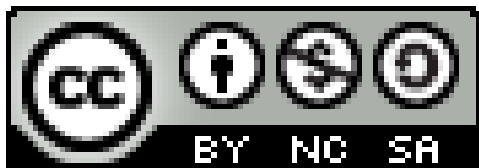


Basics to Bioinformatics

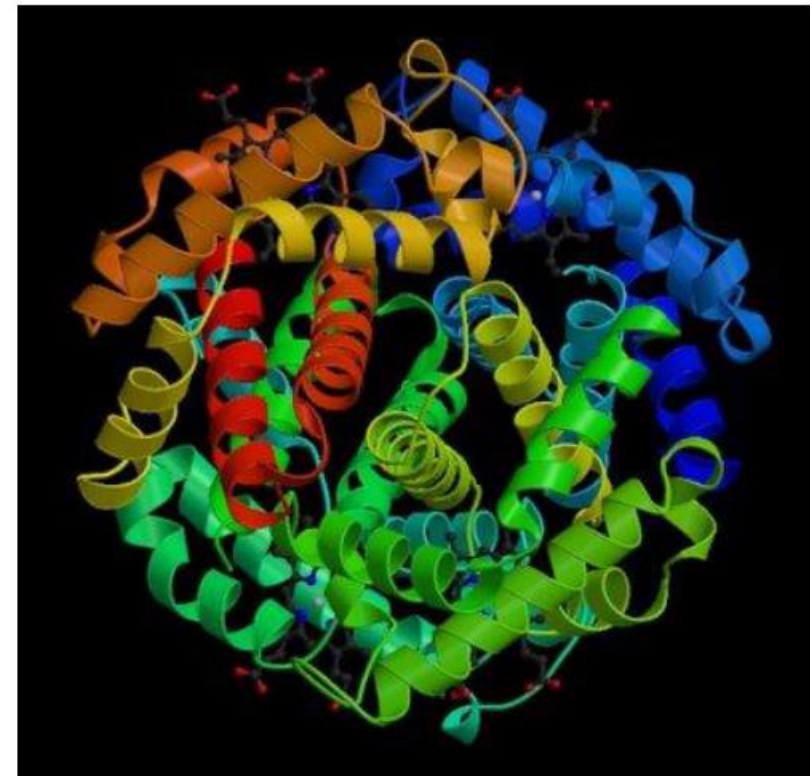


Dr. Kuldeep Sharma
Assistant Professor
Department of Botany
MLS University, Udaipur
Rajasthan, India

