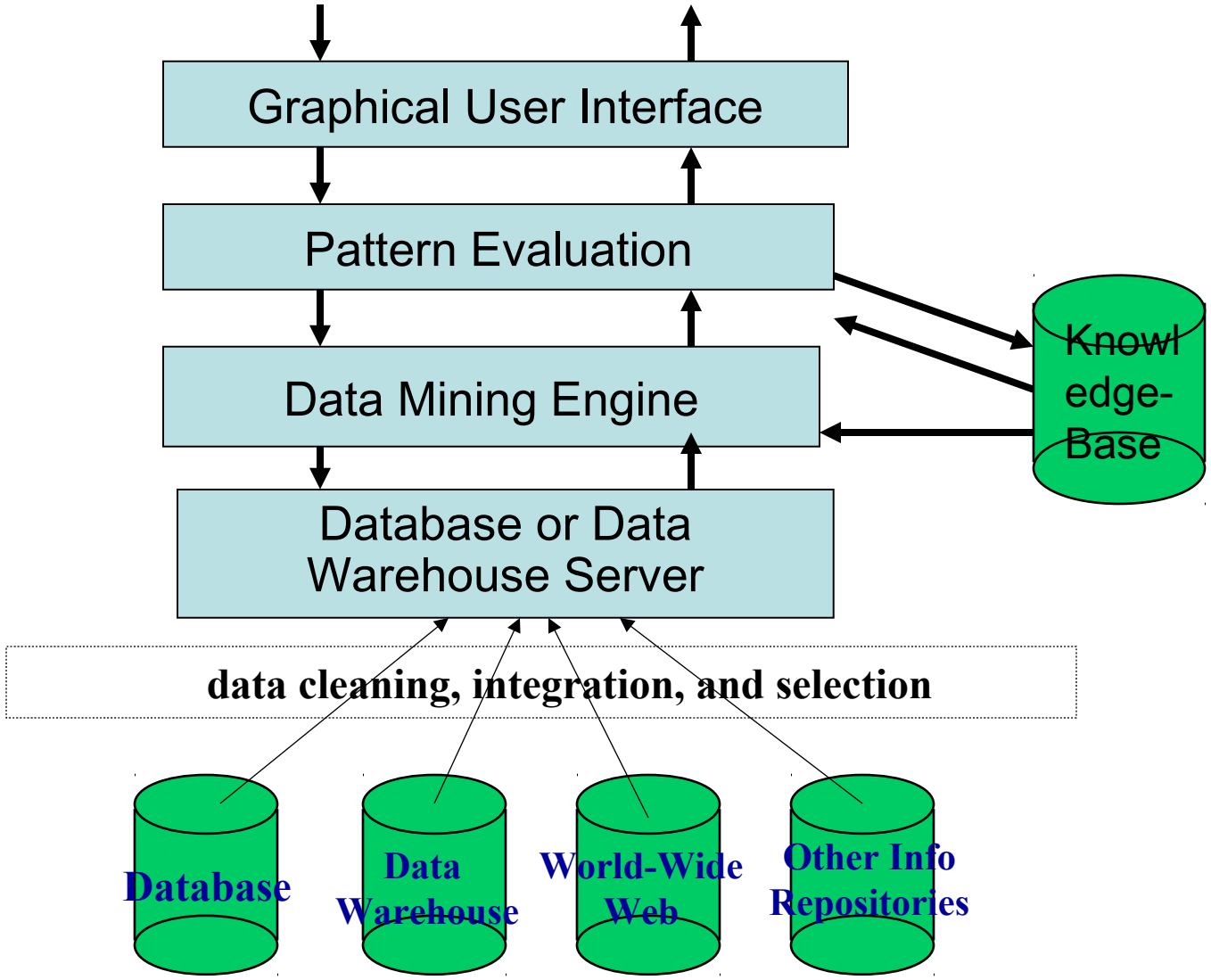
Other Applications

- Text mining (news group, email, documents) and Web mining
- Stream data mining
- Bioinformatics and bio-data analysis

