



Bioinformation up to Date

(Bioinformatics Center, Biotechnology Division)
North-East Institute of Science & Technology
Jorhat - 785 006, Assam

Contents

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

Adviser

Dr. P.G. Rao

Editors

Salam Pradeep Singh

Dr. R.L. Bezbaruah

BIF Upcoming Events

1. 6 Days Training Programme on "Bioresource Database" @ Sikkim State Council of Science & Technology, Gangtok from August 18 - 23, 2008.

2. 4 Days General Bioinformatics Training on "Bioinformatics: Current approaches & applications" @ College of Veterinary Science, AAU, Guwhati from September 23 - 26, 2008.

Cover Story

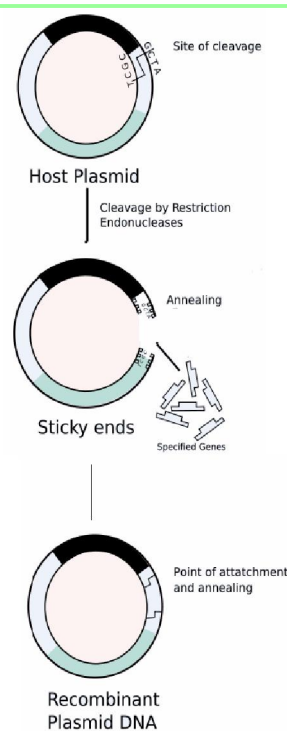
Biotechnology Information System (BTISnet)

India was the first country in the world to establish National Bioinformatics Network. Due to personal intervention and support by the then Prime Minister Late Rajiv Gandhi, India established a Distributed Bioinformatics Network in 1986-87. The culture of use of computers in biology and importance of quantization in biology in India was the result of this network. The BTISnet covers institutions under DST, CSIR, ICMR, ICAR, Universities and Institutes under Human Resource Ministry. Today an extensive Bioinformatics Network, covering 120 institutions, spread geographically all over the country, has been established. The Network is creating human resource in Bioinformatics and carrying out research in different areas of Bioinformatics. Scientists of this network have published more than 1000 bioinformatics research papers in peer reviewed journals in last five years and helped in publishing more than 3000 research papers in biology/biotechnology. The network has also helped directly or indirectly several Bioinformatics companies in India such as Biosuite of Tata Consultancy Services (TCS) or software packages for visualization of bioinformatics data by Strand Genomics. The BTISnet is the first network which established BioGrid India of large bandwidth and high speed connectivity among various institutions in the country and also high performance national computing facility. Thus, this unique network has showed that India can work in network consortium mode.

Special Interests

Recombinant DNA Technology

Recombinant DNA (rDNA) is a form of artificial DNA that is engineered through the combination or insertion of one or more DNA strands, thereby combining DNA sequences that would not normally occur together. In terms of genetic modification, rDNA is produced through the addition of relevant DNA into an existing organismal genome, such as the plasmid of bacteria, to code for or alter different traits for a specific purpose, such as immunity. It differs from genetic recombination, in that it does not occur through processes within the cell or ribosome, but is exclusively engineered. The Recombinant DNA technique was engineered by Stanley Norman Cohen and Herbert Boyer in 1973. They published their findings in a 1974 paper entitled "Construction of Biologically Functional Bacterial Plasmids in vitro", which described a technique to isolate and amplify genes or DNA segments and insert them into another cell with precision, creating a transgenic bacterium. Recombinant DNA technology was made possible by the discovery of restriction endonucleases by Werner Arber, Daniel Nathans, and Hamilton Smith, for which they received the 1978 Nobel Prize in Medicine. Because of the importance of DNA in the replication of new structures and characteristics of living organisms, it has widespread importance in recapitulating via viral or non-viral vectors, both desirable and undesirable characteristics of a species to achieve characteristic change or to counteract effects caused by genetic or imposed disorders that have effects upon cellular or organismal processes. Through the use of rDNA, genes that are identified as important can be amplified and isolated for use in other species or applications, where there may be some form of genetic illness or discrepancy, and provides a different approach to complex biological problem solving.