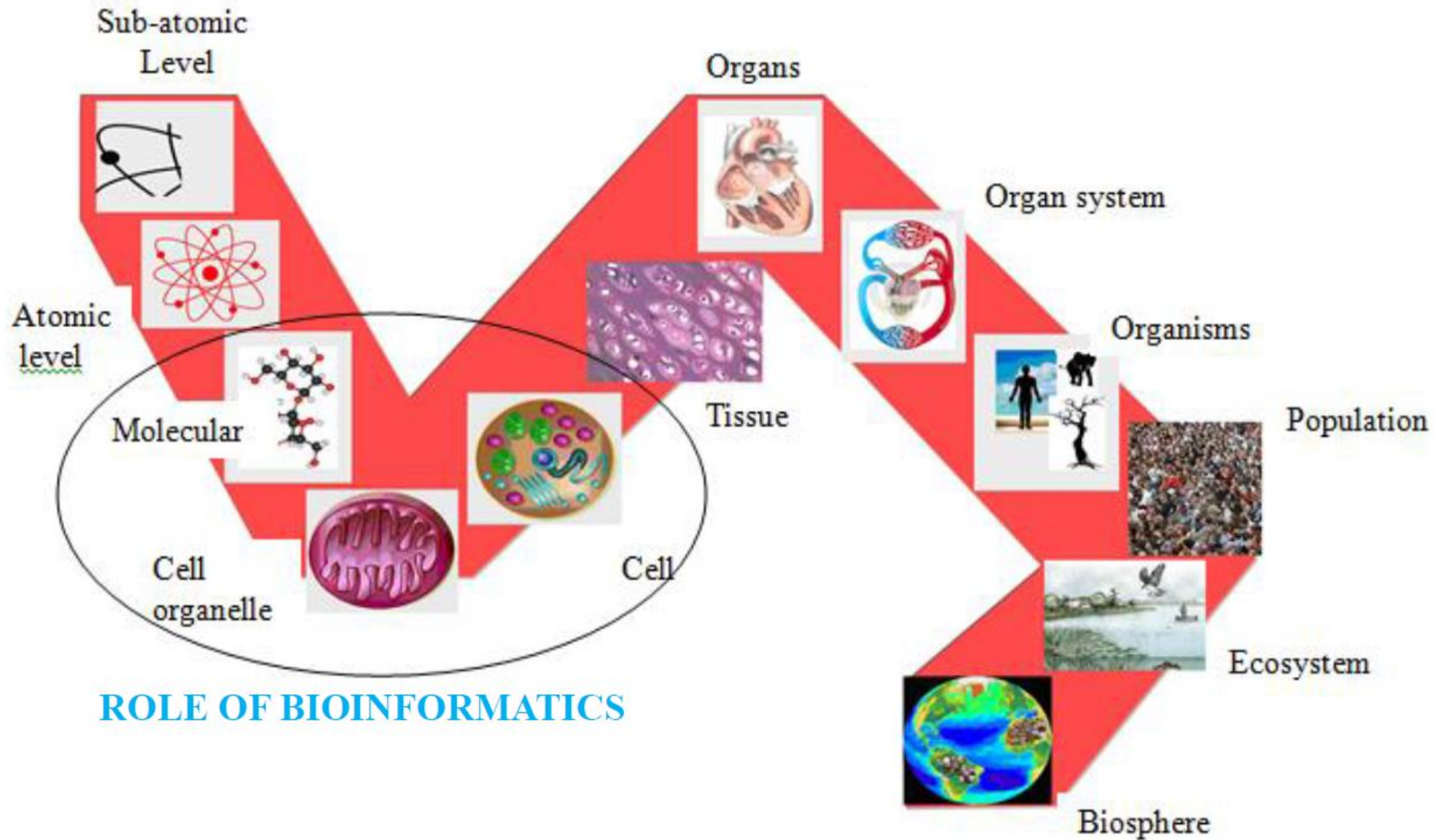


Bioinformatics - play with sequences & structures

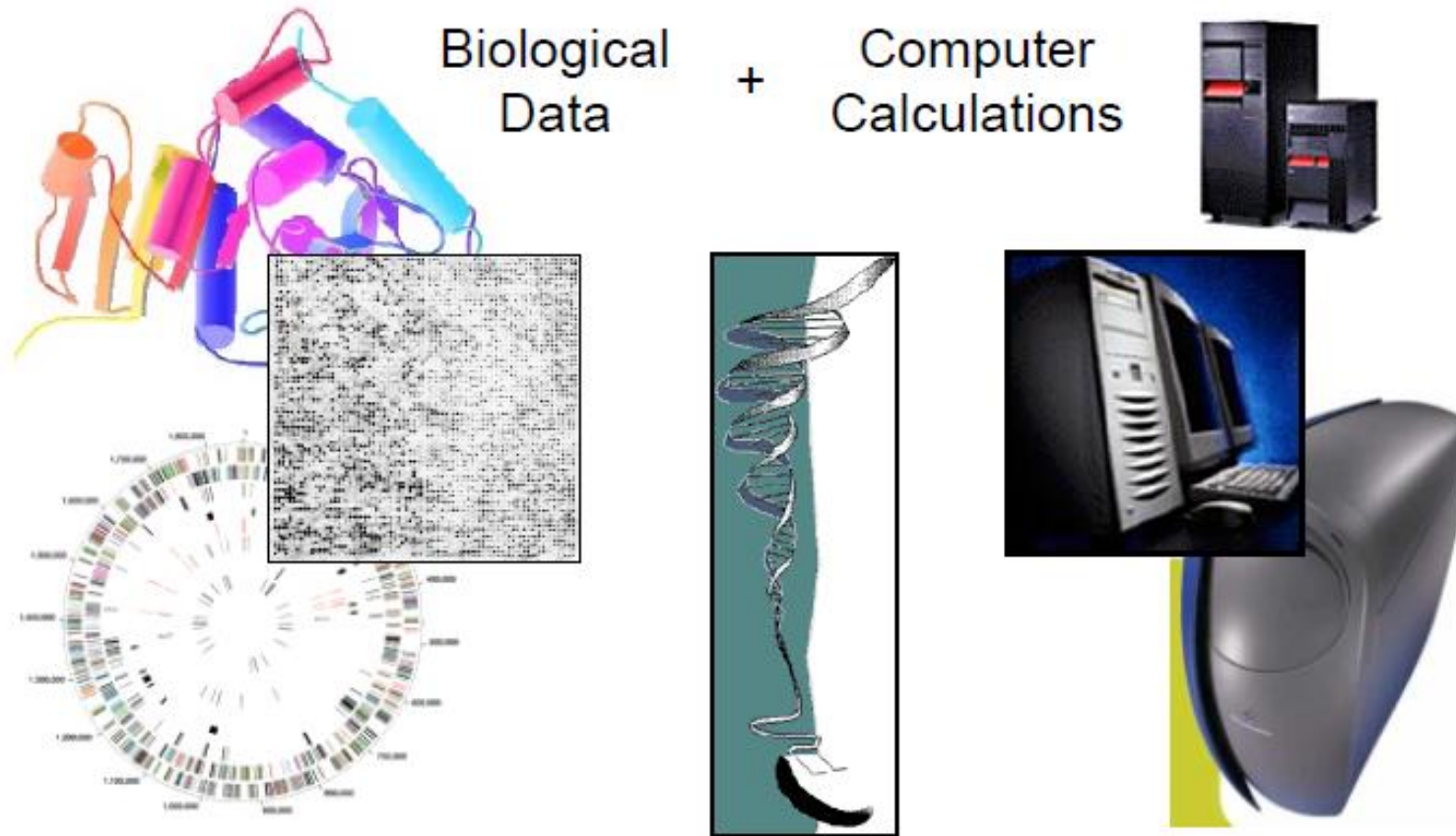
```
GAATTCCTTTGGTATCCAATGAAGAAATCGAATCCATACCCATAG|
TTCAGGAGAAAATAAGACCGAAGCTGCTCAATTAGGCGCAATTGAT|
GTGAAACTTGCCAGCTTACTTCGGCATGTCCTGGTCATTTTGGAAA.
CAACCATTATTTAAAGTCGCATTTAAAAAACTTGTTGAAAATATTT|
TTCTTTCTGTGGTGCTTTACAAAATCTTGAACCTCTGGAATTGATC.
AACGAAATACTGGAATAACAGTTAAAGATCGTGCTGCTTTTAAAAA.
ACCAAACAAAGCAAATTCAAGTGTATTGCACCTAATTGCCAAAAAC.
ACAATATTCGAAAAATAATAACTTTATATATAATTTCGGGTACTACA.
TGGATAACAGGCATGTGTTTAATATCTTACAAAATCTTCCACAAAC|
TTAACCCCTTCGAATGCTCATCAAATCGTATCTCCGAAAATGTCT|
TATCTTACTTCCACCACATAATCTACGAACATCAATGTTTATGAT|
GTTTGTTAACAAGTGATTTGAATCTGATAATGCGAAGAGTTGCTAA|
GCAAAAATACAAAAAATCTTGGATTCTATCGATAACAGCCGAGGTG|
TACAAATAAAAAGCTTACTTTGGATACTTTGACAGGTGGACACTCA.
TGCGAAGTTATATTAATGGCAAACGTATTCCTGAGACTGCCAGAGC|
TCTATGAATAAACTGGCTTTATTGAAGTACCATCTTACATTTTAA.
TGTTGTCCTTTTATAATCACGTTACGAAAGATAACATACTCAAAGT|
AAGCTTTTCTAACATATATCAAAAGTGATCATAATTCTGAAAATCC|
GATTTAGCACAGAAGAATGGATATTTAACCTTGGCTCCTAATTTTCG|
AAAAAGGAAAGAGGAAGGTGGTTTTGTAACTATTTGCAGACATCCA|
GCTTTTGGTATCCAATGAAGAAATCGAATCCATACCCATAG|
```



