



Bioinformation Up to Date

(Bioinformatics Center, Biotechnology Division)
North-East Institute of Science and Technology

Contents

Cover Story	1
Computational Biology	1
Computational Chemistry	2
Proteomics	2
Genomics	2
Software Mania	3
Bio Server	3
Current Trends	3
Bioinfo Quiz	3
Computers for Biologist	4
Molecule of the Month	4
Bioinfo Animator	4

Editors

Salam Pradeep Singh
Dr. R.L. Bezbaruah
Adviser
Dr. P.G. Rao

North East Research Labs

- North East Institute of Science and Technology, Jorhat
- North East Regional Institute of Science and Technology, Itanagar.
- Institute of Bioresource and Sustainable Development, Imphal.

Cover Story

Bioinformatics Infrastructure Facility at NEIST

The Department of Biotechnology, Government of India, has sanctioned with a release of Rs. 20 lacs for establishment of Bioinformatics Infrastructure Facility (BIF) for the promotion of Biology Teaching Through Bioinformatics (BTBI) at the North East Institute of Science and Technology, Jorhat, Assam on 14th February 2008.

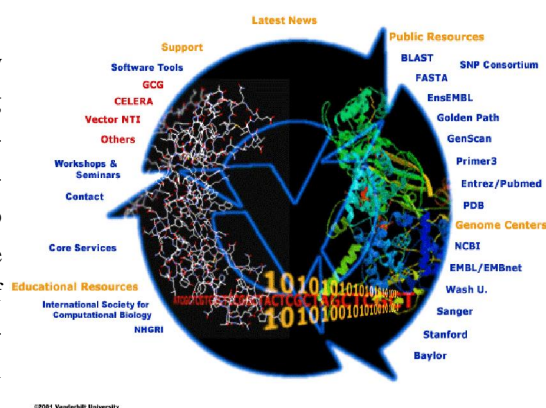
The objectives of the BIF are as follows:

- To establish BIF to support the teaching activities of biology and its allied areas of the host institute in particular and the neighboring institutions in general.
- To build up information resource, prepare database of interest to its users and to develop relevant information handling tools and techniques.
- To assess information requirements, organize creation of necessary infrastructure and to provide information and computer support service to the users.
- To establish linkages with BTISnet of DBT for sharing Information Resource and Expertise.
- To organize Training/Workshop for familiarizing the applications of Bioinformatics in Biology teaching and learning activities.

.....Biotechnology Division, NEIST

Bioinformatics—What is it ?

Bioinformatics and computational biology involve the use of techniques including applied mathematics, informatics, statistics, computer science, artificial intelligence, chemistry, and biochemistry to solve biological problems usually on the molecular level. The core principle of these techniques is using computing resources in order to solve problems on scales of magnitude far too great for human discernment. Research in computational biology often overlaps with systems biology. Major research efforts in the field include sequence alignment, gene finding, genome assembly, protein structure alignment, protein structure prediction, prediction of gene expression and protein-protein interactions, and the modeling of evolution. Bioinformatics more properly refers to the creation and advancement of algorithms, computational and statistical techniques, and theory to solve formal and practical problems arising from the management and analysis of biological data.



©2001 Veeva/KB University

Computational Chemistry

Quantitative Structure-Activity Relationship (QSAR)

Quantitative structure-activity relationship (QSAR) is the process by which chemical structure is quantitatively correlated with a well defined process, such as biological activity or chemical reactivity. For example, biological activity can be expressed quantitatively as in the concentration of a substance required to give a certain biological response. Additionally, when physiochemical properties or structures are expressed by numbers, one can form a mathematical relationship, or quantitative structure-activity relationship, between the two. The mathematical expression can then be used to predict the biological response of other chemical structures.

QSAR's most general mathematical form is: **Activity = f (physiochemical properties and/or structural properties)**

Biological Applications

The biological activity of molecules is usually measured in assays to establish the level of inhibition of particular signal transduction or metabolic pathways. Chemicals can also be biologically active by being toxic. Drug discovery often involves the use of QSAR to identify chemical structures that could have good inhibitory effects on specific targets and have low toxicity (non-specific activity). Of special interest is the prediction of partition coefficient log P, which is an important measure used in identifying "drug likeness" according to Lipinski's Rule of Five.

