



Bioinformation up to Date

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

Contents

Cover Story	1
Computational Chemistry	2
Genomics	2
Software Mania	3
Bio Server	3
Bioinfo Quiz	3
Proteomics	4
New Year Greetings	4
Contact Us	4

Adviser:

Dr. P.G. Rao

Editors:

Salam Pradeep Singh

Dr. R.L. Bezbaruah

Upcoming Events

1. Short Term Training on "Bioinformatics: Recent Trends in Biological sequence analysis" from 27th – 29th January, 2010 @ BIF, Gauhati University, Guwahati.
2. "XXI Annual BTISNET Coordinators Meeting" on 3rd & 4th Feb, 2010 @ BIF, Central Agricultural Research Institute, Port Blair, Andaman & Nicobar Islands.

Cover Story

Learning Management System

A learning management system (LMS) is a software application for the administration, documentation, tracking, and reporting of training programs, classroom and online events, e-learning programs, and training content. (Ellis 2009)

LMSs range from systems for managing training and educational records, to software for distributing courses over the Internet with features for online collaboration. Corporate training use LMSs to automate record-keeping and employee registration. Student self-service (e.g., self-registration on instructor-led training), training workflow (e.g., user notification, manager approval, wait-list management), the provision of on-line learning (e.g., Computer-Based Training, read & understand), on-line assessment, management of continuous professional education (CPE), collaborative learning (e.g., application sharing, discussion threads), and training resource management (e.g., instructors, facilities, equipment), are dimensions to Learning Management Systems.

Some LMSs are Web-based to facilitate access to learning content and administration. LMSs are used by regulated industries (e.g. financial services and biopharma) for compliance training. It is also used by educational institutions to enhance and support classroom teaching and offering courses to a larger population of learners across the globe.

Characteristics

LMSs cater to educational, administrative, and deployment requirements. While an LMS for corporate learning, for example, may share many characteristics with a VLE, or virtual learning environment, used by educational institutions, they each meet unique needs. The virtual learning environment used by universities and colleges allow instructors to manage their courses and exchange information with students for a course that in most cases will last several weeks and will meet several times during those weeks. In the corporate setting a course may be much shorter, completed in a single instructor-led or online session.

The characteristics shared by both types of LMSs include:

- Manage users, roles, courses, instructors, facilities, and generate reports
- Course calendar, Learning Path, Student messaging and notifications
- Assessment and testing handling before and after testing
- Display scores and transcripts, Grading of coursework and roster processing, including waitlisting
- Web-based or blended course delivery

Characteristics more specific to corporate learning, which sometimes includes franchisees or other business partners, include:

- Auto enrollment (enrolling Students in courses when required according to predefined criteria, such as job title or work location)
- Manager enrollment and approval
- Boolean definitions for prerequisites or equivalencies
- Integration with performance tracking and management systems
- Planning tools to identify skill gaps at departmental and individual level
- Curriculum, required and elective training requirements at an individual and organizational level
- Grouping students according to demographic units (geographic region, product line, business size).
- Assign corporate and partner employees to more than one job title at more than one demographic unit

Computational Chemistry

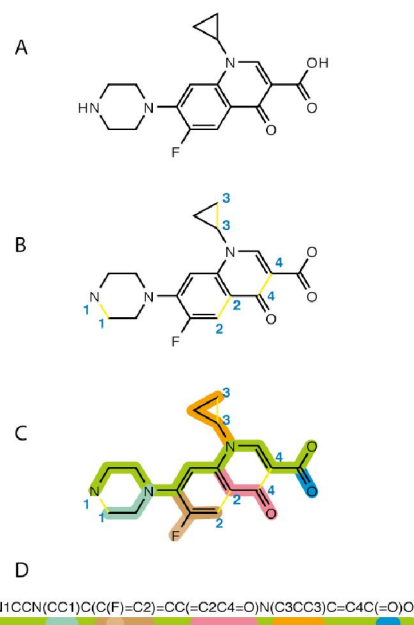
Simplified Molecular Input Line Entry Specification

The simplified molecular input line entry specification or SMILES is a specification for unambiguously describing the structure of chemical molecules using short ASCII strings. SMILES strings can be imported by most molecule editors for conversion back into two-dimensional drawings or three-dimensional models of the molecules.

The original SMILES specification was developed by Arthur Weininger and David Weininger in the late 1980s. It has since been modified and extended by others, most notably by Daylight Chemical Information Systems Inc. In 2007, an open standard called "OpenSMILES" was developed by the Blue Obelisk open-source chemistry community. Other 'linear' notations include the Wiswesser Line Notation (WLN), ROSDAL and SLN (Tripos Inc).

In August 2006, the IUPAC introduced the InChI as a standard for formula representation. SMILES is generally considered to have the advantage of being slightly more human-readable than InChI; it also has a wide base of software support with extensive theoretical backing. The term SMILES refers to a line notation for encoding molecular structures and specific instances should strictly be called SMILES strings. However, the term SMILES is also commonly used to refer to both a single SMILES string and a number of SMILES strings; the exact meaning is usually apparent from the context. The terms Canonical and Isomeric can lead to some confusion when applied to SMILES. The terms describe different attributes of SMILES strings and are not mutually exclusive.