Courtesy: Wikimedia Foundation, Inc, USA

Computational Chemistry

Molecular modeling

The Molecular modeling refers to theoretical methods and computational techniques to model or mimic the behaviour of molecules. The techniques are used in the fields of computational chemistry, computational biology and materials science for studying molecular systems ranging from small chemical systems to large biological molecules and material assemblies. The simplest calculations can be performed by hand, but inevitably computers are required to perform molecular modelling of any reasonably sized system. The common feature of molecular modelling techniques is the atomistic level description of the molecular systems; the lowest level of information is individual atoms or a small group of atoms. This is in contrast to quantum chemistry also known as electronic structure calculations where electrons are considered explicitly. The benefit of molecular modelling is that it reduces the complexity of the system, allowing many more particles to be considered during simulations. Molecular mechanics is one aspect of molecular modelling, as it refers to the use of classical mechanics/Newtonian mechanics to describe the physical basis behind the models. Molecular models typically describe atoms as point charges with an associated mass. The interactions between neighbouring atoms are described by spring-like interactions (representing chemical bonds) and van der Waals forces. The Lennard-Jones potential is commonly used to describe van der Waals forces. The electrostatic interactions are computed based on Coulomb's law. Atoms are assigned coordinates in Cartesian space or in internal coordinates, and can also be assigned velocities in dynamical simulations. The atomic velocities are related to the temperature of the system, a macroscopic quantity. The collective mathematical expression is known as a potential function and is related to the system internal energy (U), a thermodynamic quantity equal to the sum of potential and kinetic energies. Methods which minimize the potential energy are known as energy minimization techniques, while methods that model the behaviour of the system with propagation of time are known as molecular dynamics.

$$E = E_{bonds} + E_{angle} + E_{dihedral} + E_{non-bonded} \ \& \ E_{non-bonded} = E_{electrostatic} + E_{vanderWaals}$$

The above function, referred to as a potential function, computes the molecular potential energy as a sum of energy terms that describe the deviation of bond lengths, bond angles and torsion angles away from equilibrium values, plus terms for non-bonded pairs of atoms describing van der Waals and electrostatic interactions. The set of parameters consisting of equilibrium bond lengths, bond angles, partial charge values, force constants and van der Waals parameters are collectively known as a force field. Molecular modelling methods are now routinely used to investigate the structure, dynamics and thermodynamics of inorganic, biological, and polymeric systems. The types of biological activity that have been investigated using molecular modelling include protein folding, enzyme catalysis, protein stability, conformational changes associated with biomolecular function, and molecular recognition of proteins, DNA, and membrane complexes *Courtesy: Center for Molecular Modeling, USA*

Proteomics

Human Protein Reference Database

The Human Protein Reference Database (HPRD) is a protein database accessible through the internet. It is a result of an international collaborative effort between the Institute of Bioinformatics in Bangalore, India and the Pandey lab at Johns Hopkins University in Baltimore, USA. HPRD contains manually curated scientific information pertaining to the biology of most human proteins. Information regarding proteins involved in human diseases is annotated and linked to Online Mendelian Inheritance in Man database. The NCBI provides link to HPRD through its human protein databases through Entrez Gene, RefSeq protein pertaining to genes and proteins.

This resource depicts information on human protein functions including protein-protein interactions, post-translational modifications, enzyme-substrate relationships and disease associations. Protein annotation information that is catalogued was derived through manual curation using published literature by expert biologists and through bioinformatics analyses of the protein sequence. The protein-protein interaction and subcellular localization data from HPRD have been used to develop a human protein interaction network.

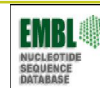
Some highlights of HPRD as follows:

- From 10,000 protein-protein interactions (PPIs) annotated for 3,000 proteins, HPRD has grown to over 36,500 unique PPIs annotated for 25,000 proteins.
- More than 50% of molecules annotated in HPRD have at least one PPI and 10% have more than 10 PPIs.