While many QSAR analyses involve the interactions of a family of molecules with an enzyme or receptor binding site, QSAR can also be used to study the interactions between the structural domains of proteins. Protein-protein interactions can be quantitatively analyzed for structural variations resulted from site-directed mutagenesis.

Source: Wikipedia® - Wikimedia Foundation Inc, USA (GNU Free Documentation License)

Proteomics

Protein Data Bank



The Protein Data Bank (PDB) is a repository for 3-D structural data of proteins and nucleic acids. These data, typically obtained by X-ray crystallography or NMR spectroscopy and submitted by biologists and biochemists from around the world, are released into the public domain, and can be accessed for free online.

It was founded in 1971 by Drs. Edgar Meyer and Walter Hamilton Brookhaven National Laboratory and contained just 7 protein structures, and it was transferred in 1998 to the Research Collaboratory for Structural Bioinformatics (RCSB). Rutgers University, New Jersey is the lead site and is currently under the direction of Helen M. Berman.

As of 15 April 2008, the database contained 50,277 released atomic coordinate entries or structures, 46,400 of that proteins, the rest being nucleic acids, nucleic acid-protein complexes, and a few other molecules. About 5,000 new structures are released each year. The database stores information about the exact location of all atoms in a large biomolecule.

Source: RCSB Protein Data Bank, USA

Genomics

National Center for Biotechnology Information



The National Center for Biotechnology Information (NCBI) is part of the United States National Library of Medicine, a branch of the National Institutes of Health. The NCBI is located in Bethesda, Maryland and was founded in 1988.

The NCBI houses genome sequencing data in GenBank and an index of biomedical research articles in PubMed Central and PubMed. All these databases are available online through the Entrez search engine.

NCBI is directed by David Lipman, one of the original authors of the BLAST sequence alignment program. The NCBI has had responsibility for making available the GenBank DNA sequence database since 1992. GenBank coordinates with the European Molecular Biology Laboratory (EMBL) and the DNA Database of Japan (DDBJ).

Since 1992, NCBI has grown to provide other databases in addition to GenBank. NCBI provides Online Mendelian Inheritance in Man, the Molecular Modeling Database (3D protein structures), dbSNP a database of Single Nucleotide Polymorphisms, the Unique Human Gene Sequence Collection, a Gene Map of the Human genome and a Taxonomy Browser.

Source: National Center for Biotechnology Information, USA

Software Mania

DeepView - Swiss PDB Viewer

DeepView - Swiss-PdbViewer is an interactive molecular graphics program for viewing and analyzing protein and nucleic acid structures. It has been developed by Nicolas Guex (GlaxoSmithKline R&D). Swiss-PdbViewer is tightly linked to SWISS-MODEL, an automated homology modeling server developed within the Swiss Institute of Bioinformatics (SIB) at the Structural Bioinformatics Group at the Biozentrum in Basel. Some of the important features available are:

Graphic window - To visualize loaded molecules, which can be rotated, translated and zoomed.

Control Panel - Control the visual representation. Display of backbones, side chains, labels, molecular surfaces, and ribbons for each group; and set the colors for the different objects on display.

Alignment Window - Shows the amino-acid sequence of proteins in one-letter code to compare & align sequences of two or more proteins.

Ramachandran Plot window - Displays a Ramachandran plot & gives the ϕ and ψ angles of one selected residue of the loaded protein.

Source: GlaxoSmithKline, Geneva



Bio Servers

Swiss-Model Server

SWISS-MODEL is a fully automated protein structure homology-modeling server, accessible via the ExPASy web server, or from the program DeepView (Swiss Pdb-Viewer). The purpose of this server is to make Protein Modelling accessible to all biochemists and molecular biologists World Wide. It assists and guides the user in building protein homology models at different levels of complexity.

