

## Bioinformatics

The bioinformatics can be defined as conceptualizing “Biology” in terms of macromolecules (i.e. DNA, RNA, and proteins) in the sense of physical chemistry and then applying informatics techniques (like applied mathematics, computer science and statistics) in order to understand and organize the information associated with these molecules on a large scale basis.

The bioinformatics has led to the discovery of the genetic determinants of human disease (molecular medicine), the basis of the various production traits in animals and plants, the development of drug discovery tools (pharmacogenomics), the discovery of gene polymorphism including the single nucleotide polymorphism (SNPs) which determine the varied individual response to various drug (pharmacogenetics) and led to the emergence of new fields like comparative genomics, genetic epidemiology and evolutionary genetics.

### Biological databases and database searching

A database is a collection of interrelated data from which some information can be extracted. The collection of data is logically coherent with some inherent meaning. A database is designed and built for a specific purpose, keeping in mind the needs of the applications in which the data is going to be utilized as also the end users of those applications.

*The biological database contain usually DNA or protein sequence.*

Major databases:

|         |                                                                      |                                                                              |
|---------|----------------------------------------------------------------------|------------------------------------------------------------------------------|
| GenBank | <a href="http://www.ncbi.nlm.nih.gov">www.ncbi.nlm.nih.gov</a>       | DNA and protein sequences                                                    |
| EMBL    | <a href="http://www.ebi.ac.uk/embl.html">www.ebi.ac.uk/embl.html</a> | DNA and protein sequences along with the tertiary structure of some proteins |
| DDBJ    | <a href="http://www.ddbj.nig.ac.jp">www.ddbj.nig.ac.jp</a>           | All known nucleotide and protein sequences.                                  |

### Data retrieval systems

The maintenance of database is not sufficient. The system of retrieval of specific data from a sea of data is needed. For the retrieval of specific data, various database search tools have been devised such as BLAST (Basic Local Alignment Search Tool) that is maintained by National center of Biotechnology information (NCBI) and can be obtained from the website [www.ncbi.nlm.nih.gov/Blast](http://www.ncbi.nlm.nih.gov/Blast). *The term alignment indicates the similarity of sequences between or among two or more different either DNA or protein sequence.* Other retrieval system includes, T-cell and B-cell epitope finders ([www.immuneepitope.com](http://www.immuneepitope.com))

1. **BLAST:** BLAST searches for the similarity of the sequences entered. For example, if you input an unknown sequence of DNA or protein the BLAST, the BLAST would search for the similarity of the input DNA or protein sequence in the database containing large numbers of various protein and DNA sequences and would display the sequences that are similar to the input sequence.
2. **T-cell and B-cell epitope finders:** This tool finds the B-cell epitope and T-cell epitope of the input protein sequence. From this tool the prediction of these epitopes can be made with some accuracy though not 100% accurate. From this tool, the immunogenicity of a protein can be predicted.

### 3. Examples of Software programs:

- a. **DNA-STAR:** DNA-STAR can be used to perform multiple analysis of input data. For example, the software DNA-STAR can be used to predict the secondary structure of protein sequences along with the antigenic index of the protein. Apart from this software can also be used for drawing restriction enzyme map, PCR primer designing and finding out the protein sequence from the input DNA sequence.
- b. **TREECON software:** This software can make phylogenetic tree of different species of animals based on the pattern of the RFLP or number of different amplified DNA generated during PCR of microsatellite (STR) or RAPD.

### Applications of Bio-informatics

1. **Comparative genome analysis:** comparison of genome sequence of different animals belonging to different species or different breeds of same species. This can be utilized for construction of phylogenetic tree.
2. **Classification of protein:** The protein classification is performed based on sequence of protein. This can be easily retrieved by BLAST. This system can be useful in finding out the function of unknown protein.
3. **Vaccine designing:** the tools such as B-cell and T-cell epitope finder can be used to choose the proteins having more number of immunogenic epitope for designing a vaccine.
4. **Bioinformatics aided Drug Designing:** there are two approaches of Bioinformatics aided Drug Designing viz. *structure based approach* and *ligand based approach*.
  - a. **Structure based approach:** in this approach, the three dimensional structure of receptor to which the particular drug binds is found out. Many innovative approaches have been developed to construct drugs that would fit the receptor in question.
  - b. **ligand based approach:** in this approach, the physico-chemical or structural properties of ligand molecules are characterized. A classical example of this concept is a quantitative structure-activity relationship model construction that help in finding out its prospective receptor to which it would bind efficiently.
5. **PCR primers and oligonucleotide probe designing:** in today's date the PCR has become a routine tool for molecular biologists and which necessitates the designing the primers for specific DNA sequences. These primers can be designed online in NCBI or using some softwares such as DNA-STAR, Genetool etc.