Organization of Life



Bioinformatics



Paulien Hogeweg & Ben Hesper coined the term 'Bioinformatics' in 1970.

WHAT IS BIOINFORMATICS?

Computational Biology/Bioinformatics is the application of computer sciences and allied technologies to answer the questions of Biologists, about the mysteries of life. It has evolved to serve as the bridge between:

- ☐ Observations (data) in diverse biologically-related disciplines and
- ☐ The derivations of understanding (information)

APPLICATIONS OF BIOINFORMATICS:

- ☐ Computer Aided Drug Design
- ☐ Microarray Bioinformatics
- ☐ Proteomics
- ☐ Genomics
- ☐ Biological Databases
- ☐ Phylogenetics
- ☐ Systems Biology

What is Bio- sequence

DNA, RNA or protein information represented as a series of bases (or amino acids) that appear in bio-molecules. The method by which a biosequence is obtained is called Bio-sequencing.

GTCTGATAAGTCAGTGTCTCC
GAGTCTAGCTTCTGTCCATGCT
GATCATGTCCATGTTCTAGTCAT
GATAGTTGATTCTAGTGTCTCT

DNA/ RNA
SEQUENCE

TPPUQWRDCCLKSWCUWMF
ESPWYZWEGHILDDFPTCTWF
CCDTWCUWGHISTDTKK SUN
RGHPPHLLDTWQESRND CQE

PROTEIN
SEQUENCE

What is Sequence Alignment

Arranging DNA/protein sequences side by side to study the extent of their similarity

BIOLOGICAL DATABASES

Biological Databases:

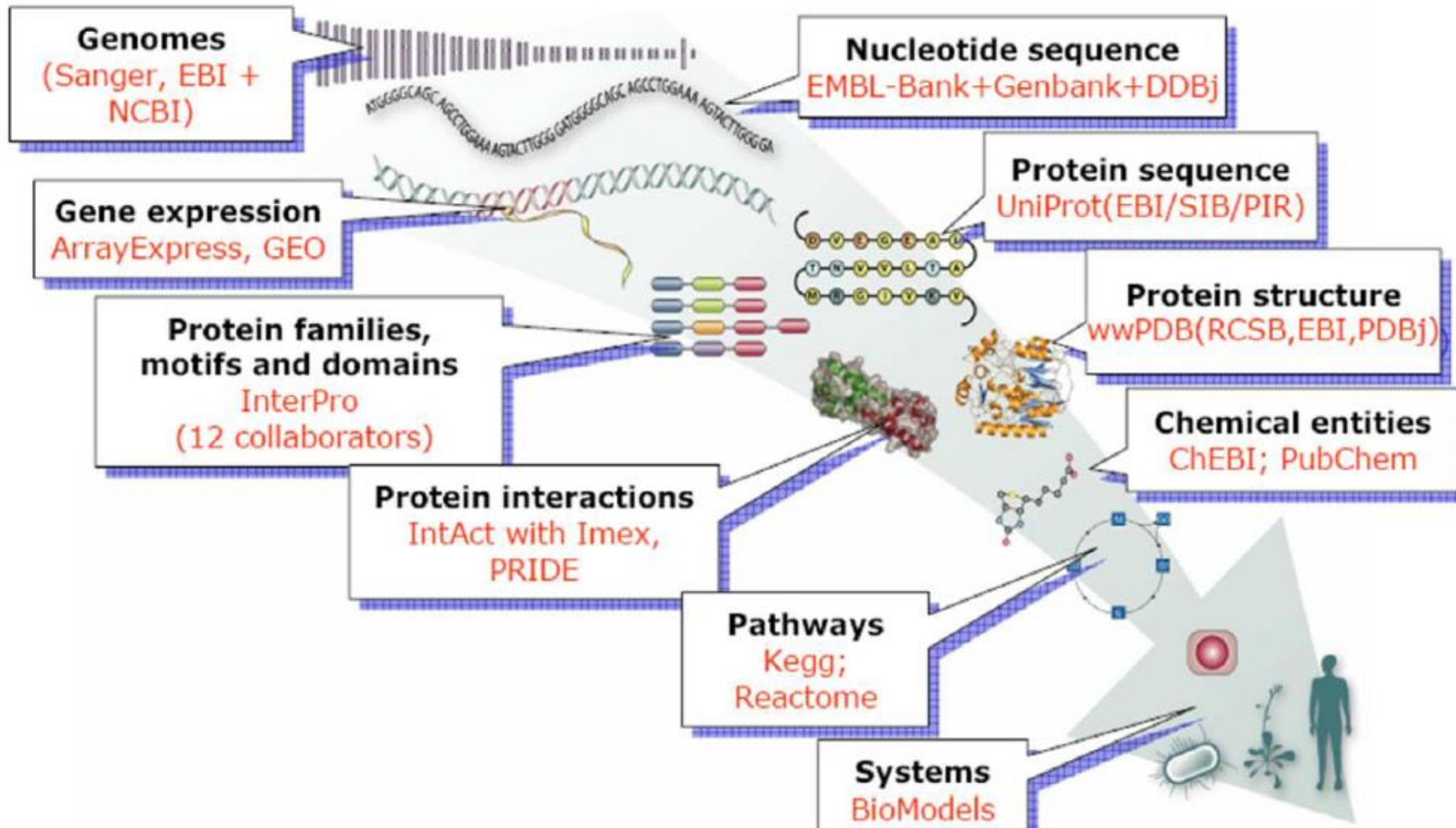
A collection of biological data and biological informations in electronic form and organized manner, which is accessible in various way, structures, searchable, updated periodically and cross referenced.

Purpose of biological databases:

- ❖ Availability and of biological data.
- ❖ Computation and systematization of biological information and data.
- ❖ Efficient manipulation of large data sets
- ❖ Structural and functional analysis of different biological data sets (Analysis of Nucleic acids & Protein Sequences. Sequence alignment and molecular modeling)

BIOLOGICAL DATABASES

Core Databases



Classification of Biological Databases

```
graph LR; Root[Classification of Biological Databases] --> Source[Based on Data Source]; Root --> Type[Based on Data Type]; Source --> Primary[Primary Databases]; Source --> Secondary[Secondary Databases]; Type --> N1[NUCLEOTIDE SEQUENCE DATABASE]; Type --> N2[PROTEIN SEQUENCE DATABASE]; Type --> N3[GENOME DATABASE]; Type --> N4[GENE EXPRESSION DATABASE]; Type --> N5[ENZYME DATABASE]; Type --> N6[STRUCTURE DATABASE]; Type --> N7[PROTEIN INTERACTION DATABASE]; Type --> N8[PATHWAY DATABASE]; Type --> N9[LITERATURE DATABASE];
```

Based on
Data Source

Primary Databases

Secondary Databases

Based on
Data Type

- NUCLEOTIDE SEQUENCE DATABASE
- PROTEIN SEQUENCE DATABASE
- GENOME DATABASE
- GENE EXPRESSION DATABASE
- ENZYME DATABASE
- STRUCTURE DATABASE
- PROTEIN INTERACTION DATABASE
- PATHWAY DATABASE
- LITERATURE DATABASE