Courtesy: Human Protein Reference Database, Indo-US

Genomics

EMBL Nucleotide Sequence Database



The EMBL Nucleotide Sequence Database also known as EMBL-Bank (<http://www.ebi.ac.uk/embl/>), maintained at the European Bioinformatics Institute (EBI), 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). Each of the three groups collects a portion of the total sequence data reported worldwide, and all new and updated database entries are exchanged between the groups on a daily basis.

The web-based tool, Webin, is the preferred system for individual submission of nucleotide sequences, including Third Party Annotation and alignment data. Automatic submission procedures are used for submission of data from large-scale genome sequencing centres and from the European Patent Office. Database releases are produced quarterly. The EBI's Sequence Retrieval System (SRS) integrates and links the main nucleotide and protein databases as well as many other specialist molecular biology databases. For sequence similarity searching, a variety of tools (e.g. FASTA and BLAST) are available that allow external users to compare their own sequences against the data in the EMBL Nucleotide Sequence Database, the complete genomic component subsection of the database, the WGS data sets and other databases. All available resources can be accessed via the EBI home page at <http://www.ebi.ac.uk>.

Software Mania

Molegro Virtual Docker

Molegro Virtual Docker (MVD) is an integrated environment for studying and predicting how ligands interact with macromolecules. The identification of ligand binding modes is done by iteratively evaluating a number of candidate solutions (ligand conformations) and estimating the energy of their interactions with the macromolecule. The highest scoring solutions are returned for further analysis. MVD requires a three-dimensional structure of both protein and ligand usually derived from X-ray/NMR experiments or homology modeling. MVD performs flexible ligand docking, so the optimal geometry of the ligand will be determined during the docking. The user interface in MVD is composed of :

- 1. Visualization Window** - Visualizes all the selected molecules in the workspace and all custom graphical objects (e.g. labels, surfaces etc).
- 2. Tool Bar** - Provides easy & fast access to commonly used actions (e.g. import molecules, docking wizard, pose organizer etc).
- 3. Workspace Explorer** - Displays information about the 3D-objects such as proteins, ligands, and water molecules & also objects such as labels, surfaces, backbones, and cavities.
- 4. Properties Window** - Displays information about the currently selected or highlighted 3D object(s) in the Visualization Window and provides information such as residue position, residue label, element, covalent radius, hydrogen bonding, partial charge, hybridization etc.
- 5. Console Window** - Displays information, warnings and errors. The input field at the bottom of the console window allows the user to enter console commands..

Courtesy: Molegro ApS, Aarhus, Denmark



Bio Servers

Primer Fox

PrimerFox is a free tool for primer generation. It works very simple in an easy to use fashion for fast generation of primer. When we entered a sequence, we get a primer pair which is optimized for the optimal PCR product. PrimerFox can be accessed freely online at <http://www.primerfox.com/>

The PrimerFox output displays in 3 different result views i.e. Sequence view, List view and Table view.

The Sequence view shows the best fitting primer pair according to the optimal PCR product length we entered & highlights the primer of the sequence .

The List view shows all forward and reverse primer as a list. It also listed the positions of the primer within our sequence corresponding to the starting point. This view is very useful to transfer data into other applications such as MS Excel.

The Table view provides a comparison of all primer combinations in the sequence depending on the given parameters. In each column there is a forward primer & in each row a reverse primer.

