

Bioinformation up to Date

(Bioinformatics Infrastructure Facility, Biotechnology Division)
North-East Institute of Science & Technology
 Jorhat - 785006, Assam

Contents	Cover Story																						
<table border="1"> <tr> <td style="width: 80%;">Cover Story</td> <td style="text-align: right;">1</td> </tr> <tr> <td>Computational Chemistry</td> <td style="text-align: right;">2</td> </tr> <tr> <td>Proteomics</td> <td style="text-align: right;">2</td> </tr> <tr> <td>Software Mania</td> <td style="text-align: right;">3</td> </tr> <tr> <td>Bio Server</td> <td style="text-align: right;">3</td> </tr> <tr> <td>Bioinfo Quiz</td> <td style="text-align: right;">3</td> </tr> <tr> <td>Genomics</td> <td style="text-align: right;">4</td> </tr> <tr> <td>Molecule of the Month</td> <td style="text-align: right;">4</td> </tr> <tr> <td>Bioinfo Animator</td> <td style="text-align: right;">4</td> </tr> <tr> <td>Answer to Quiz</td> <td style="text-align: right;">4</td> </tr> <tr> <td>Contact Us</td> <td style="text-align: right;">4</td> </tr> </table>	Cover Story	1	Computational Chemistry	2	Proteomics	2	Software Mania	3	Bio Server	3	Bioinfo Quiz	3	Genomics	4	Molecule of the Month	4	Bioinfo Animator	4	Answer to Quiz	4	Contact Us	4	<p>Chimera</p> <p>A chimera is a spurious gene sequence derived from more than one DNA template. Chimeras form during PCR amplification, where sequence synthesis starts from one template, is interrupted for example, through damage or degradation of the template, and then continues from another template sharing a degree of localised homology to the original. In this way, a chimera can form from two or more phylogenetically distinct gene fragments, derived from different parental templates. Clearly, chimera formation has the potential to occur when there is more than one DNA template available within the PCR mixture; as is the case when PCR amplification is performed on pure DNA samples contaminated with foreign genetic material, or, as is increasingly common, on DNA extracted from environmental or medical samples.</p> <p>Chimera formation is the most frequent 16S rDNA artifact currently displayed by new submissions to the public repositories. Chimeras tend also to be the most insidious artifacts, since if undetected, they can be responsible for spurious phylogenetic reconstructions, inaccurate taxonomic identifications and over-estimations of microbial diversity.</p> <p>Chimera formation becomes more likely, the greater the amount of the homology that exists between the different DNA templates present in the PCR mixture. All 16S rDNA sequences share a degree of homology due to the highly conserved nature of the 16S rRNA gene; so much so, chimeras straddling different phyla can be found within clone libraries with surprising frequency. Chimeras are even more likely to form when different templates within the PCR mixture derive from the same phylum, and the likelihood of chimera formation increases as evolutionary distance between templates decreases. Unfortunately, this also means that the most common chimeras will be those formed from the most closely related species and thus the most difficult to detect.</p> <p>Some researchers were able to show that up to 30% of clones within a clone library of 16S rRNA will be chimeric - depending on methodology used. Factors known to increase the likelihood of chimera formation include the number of PCR cycles used, the stringency of the reaction conditions, and the harshness of the DNA extraction method employed. Chimeras have been recognised as a problem for almost as long as PCR amplification has been used, and in that time a number of methods for detecting chimeras have been developed, including Chimera_Check, Bellerophon and ChimeraBuster. Recently, the programs Pintail and Mallard have been made available.</p> <p>Finally it should be noted that, whilst chimera formation discussed here is clearly an artificial recombination event, there is some evidence to suggest that natural recombination between 16S rDNA, presumably as a consequence of horizontal gene transfer, does occur, leading to the formation of naturally occurring chimera-like sequences, though this would appear to be a rare event and it is unclear whether such genes are functional.</p> <p>A few chimera detection programs are available freely</p> <ol style="list-style-type: none"> 1) Pintail - a tool for screening single sequences for chimeras and other anomalies. 2) Mallard - a tool for screening multiple sequence alignments for chimeras and other anomalies.
Cover Story	1																						
Computational Chemistry	2																						
Proteomics	2																						
Software Mania	3																						
Bio Server	3																						
Bioinfo Quiz	3																						
Genomics	4																						
Molecule of the Month	4																						
Bioinfo Animator	4																						
Answer to Quiz	4																						
Contact Us	4																						
<p>Adviser: Dr. P.G. Rao</p> <p>Editors: Salam Pradeep Singh Dr. R.L. Bezbaruah</p>																							
<p>Upcoming Events</p> <ol style="list-style-type: none"> 1. “Bangalore India Bio: India’s Biggest Biotech Show” @ Lalit Ashok Bangalore from 02-04 June, 2010 2. International Conference on “Recent Advances in Bioinformatics” organized by KIIT School of Biotechnology from 3rd-5th Sept 2010 																							