Primary Data

S. No.	DATABASE	WEB- LINK
1	NCBI- National Centre for Biotechnology Information	www.ncbi.nlm.nih.gov
2	EMBL- European Molecular Biology Laboratory	www.ebi.ac.uk/embl
3	DDBJ- DNA Data Bank of Japan	www.ddbj.nig.ac
4	NDB- Nucleotide Data Bank	www.ndbserver.rutgers.edu
5	PIR- Protein Information Resource	www.pir.georgetown.edu
6	PDB- Protein Data Bank	www.rcsb.org/pdb
7	UniProt- Universal Protein Resource	

Primary Databases:

- ☐ Contain bio-molecular data in its original form.
- ☐ First-hand information of experimental data from scientists and researchers
- ☐ Data not edited or validated
- ☐ Raw and original submission of data
- ☐ Experimental results are submitted directly into the database by researcher,
- ☐ the data are archival in nature
- ☐ Made available to public for annotation
- ☐ Once given a database accession number the data in primary database are never changed

Secondary Databases

- ☐ Contains data derived from the results on analysing primary data
- ☐ Manually created or automatically generated.
- ☐ Contains more relevant and useful information structured to specific requirement.
- ☐ Data of highest quality as it is double checked.

Secondary Databases

S. No.	DATABASE	WEB- LINK
1	PROSITE (Protein domains, families, functional sites)	www.expasy.org/prosite
2	Pfam (Protein families)	www.sanger.ac.uk/pfam
3	SCOP (Structural Classification Of Proteins)	www.scop.mrc-lmb.cam.ac.uk/scop
4	CATH (Class, Architecture, Topology, Homologous Super Family of Proteins)	www.cathdb.info
5	OMIM (Online Mendelian Inheritance in Man)	www.ncbi.nlm.nih/omim
6	KEGG (Kyoto Encyclopedia of Genes and Genome)	www.genome.jp/kegg/pathway.html
7	MetaCyc (Enzyme Metabolic Pathways)	www.metacyc.org
8	<i>ExPASY</i>	http://web.expasy.org
9	<i>SWISS-Prot</i>	www.ebi.ac.uk/swissprot/access.html
10	<i>Enzyme structure</i>	http://www.ebi.ac.uk/thornton-srv/databases/enzymes
	Interpro	