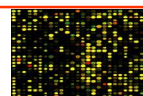
It was initiated in 1993 by Manuel Peitsch, and further developed at Glaxo Wellcome Experimental Research in Geneva and the SIB Swiss Institute of Bioinformatics. Since 2001, SWISS-MODEL is being developed by Torsten Schwede's Structural Bioinformatics Group at the SIB & Biozentrum.

Building a homology model with the SWISS-MODEL sever comprises four main steps: identification of structural template(s), alignment of target sequence and template structure(s), model building, and model quality evaluation. These steps can be repeated until a satisfying modelling result is achieved. Each of the four steps requires specialized software and access to up-to-date protein sequence and structure databases.

Source: Swiss Institute of Bioinformatics & Biozentrum, Switzerland

Current Trends

DNA Micro Array Technology



DNA microarrays are created by robotic machines that arrange minuscule amounts of hundreds or thousands of gene sequences on a single microscope slide. Researchers have a database of over 40,000 gene sequences that they can use for this purpose. When a gene is activated, cellular machinery begins to copy certain segments of that gene. The resulting product is known as messenger RNA (mRNA), which is the body's template for creating proteins. The mRNA produced by the cell is complementary, and therefore will bind to the original portion of the DNA strand from which it was copied. To determine which genes are turned on and which are turned off in a given cell, a researcher must first collect the messenger RNA molecules present in that cell. The researcher then labels each mRNA molecule by attaching a fluorescent dye. Next, the researcher places the labeled mRNA onto a DNA microarray slide. The messenger RNA that was present in the cell will then hybridize or bind to its complementary DNA on the microarray, leaving its fluorescent tag. A researcher then use a special scanner to measure the fluorescent areas on the microarray. If a particular gene is very active, it produces many molecules of messenger RNA, which hybridize to the DNA on the microarray and generate a very bright fluorescent area. Genes that are somewhat active produce fewer mRNAs, which results in dimmer fluorescent spots. If there is no fluorescence, none of the messenger molecules have hybridized to the DNA, indicating that the gene is inactive. This technique is frequently use to examine the activity of various genes at different times.

Source: National Human Genome Research Institute, USA

Bioinfy Quiz - 001

1. Which of these codons translate to Trp in the mammalian mitochondrial code?

- a) UAA
- b) UAG
- c) UGA

2. What reagent is used to measure the number of thiol groups in a protein?

- a) Ellman's reagent
- b) Mercaptoethanol
- c) Ninhydrin

3. Who demonstrated the semi-conservative nature of DNA replication?

- a) Inman R. and Schn's M
- b) Meselson M. and Stahl H
- c) Okazaki R

4. What is the approximate percentage of repetitive DNA sequences in human DNA?

- a) 10 to 20%
- b) 20 to 30%
- c) 30 to 40%

5. The disease Beriberi is due to a dietary deficiency in:

- a) thiamine
- b) riboflavin
- c) pyridoxine

Computers for Biologist

BIO-PERL - Practical Expansion and Reporting Language



BioPerl is a collection of Perl modules that facilitate the development of Perl scripts for bioinformatics applications. It has played an integral role in the Human Genome Project. It is an open source software project which is still under active development. The first stable release was on June 11, 2002. Version 1.5.2 is considered to be the most stable (in terms of bugs) version of BioPerl and believed to be suitable for most peoples requirements.

In order to take advantage of BioPerl, the user needs a basic understanding of the Perl programming language including an understanding of how to use Perl references, modules, objects and methods. BioPerl provides software modules for many of the typical tasks of bioinformatics programming. These include:

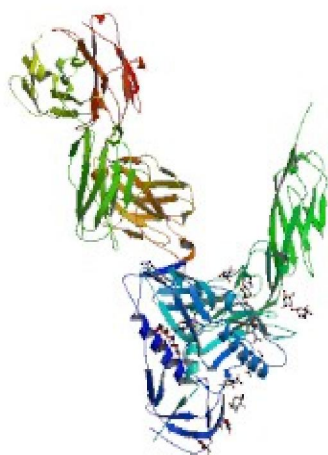
Accessing sequence data from local and remote databases , Transforming formats of database/ file records , Manipulating individual sequences; Searching for similar sequences; Creating and manipulating sequence alignments; Searching for genes and other structures on genomic DNA and Developing machine readable sequence annotations