Computational Chemistry

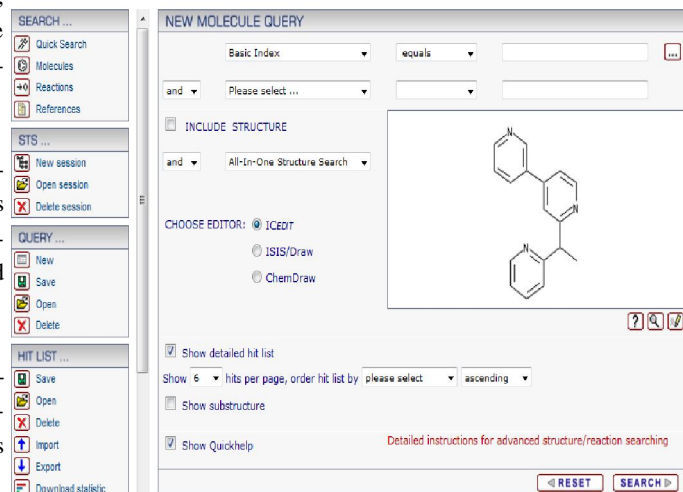
SPRESI Database

The SPRESI data collection is one of the largest databases for organic chemistry worldwide. The database covers the scientific literature from 1974 to the present, focusing on organic synthesis. It contains information on 7 millions chemical structures and 3.9 millions chemical reactions abstracted from 645,000 references.

The database was jointly built from VINITI (All-Russian Institute of Scientific and Technical Information of the Russian Academy of Sciences, based in Moscow) and ZIC (Zentrale Informationsverarbeitung Chemie, based in east Berlin, up to 1989) and the data are now maintained by the VINITI Institute. Since 1990 InfoChem GmbH, based in Munich, Germany, has been the distributor of this data collection.

The SPRESI database contains information on organic substances, including coverage of reactions, structures and properties. Over 32 millions of factual data, such as physical properties (boiling/melting points, refractive indexes, etc.), reaction conditions (catalysts, yields, etc.) and keywords have also been abstracted.

The SPRESI data collection can be accessed online via the web-application SPRESIweb, developed and distributed by InfoChem. Alternatively the complete set or subsets of the database can be acquired as raw data in SDF/RDF chemical file format.



Proteomics

Surface-enhanced laser desorption/ionization- Time of flight Mass Spectrometry

In mass spectrometry Surface-enhanced laser desorption/ionization (SELDI) is an ionization method that is used for the analysis of protein mixtures. Typically SELDI is used with time-of-flight mass spectrometers and is used to detect proteins in tissue samples, blood, urine, or other clinical samples.

It was developed by T. William Hutchens at Baylor College of Medicine in 1993. In 1997 the technology was commercialized by Ciphergen Biosystems as the ProteinChip system. SELDI technology now is produced and marketed by Bio-Rad Laboratories.

SELDI-TOF-MS is a variation of matrix-assisted laser desorption/ionization (MALDI) that uses a target modified to achieve biochemical affinity with the analyte compound.