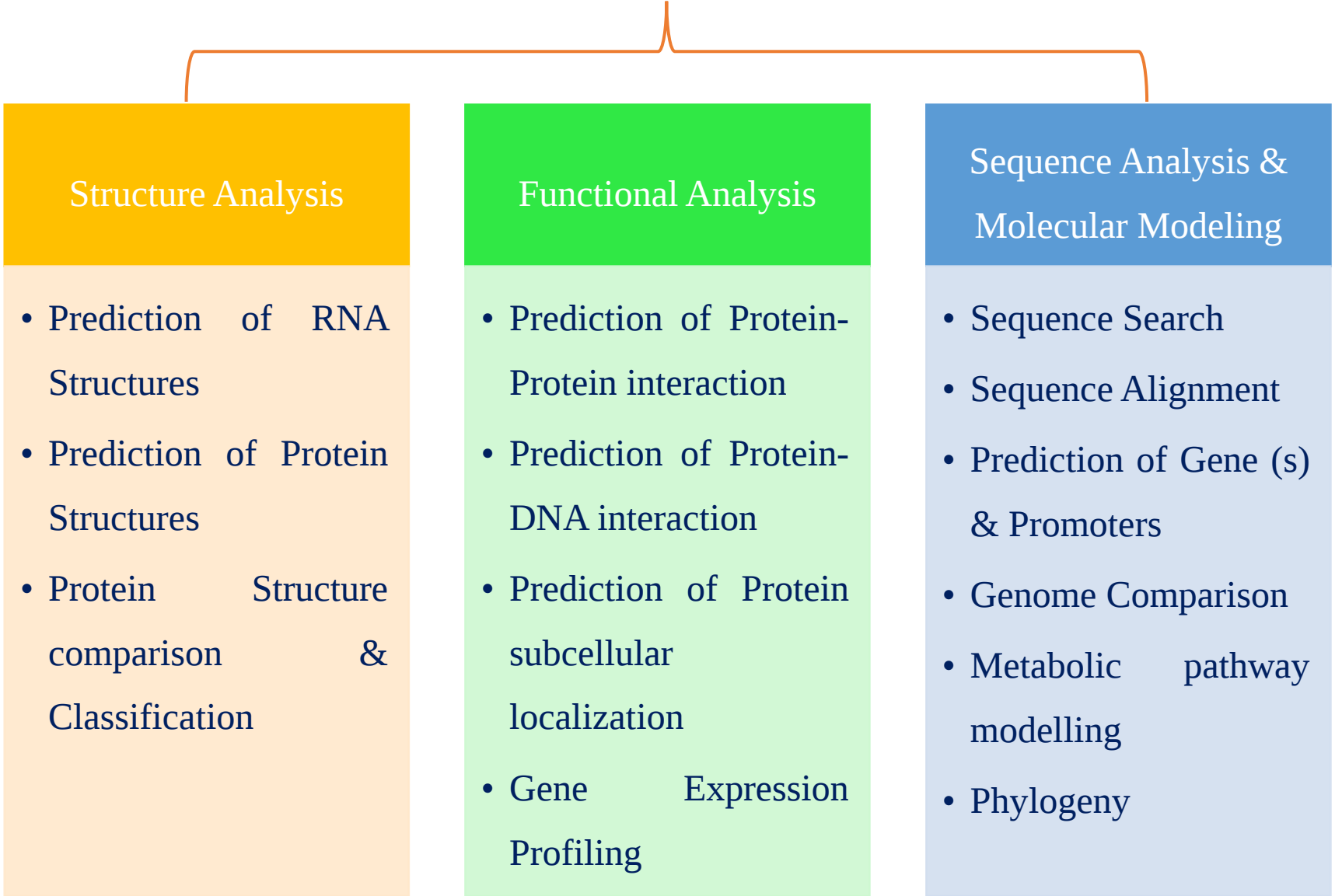
Significance of Biological Databases:

- ❖ Structural and functional analysis of different biological data sets (Analysis of Nucleic acids & Protein Sequences. Sequence alignment and molecular modeling)
 - DNA and Protein Sequence Assembly: A process through which short DNA sequence fragments (called reads or samples) are merged into longer DNA sequence to reconstruct the original DNA sequences.
 - Sequence searches
 - Sequence alignments
 - Motif discovery
 - Gene and promoter prediction

Significance of Biological Databases:

- Genome Annotation: It is the process of identifying the location of genes and all of the coding regions in a genome and determining what those genes do.
 - It consists of 3 main steps:
 - ✓ Identifying portion of the genome that do not code for proteins
 - ✓ Identifying elements on the genome.
 - ✓ Attaching biological information to these elements.
- Inferences of Protein and RNA structure:
 - Prediction of three dimensional structure of a protein from its amino acid sequence
 - Comparison of protein structures
 - Protein structure classification
 - Prediction of 3- D structure of different RNA molecules from ribonucleotide sequences.
- ❖ Reduce the time consuming experimental procedures.

Applications of Bioinformatics



Structure Analysis

- Prediction of RNA Structures
- Prediction of Protein Structures
- Protein Structure comparison & Classification

Functional Analysis

- Prediction of Protein-Protein interaction
- Prediction of Protein-DNA interaction
- Prediction of Protein subcellular localization
- Gene Expression Profiling

Sequence Analysis & Molecular Modeling

- Sequence Search
- Sequence Alignment
- Prediction of Gene (s) & Promoters
- Genome Comparison
- Metabolic pathway modelling
- Phylogeny

Thank You