Typically, a number of equally valid SMILES can be written for a molecule. Algorithms have been developed to ensure the same SMILES is generated for a molecule regardless of the order of atoms in the structure. The SMILES is unique for each structure, although dependent on the canonicalisation algorithm used to generate it, and is termed the Canonical SMILES. These algorithms first convert the SMILES to an internal representation of the molecular structure and do not simply manipulate strings as is sometimes thought. Various algorithms for generating Canonical SMILES have been developed, including those by Daylight Chemical Information Systems, OpenEye Scientific Software, MEDIT and Chemical Computing Group. A common application of Canonical SMILES is indexing and ensuring uniqueness of molecules in a database. SMILES notation allows the specification of configuration at tetrahedral centers, and double bond geometry. These are structural features that cannot be specified by connectivity alone and SMILES which encode this information are termed Isomeric SMILES. A notable feature of these rules is that they allow rigorous partial specification of chirality. The term Isomeric SMILES is also applied to SMILES in which isotopes are specified.



Genomics

GENSAT

The GENSAT project aims to map the expression of genes in the central nervous system of the mouse, using both in situ hybridization and transgenic mouse techniques.

The GENSAT database contains a gene expression atlas of the central nervous system of the mouse based on bacterial artificial chromosomes (BACs). In each of the BAC transgenic vectors, endogenous protein coding sequences have been replaced by sequences encoding the EGFP reporter gene. As in any gene replacement experiment, the stability of the reporter gene can vary somewhat from the endogenous gene. Thus these results measure the relative rates of transcription for each gene; they are not a direct measure of mRNA accumulation or of protein abundance for the endogenous gene products. Furthermore, the enhanced sensitivity of reporter gene assays, particularly in BAC lines carrying multiple copies of the BAC transgene, may allow detection of sites of expression that are not evident in situ hybridization experiments.

Gene Name	<input type="text"/>
Gene Symbol	<input type="text"/>
Age	Embryonic day 10.5 <input type="button" value="v"/>
Image Format	BAC confocal <input type="button" value="v"/>
Image Section	sagittal <input type="button" value="v"/>
Expressed in	Cerebral cortex <input type="button" value="v"/>
Localized to	glial cell, CA1 pyramidal cell <input type="button" value="v"/>
Expression Level	weak signal <input type="button" value="v"/>
Display Type <input type="button" value="Gene List"/> <input type="button" value="Search!"/>	

This database contains histological data from given BAC transgenic mouse lines at three developmental stages - embryonic day 15.5 (E15.5), postnatal day 7 (P7) and adult; in all cases the data represent results of multiple transgenic lines. EGFP is visualized by staining with an anti-EGFP antibody using the DAB method, or by confocal microscopy of unstained tissue sections.

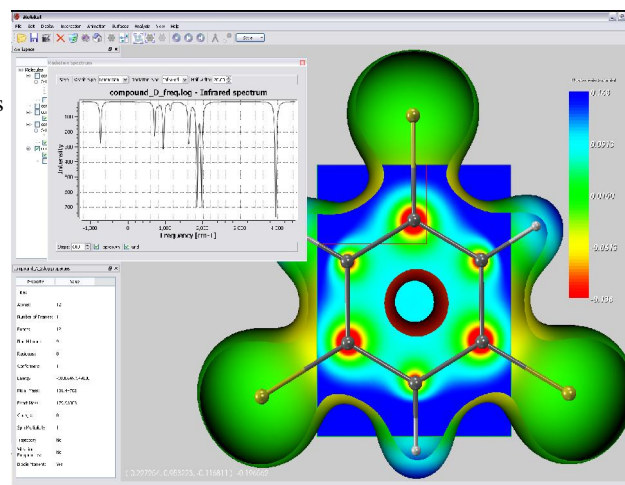
Software Mania

Molkel

Molekel is a free software multiplatform molecular visualization program. It was originally developed at the University of Geneva by Peter F. Flükiger in the 1990s for Silicon Graphics Computers. In 1998, Stefan Portmann took over responsibility and released Version 3.0. Version 4.0 was a nearly platform independent version. Further developments lead to version 4.3, before Stefan Portman moved on and ceased to develop the codes. In 2006, the Swiss National Supercomputing Centre (CSCS) restarted the project and version 5.0 was released on 21 December 2006.