Data Mining Techniques

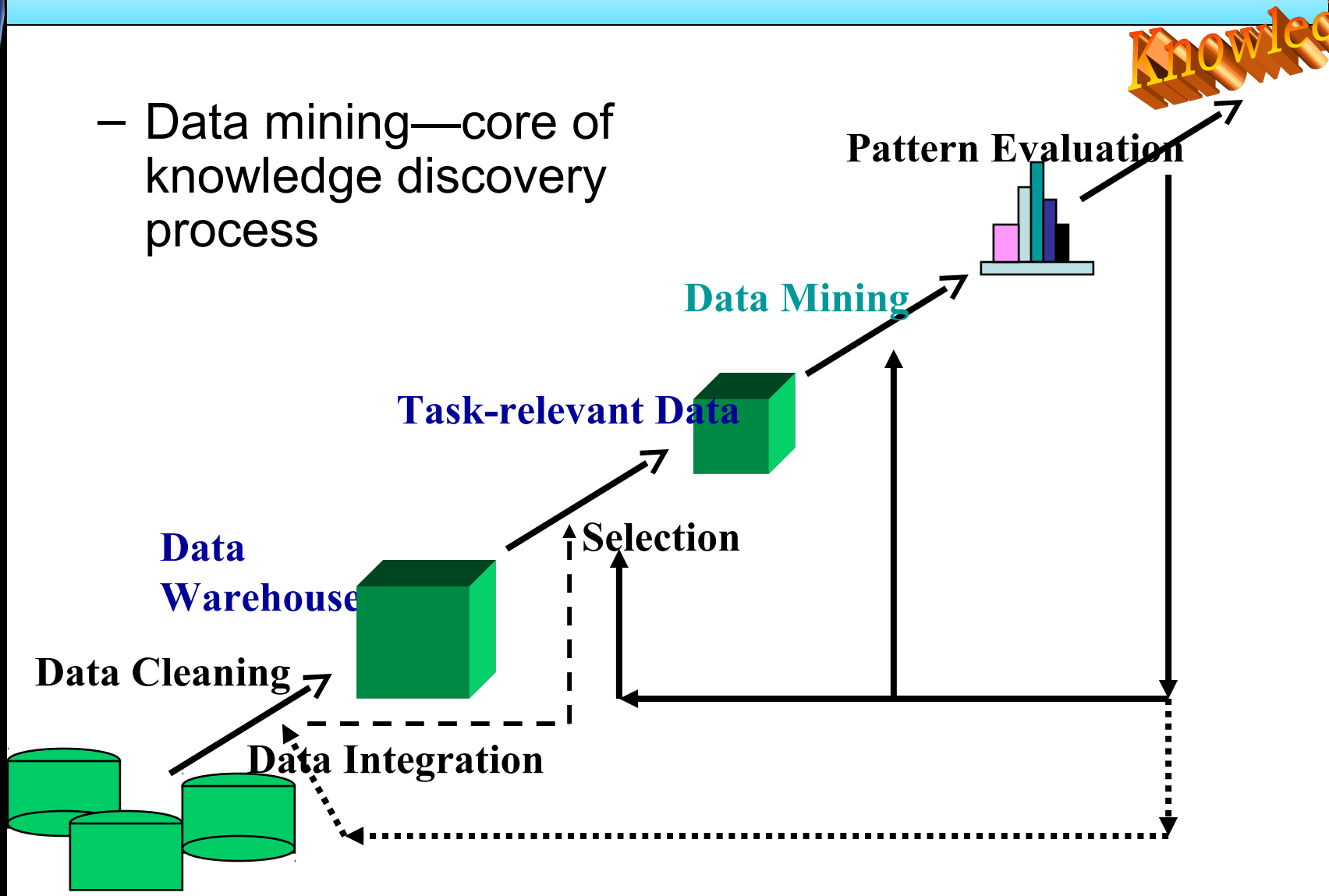


Architecture: Typical Data Mining System

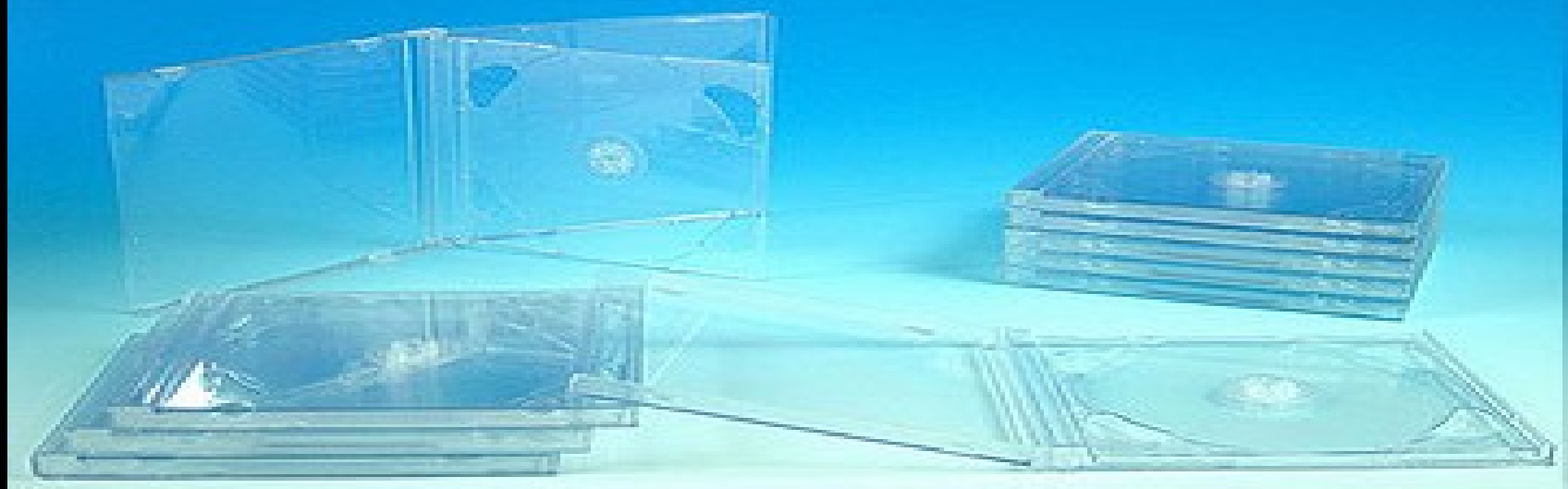


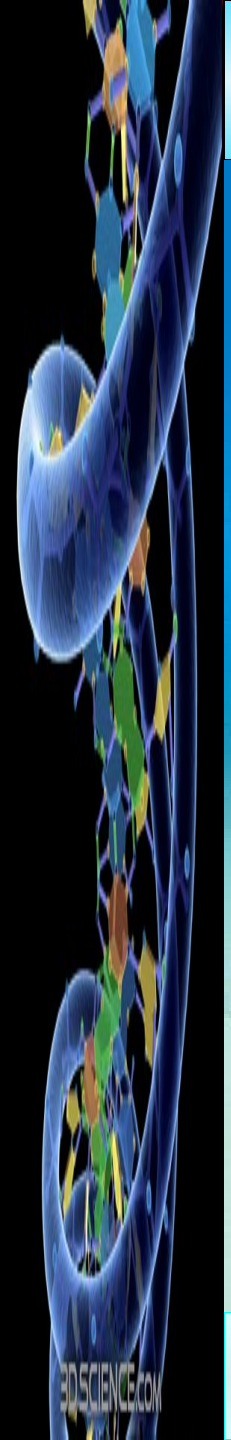
Knowledge Discovery (KDD) Process

– Data mining—core of knowledge discovery process

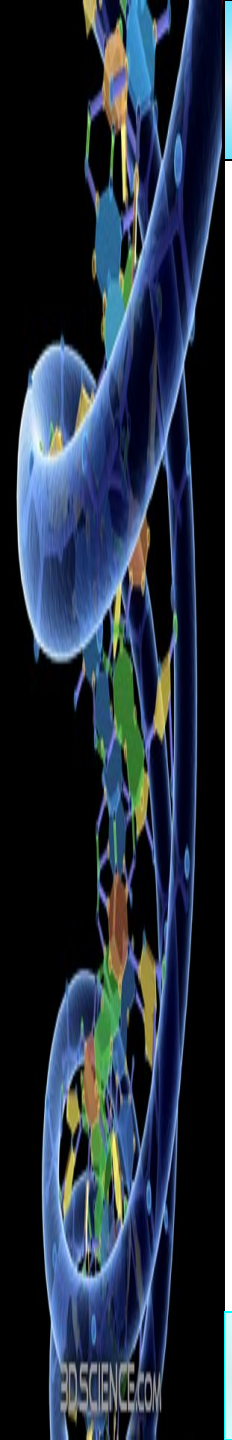


Part VII-Career Prospect In Bioinformatics





BIOINFORMATICS INSTITUTE OF INDIA



Develop your carrier ……as

- SCIENTIST , RESEARCH ASSOCIATE
- PROF. READER & LECT. IN UNIV \COLLEGE
- TECHNICAL EXCUTIVE IN BUSSINESS & DEV
- DATA ANALYST IN SCIENTIFIC ORG & LAB
- BIOSOFT DEVELOPER IN IT INDUSTRY

BIOINFORMATICS INSTITUTE OF INDIA