Courtesy: <http://www.primerfox.com/>

Current Trends

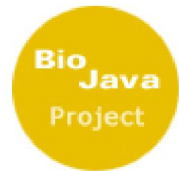
RNA interference

A RNA interference (RNAi) is a mechanism that inhibits gene expression at the stage of translation or by hindering the transcription of specific genes. RNAi targets include RNA from viruses and transposons which is significant for some forms of innate immune response, and also plays a role in regulating development and genome maintenance. Small interfering RNA strands (siRNA) are key to the RNAi process, and have complementary nucleotide sequences to the targeted RNA strand. Specific RNAi pathway proteins are guided by the siRNA to the targeted mRNA, where they "cleave" the target, breaking it down into smaller portions that can no longer be translated into protein. A type of RNA transcribed from the genome itself, microRNA, works in the same way. The RNAi pathway is initiated by the enzyme dicer, which cleaves long, dsRNA molecules into short fragments of 20–25 base pairs. One of the 2 strands of each fragment, known as the guide strand, is then incorporated into the RNA-induced silencing complex (RISC) and pairs with complementary sequences. The most well-studied outcome of this recognition event is post-transcriptional gene silencing. This occurs when the guide strand specifically pairs with an mRNA molecule and induces cleavage by argonaute, the catalytic component of the RISC complex. Another outcome is epigenetic changes to a gene – histone modification and DNA methylation – affecting the degree the gene is transcribed. The selective and robust effect of RNAi on gene expression makes it a valuable research tool, both in cell culture and in living organisms because synthetic dsRNA introduced into cells can induce suppression of specific genes of interest. RNAi may also be used for large-scale screens that systematically shut down each gene in the cell, which can help identify the components necessary for a particular cellular process or an event such as cell division. Exploitation of the pathway is also a promising tool in biotechnology and medicine.

Courtesy: Wikimedia Foundation, Inc, USA

Bioinfy Quiz - 003

1. Which eukaryotic RNA polymerase makes tRNA's?
 - a) RNA polymerase I
 - b) RNA polymerase II
 - c) RNA polymerase III
2. What is the average time it take to Mycobacterium tuberculosis to divide?
 - a) 20 minutes
 - b) 2 hours
 - c) 24 hours
3. Jaundice is due to an accumulation of which heme degradation product?
 - a) Porphobilinogen
 - b) Bilirubin
 - c) Biliverdin
4. Klinefelter syndrome is a manifestation of which chromosome aneuploidy?
 - a) XXX
 - b) XYY
 - c) XXY
5. Which of the compounds in an intermediate in the biosynthesis of Epinephrine?
 - a) Dihydroxymandelic acid
 - b) Dopamine
 - c) Metanephrine



BioJava is an open-source project dedicated to providing a Java framework for processing biological data. It includes objects for manipulating biological sequences, file parsers, DAS client and server support, access to BioSQL and Ensembl databases, tools for making sequence analysis GUIs and powerful analysis and statistical routines including a dynamic programming toolkit.

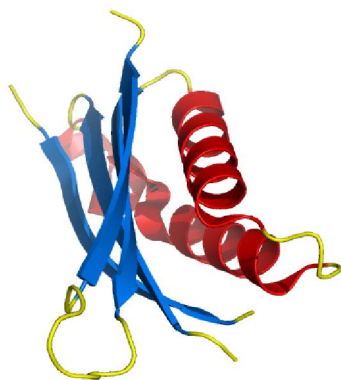
The BioJava library is useful for automating many daily and mundane bioinformatics tasks. As the library matures, the BioJava libraries will provide a foundation upon which both free software and commercial packages can be developed. The latest release is BioJava 1.6. and runs on Java 2 Standard Edition 1.5 platforms. A complete BioJava download is available at <http://www.biojava.org/>.

The latest release contains all binaries, required jars, docs, source, test, demos and apps as gzipped tar files [Courtesy: www.biojava.org/](http://www.biojava.org/)