SELDI-TOF MS combines two powerful techniques: chromatography and mass spectrometry and it is a novel approach to biomarker discovery. SELDI-TOF MS has the ability to provide a rapid protein expression profile from a variety of biological and clinical samples. It has been used for biomarker identification as well as the study of protein-protein, and protein-DNA interaction. Because of its versatility it has been used in wide range of projects ranging from the identification of potential diagnostic markers for prostate, bladder, breast, and ovarian cancers, interstitial cystitis, Alzheimer's disease, to the study of biomolecular interactions and the characterization of post-translational modifications. SELDI technique has also been used to characterize phosphorylated and glycosylated proteins, transcription factors, and peptides and proteins shed or secreted by various cancer cell lines.

SELDI-TOF MS has three major components which constitute the instrumentations:

(1) Protein chip arrays, (2) A mass analyzer and (3) Data analysis software.

Biological Application

The surfaces of proteins chips can be covalently modified with bait molecules, such as antibodies or recombinant proteins, to capture proteins that have an affinity to the bait. If the affinity reagent being used as bait recognizes an epitope in different isoforms of the protein, then these proteins can be separated by their molecular mass, resulting in multiple discrete signals. The direct sampling has a unique advantage over the traditional enzyme-linked immunosorbent assay, in which the resultant signal is a weighted average of all of the bound species. Post-translationally modified proteins can also be detected using SELDI-TOF MS. Another important application is studying phosphorylation states of proteins involved in signal transduction pathways.

.....Lipee Boruah, Dibrugarh University

Software Mania

GenDoc

GeneDoc is a powerful editor which allows you to manually edit and add a wide range of attributes to multi-sequence alignments generated by Clustal W.

Some important features of GenDoc:

Alignment view: In this view, we will be able to see the properties of each residue in the alignment with the various shading modes, also able to arrange the residues to improve our alignment. We will be able to score various sections on the alignment and add manual comments and manual shading.

Summary view: GeneDoc will also apply shading to the Summary view. Users can control the summary view settings in the Project Configuration dialog. GeneDoc can compress the Summary View up to one dot per column, so on a high resolution printer the amount of compression can be considerable.

Tree view: GeneDoc provides a phylogenetic tree view. This view is can be used to construct the phylogenetic tree.

Plot view: GeneDoc displays a few Cumulative Distribution plots in the Plot view. At its simplest, the plot view can be used to show Percent Identity or Percent Favorable Substitutions for the alignment as a whole. Using the Super Family Groups of GeneDoc, the plot can be used to show that the scoring within groups is significantly better than scoring between different groups, thus demonstrating that our super family groups are well chosen. These functions are controlled through the Plot menu.

Gel view: GeneDoc displays a simple Enzyme Gel Simulation view. This view contains two lists, the list of sequences and the list of loaded enzymes. Select multiple sequences and multiple enzymes and click the 'Run Gel' button. The resulting sequence fragments are plotted on the view on a log scale.

.....Runuma, Dibrugh University

Bio Servers

GlycoMod Tool

GlycoMod is a program designed to find all possible compositions of a glycan structure from its experimentally determined mass. This is done by comparing the mass of the glycan to a list of pre-computed masses of glycan compositions. The program can be used with free or derivatised glycans and for glycopeptides where the peptide mass or protein is known. Compositional constraints can be applied to the output. The program can be accessed at the url <http://expasy.org/tools/glycomod/>. The input parameters are:

1. Experimental masses

The user may enter the experimental masses to be analyzed, separating them by spaces or new lines. It is also possible to enter the masses from a text file provided each mass is on a new line. These mass values may be average or monoisotopic, but the user must select the appropriate button, and the mass values must all be in agreement. A mass tolerance level should be selected in either Daltons or ppm. Note that the higher the mass tolerance, the greater the number of compositions returned.

The experimental masses may correspond to glycopeptides or free oligosaccharides, which may be derivatised.

2. Ion mode and adducts

The user may enter the masses as neutral ions, positive ions, or as negative ions such as [M], [M+H].

3. Glycan form

GlycoMod can calculate the possible compositions of N-linked oligosaccharides, linked via the amide nitrogen of an asparagine residue, or O-linked oligosaccharides, linked via the hydroxyl group of serine or threonine. Oligosaccharides may also be O-glycosidically linked via the hydroxyl group of hydroxylsine, hydroxyproline and tyrosine. These amino acid linkages are less common and are not considered in this version of GlycoMod..