Major features

- Visualization of residues (ribbon or schematic)
- Complete control over the generation of molecular surfaces (bounding box and resolution)
- Visualization of the following surfaces:
 - Orbitals
 - Isosurface from electron density data
 - Isosurface from Gaussian cube grid data
 - Solvent-accessible surface (SAS)
 - Solvent excluded surface (SES)
 - Van der Waals radii
- Animation of molecular surfaces
- Export to PostScript or TIFF



Bio Servers

3D JIGSAW

3D-JIGSAW is an automated system to build three-dimensional models for proteins based on homologues of known structure. The server comprises of two programs

a) Automatic Mode & b) Interactive Mode

a) Automatic Mode: The program looks for homologous templates in our sequence databases (PFAM+PDB+nr) and splits the query sequence into domains. If good templates are found, the best covered domain is then modelled using a maximum of 2. This process can take up to an hour, depending on the load of the system. You will receive an e-mail with the alignment between query and template/s and a PDB formatted set of coordinates, that you can display easily with Rasmol.

b) Interactive Mode: The program looks for homologous templates in our sequence databases (PFAM+PDB+nr) and splits the query sequence into domains. An e-mail is sent back to you with a link to a graphical display of this domain arrangement and useful information extracted from the PFAM database. From this link you can choose the domains you need to model and you can select the templates and correct the alignments before submitting a modelling job. Templates are ranked according to the coverage of the query, their sequence identity and their crystallographic resolution. Information from each template is easily accessed, including its alignment to the query sequence.

Warning: You must provide a valid E-mail address to retrieve the results of your query.

Your name
 Your E-Mail Address
 Your E-Mail Address (verification)

Protein identifier Automatic Interactive!
 Split your sequence into domains, choose the modelling templates and edit the alignments

3D-JIGSAW

Protein amino acid sequence in one letter code

```
>Unknown
MTHNIPWRTYEWFGHWIACVGHNI PNIPWRTGHINI PWRTACVNI PNWRWINGHI
ACVNI PFWRTGHIACVNI PWRNTGHI PWRNTGHIACVNI GHPNI PNWRNTNWRIGH
IANIPWNINPRTCVNWINNIWPPWRT]
```

Bioinfy Quiz - 018

1. Which of those three termination codons is called 'amber'?

- UAA
- UAG
- UGA

2. What is approximately the mass of a kilobase of double stranded DNA?

- 315 kdal
- 480 kdal
- 660 kdal

3. What is the approximate size (in bp) of a mammalian mitochondrial genome?

- 14'000 bases
- 16'000 bases
- 18'000 bases

4. Which disease is not caused by a member of the Poxvirus family?

- Smallpox
- Chickenpox
- Cowpox

5. Which one of those proteins is a membranal protein?

- Tropomyosin A
- Flavodoxin
- Glycophorin

Answers on Page 4

Proteomics:

ENZYME

The ENZYME database is a repository of information relative to the nomenclature of enzymes. It is primarily based on the recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (IUBMB), and it contains the following data for each type of characterized enzyme for which an EC (Enzyme Commission) number has been provided:

- EC number
- Recommended name
- Alternative names (if any)
- Catalytic activity
- Cofactors (if any)
- Pointers to the Swiss-Prot protein sequence entrie(s) that correspond to the enzyme (if any)
- Pointers to human disease(s) associated with a deficiency of the enzyme (if any)

The ENZYME database can be useful to anybody working with enzymes and that it can be of help in the development of computer programs involved with the manipulation of metabolic pathways.

NiceZyme View of ENZYME: EC 2.3.1.43

Accepted Name	
Phosphatidylcholine--sterol O-acyltransferase.	
Alternative Name(s)	
LCAT. Lecithin--cholesterol acyltransferase. Phospholipid--cholesterol acyltransferase.	
Reaction catalysed	
Phosphatidylcholine + a sterol <=> 1-acylglycerophosphocholine + a sterol ester	
Comment(s)	
<ul style="list-style-type: none">• Palmitoyl, oleoyl, and linoleoyl can be transferred; a number of sterols, including cholesterol, can act as acceptors.• The bacterial enzyme also catalyzes the reactions of EC 3.1.1.4 and EC 3.1.1.5.	
Cross-references	
PROSITE	PDOC00110
BRENDA	2.3.1.43
EC2PDB	2.3.1.43
PRIAM enzyme-specific profiles	2.3.1.43
KEGG Ligand Database for Enzyme Nomenclature	2.3.1.43
IUBMB Enzyme Nomenclature	2.3.1.43
IntEnz	2.3.1.43
MEDLINE	Find literature relating to 2.3.1.43
MetaCyc	2.3.1.43
UniProtKB/Swiss-Prot	P10480, GCAT_AERHY; P53760, LCAT_CHICK; O35573, LCAT_ELIIQU; P04180, LCAT_HUMAN; O35724, LCAT_MICMN; P16301, LCAT_MOUSE; O35502, LCAT_MYOGA; Q08758, LCAT_PAPAN; P30930, LCAT_PIG; P53761, LCAT_RABIT; P18424, LCAT_RAT; O35840, LCAT_TATKG;

**HAPPY
NEW YEAR
2010**

For suggestions & contributions contact:

Salam Pradeep Singh
CSIR D.J. Res. Intn; Email: salampradeep@gmail.com.

Bioinfy Quiz 017 Answers

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