Molecule of the Month

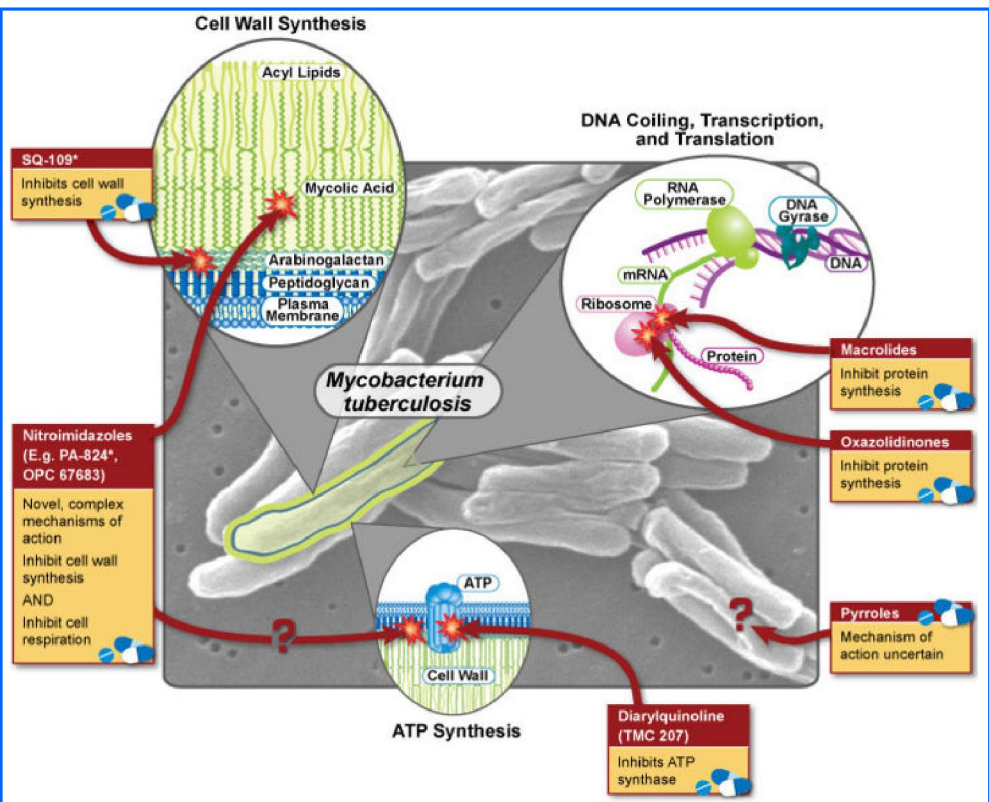
Apo-7,8-Dihydroneopterin Aldolase (Mycobacterium tuberculosis)

Mycobacterium tuberculosis 7,8-dihydroneopterin aldolase (Mtb FolB, DHNA) is the second enzyme in the folate biosynthetic pathway, which catalyzes the conversion of 7,8-dihydroneopterin to 6-hydroxymethyl-7,8-dihydropterin and glycoaldehyde. Folate derivatives are essential cofactors in the biosynthesis of purines, pyrimidines & amino acids across all forms of life. Mammals uptake folate from their diets, whereas most bacteria must synthesize folate de novo. Therefore, the enzymes in the folate biosynthetic pathway are attractive drug targets against bacterial pathogens such as Mycobacterium tuberculosis.



Molecular Data

PDB ID : 1Z9W
 Amino acids : 822
 Atoms : 1320
 Exp. Method : X-Ray
 Chains : A (1)
 Deposition : 2005-04-05
 E.C. no. : 4.1.2.25



Bioinfy Animator - New T.B. Drugs Under Development .

Courtesy: National Institute of Allergy & Infectious Diseases, USA

Nuclei Acid Sequence Single Letter Representation

Nuclei Acid Code	Meaning	S	GC (Strong interacion)
A	Adenosine	W	AT (Weak interaction)
C	Cytosine	B	GTC
G	Guanine	D	GAT
T	Thymidine	H	ACT
U	Uracil	V	GCA
R	GA (puRine)	N	AGCT (aNy)
Y	TC (pYrimidine)	X	masked
K	GT (Ketone)	-	gap
M	AC (aMino group)	Answer to Bioify Quiz - 03	1-c ; 2-c ; 3-b ; 4-c; 5-b

Please contribute to this bulletin, please contact:

Salam Pradeep Singh ; CSIR Diamond Jubilee Research Intern; Email: salampradeep@gmail.com; C/o Dr. R.L. Bezbaruah, Scientist E-II ; Biotechnology Division, NEIST, Jorhat 785006, Assam.