GlycoMod can calculate the composition of the glycans from the masses of glycopeptides or of glycans released from the peptide moiety by enzymatic or chemical means.

4. Monosaccharide residues

GlycoMod is designed to calculate the masses of oligosaccharides using underivatised, permethylated or peracetylated monosaccharides since mass spectrometric data is often obtained from these later derivatised oligosaccharides. The user may stipulate which monosaccharides are/are not/or may possibly be present in the glycan.

Bioinfy Quiz - 022

1. What is the name of the database on 7-transmembrane G-linked receptors?

- A) 7TM
- B) GCRDb
- C) GLR7

2. Which gene is responsible for X chromosome inactivation?

- A) Hprt
- B) Xip
- C) Xist

3. At which wavelength are DNA/RNA concentrations measured?

- A) 240 nm
- B) 260 nm
- C) 300 nm

4. In which of these organisms does UAA and UAG code for Gln instead of Stop?

- A) Belpharisma japonicum
- B) Paramecium caudatum
- C) Paramecium primaurella

5. The number of nucleotides on each chain turn of the DNA double helix is?

- A) 8
- B) 10
- C) 12

Answers on Page 4

Genomics:

Mouse Genome Database (MGD) Project

Mouse Genome Database (MGD) Project is a project contributed by Mouse Genome Informatics (MGI). MGI is the international database resource for the laboratory mouse, providing integrated genetic, genomic, and biological data to facilitate the study of human health and disease.

MGD includes information on mouse genetic markers, molecular clones (probes, primers, and YACs), phenotypes, sequences, comparative mapping data, graphical displays of linkage, cytogenetic and physical maps, experimental mapping data, as well as strain distribution patterns for recombinant inbred strains (RIs) and cross haplotypes. Sources for information in the database include the scientific literature, DNA mapping panel data produced by various researchers, the MIT Whitehead Institute (MIT genetic and physical mapping data), WashU, and the IMAGE consortium. MGD is closely linked with GXD's Gene Expression Literature index. This is a searchable index of references to the scientific literature reporting data on endogenous gene expression during mouse embryonic development. MGD provides links to Gene Expression Literature content records through genetic marker detail records, and in turn, the content records provide links to MGD references and genetic marker records. Since it first became available on the web, MGD has continued to evolve, expanding its data coverage, improving data handling, and providing new data manipulation and display tools. MGD contains the following kinds of information:

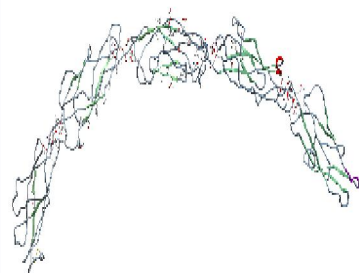
- Gene, DNA marker, QTL and Cytogenetic marker descriptions, Mouse genetic phenotypes, genetic interrelationships, and polymorphic loci
- Polymorphic loci related to specified strains, SNPs and other sequence polymorphisms, Mammalian orthology data, Sequence data,
- Molecular probes and clones (probes, clones, primers and YACs)' Genetic and physical mapping data, Human disease data (OMIM)
- Information on inbred strains (M. Festing's listing)