Source: www.bioperl.org/

Molecule of the Month

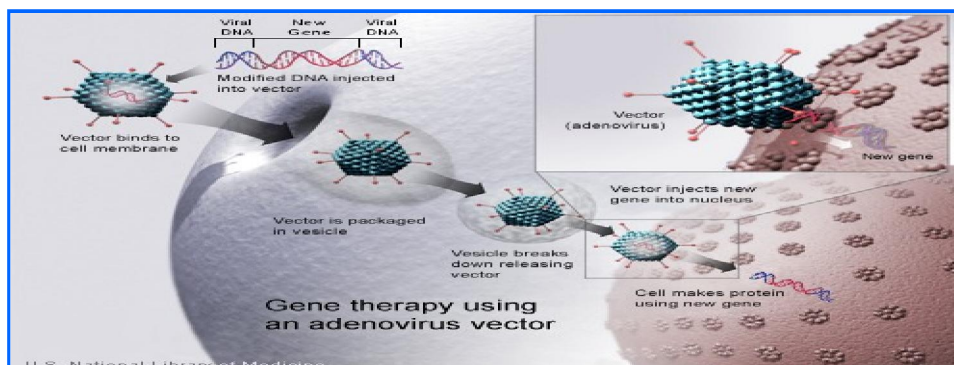
GP 120

This is the envelope glycoprotein of HIV, that causes AIDS. It forms the coat for HIV. If the virus has to enter the host cell then the GP120 has to interact with the human T-cell surface glycoprotein CD4 and a chemokine receptor on the surface. This type of protein i.e. GP120 are found only in HIV type I, type II and Simian Immunodeficiency Virus (SIV). This is the protein that is targeted to combat HIV. Drugs and method are being devised to neutralize the effect of GP 120.



Molecular Data

PDB ID : 1G9M
Amino acids : 949
Atoms : 8322
Mol. Wt. : 103728 Da
Chains : G, C, L, H (4)



Bioinfo Animator - A new gene is inserted into an adenovirus vector, which is used to introduce the modified DNA into a human cell. If the treatment is successful, the new gene will make a functional protein.

Source: National Library of Medicine, USA



NORTH EAST INSTITUTE OF SCIENCE AND TECHNOLOGY
JORHAT- 785006, ASSAM

NEIST
A National R & D Institute of CSIR

The North East Institute of Science and Technology, formerly (RRL) , Jorhat was established in the year 1961 as one of the multidisciplinary laboratories of Council of Scientific & Industrial Research (CSIR) under its Chemical Science Group of laboratories. Its major thrust of R&D activities has been to develop indigenous technologies by utilizing the immense natural wealth of India.

The North Eastern Region of the country being bestowed with an abundance of material resources like petroleum, natural Gas, minerals, tea as well as aromatic and Medicinal plants and hence the laboratories was targeted to undertake research for development of Know-How for a wide a range of industries and extension works. Over the years, the laboratory has generated more than 100 technologies in the areas of agro technology, biological and oil field chemicals of which about 40% were of commercial success culminating in setting up of various industries through out the country. The laboratory also developed expertise in the areas like Natural Products Chemistry, drug and drug intermediates, VSK cement, Plant Technology, Agro-technologies, Petroleum Microbiology and Petrochemicals, Crude oil transportation, Paper and Paper Products, beneficiation Chemicals, ecology and environmental studies, Geotechnical investigations, foundation design engineering, soil and building materials etc.

Please make your own contributions, please contact:

Salam Pradeep Singh (CSIR Diamond Jubilee Research Internship)
Email: salampradeep@gmail.com;
C/o Dr. R.L. Bezbaruah ; Biotechnology Division, NEIST, Jorhat

Answers to Bioinfy Quiz

1-c ; 2-a ; 3-b ; 4-b; 5-a