....Dhrubajyoti, Dibrugarh University

Molecule of the Month

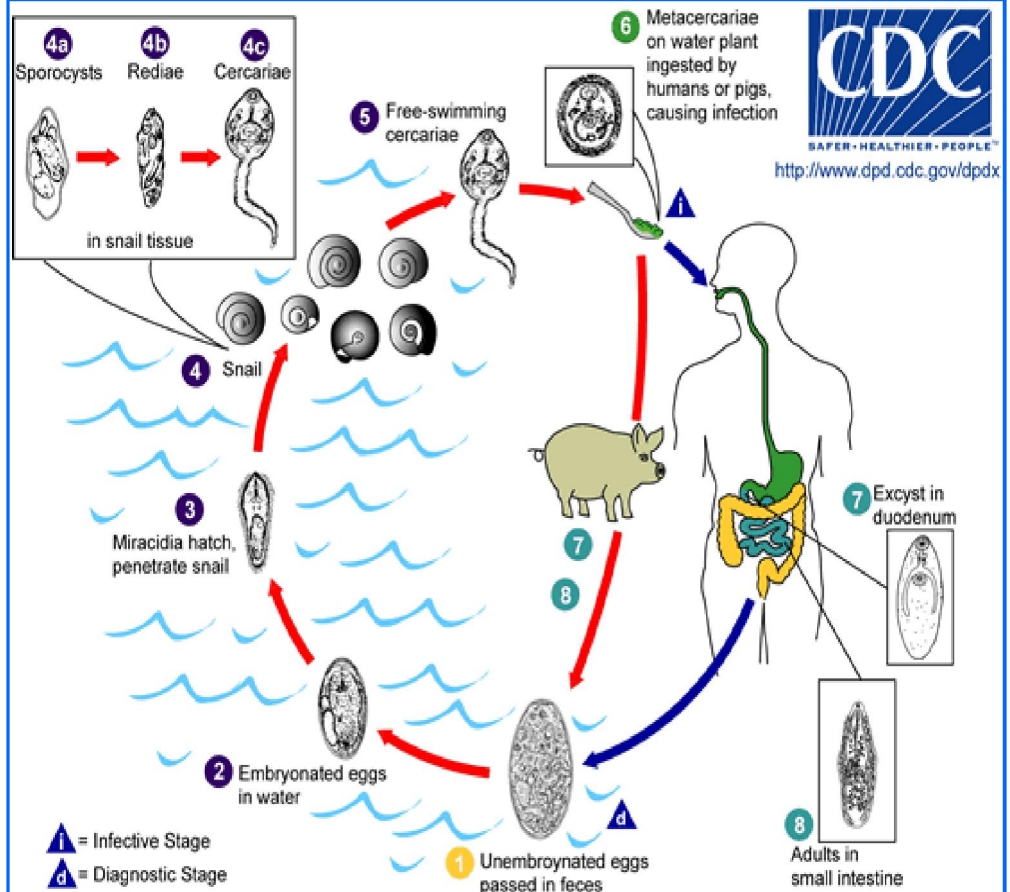
Cadherins

Cadherins are one of the many molecules that glue cells together. They are long proteins that extend from the surface of the cell. The outer portion of the protein is composed of a series of folded domains arranged one after the next, and calcium ions bind between each domain, rigidifying the whole structure. If calcium is removed, however, the chain becomes floppy and is easily destroyed by protein-cutting enzymes. The tip of the chain has a special tryptophan amino acid, colored red here, that binds to cadherins on neighboring cells, adhering the two cells together. There are many different types of cadherins, made by different types of cells. Each one is slightly different and sticks only to its proper partners. The one shown here is made by epidermal cells. These different cadherins play an important role in the development of embryos. They allow the growing cells to find their proper tissues and fix themselves in place.



Molecular Data

PDB ID : 3L3W
 Amino acids : 540
 Exp. Method : X-Ray Diff
 Chains : A



Bioinfy Animator:- Life Cycle of Fasciolopsis

Fasciolopsis is a genus of trematodes. It consists only of two species, *F. buski* and *F. magna*, of which, *F. buski*, is more notable in terms of prevalence and pathogenicity as it causes the disease fasciolopsiasis. *Fasciolopsis buski* is commonly called the giant intestinal fluke, being the largest known parasitic fluke in humans. The body can be up to 7.5 cm in length and 2.5 cm in width. It is a common parasite of humans and pigs and is most prevalent in Asia, mainly central and southeast Asia. It belongs to the class Trematoda, family Fasciolidae. The worm inhabits the upper region of the small intestine and, when abundant, can also be found in the lower areas of the intestine and the stomach. *Fasciolopsis buski* is the cause of the disease fasciolopsiasis.

.....Sanchaita Rajkhowa, Center for Bioinformatics Studies, Dibrugarh University

For suggestions & contributions contact:

Salam Pradeep; Project Assistant Level-II
 Email: salampradeep@gmail.com.

Bioinfy Quiz

022

Answers

1 - B ; 2 - C ; 3 - B ; 4 - C ; 5 - B