



Bioinformation up to Date

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

Contents	Cover Story																				
<table style="width: 100%; border-collapse: collapse;"> <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;">1</td> </tr> <tr> <td>Genomics</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>Bioinfy Quiz</td> <td style="text-align: right;">3</td> </tr> <tr> <td>Proteomics</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>Bioinfy Animator</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	1	Genomics	2	Software Mania	3	Bio Server	3	Bioinfy Quiz	3	Proteomics	4	Molecule of the Month	4	Bioinfy Animator	4	Contact Us	4	<p style="text-align: center;">Proceedings of 2nd North-East Bioinformatics Network Coordinator's Meeting</p> <p>A meeting on North-East Bioinformatics Network (NEBInet) Coordinator's Meet was held on 12th & 13th November, 2009 at College of Veterinary Science, Assam Agricultural University, Khanapara, Guwahati. The meeting was sponsored by the Department of Biotechnology (DBT), Government of India, New Delhi. The meeting was organized by the Bioinformatics Centre, College of Veterinary Science, Assam Agricultural University, Khanapara, Guwahati.</p> <p>On the basis of the deliberations during the 2nd NEBInet Coordinator's Meeting, the following recommendations were made:</p> <ol style="list-style-type: none"> 1. The DBT may provide financial support to all NEBInet centres to have at least 2 mbps internet connectivity for their effective functioning. 2. An additional position of a Technical Assistant may be created under each of the centres (exclusively for BTISnet centres of the North-East) with adequate funding from DBT. 3. All the centers should take adequate steps to ensure maximum use of the facilities by students and faculty members of the host institute and other neighboring institutions. 4. The Utilization Certificate and Statement of Expenditure of the centers should be submitted within the stipulated date of 31st May every year to ensure early release of fund for the next financial year. 5. It was suggested to evolve network projects in Biotechnology and Bioinformatics by the NEBInet centers. 6. The computer servers provided to NEBInet centers to be used effectively. The centers are requested to maintain usage profile of the BIF resources. 7. The NEHU Bioinformatics centers would be the NEBInet nodal centre as recommended in the previous NEBInet Coordinator's meeting held in Gangtok, Sikkim. DBT has to issue a Notification in this regard. 8. The NEBInet nodal centre will access the software requirements for the NEBInet centres and propose funding from DBT for the same. 9. North East Region – DBT e-Library Consortium (NER-DeLCON) is now for free trials for the centers those were selected for this Consortium. Therefore, it was suggested to use these resources effectively and issue the merits and demerits of the resources before going in for actual subscription of these journals. 10. NEHU, TURA Campus to be networked as early as possible. 11. Based on the recommendations emerged out of 2nd NEBInet meeting, Dr. Borah shall make presentation in the next BTIS coordinators meeting to be held on 3rd and 4th February, 2010 in Port Blair, Andaman & Nicobar Islands. The NEBInet Coordinators will also be participating in this meeting to interact with other BTISnet Coordinators.
Cover Story	1																				
Computational Chemistry	1																				
Genomics	2																				
Software Mania	3																				
Bio Server	3																				
Bioinfy Quiz	3																				
Proteomics	4																				
Molecule of the Month	4																				
Bioinfy Animator	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. “National Symposium on Recent Developments & Trends in Computational Chemistry” @ Dep't of Chemistry, NEHU, Shillong from 12th-13th March, 2010. 2. National Seminar on “Shape and Electrostatics Best Practices for Virtual Screening and Lead-Hopping” @ School of Information Tech, JNU on 28th & 29th Jan 2010. 																					

Computational Chemistry

Austin Model 1 (AM1)

Austin Model 1, or AM1, is a semi-empirical method for the quantum calculation of molecular electronic structure in computational chemistry. It is based on the Neglect of Differential Diatomic Overlap integral approximation. Specifically, it is a generalization of the modified neglect of differential diatomic overlap approximation. Related methods are PM3 and the older MINDO.

AM1 was developed by Michael Dewar and co-workers and published in 1985. AM1 is an attempt to improve the MNDO model by reducing the repulsion of atoms at close separation distances. The atomic core-atomic core terms in the MNDO equations were modified through the addition of off-center attractive and repulsive Gaussian functions.

The complexity of the parameterization problem increased in AM1 as the number of parameters per atom increased from 7 in MNDO to 13-16 per atom in AM1.

The results of AM1 calculations are sometimes used as the starting points for parameterizations of forcefields in molecular modelling.

AM1 is implemented in the MOPAC, AMPAC, GAUSSIAN, CP2K, GAMESS (US), PC GAMESS and GAMESS (UK) programs.

An extension of AM1 is SemiChem Austin Model 1 (SAM1), which is implemented in the AMPAC program and which explicitly treats d-orbitals.

A model for the AM1 calculation of lanthanide complexes, called Sparkle/AM1, was also introduced and is implemented in MOPAC2007.

AM1 has been recently reparameterized, leading to the new RM1, or Recife Model 1, available in MOPAC2007, Spartan'06, Hyperchem 8, etc.

Genomics

The Genome Reference Consortium

At the time the human reference was initially described, it was clear that some regions were recalcitrant to closure with existing technology. What was not as clear was the degree to which structural variation affected our ability to produce a truly representative genome sequence at some loci. It is now apparent that some regions of the genome are sufficiently variable that they are best represented by multiple sequences in order to capture all of the sequence potentially available at these loci. In order to improve the representation of the reference human genome the Genome Reference Consortium (GRC) have formed. The goal of this group is to correct the small number of regions in the reference that are currently misrepresented, to close as many remaining gaps as possible and to produce alternative assemblies of structurally variant loci when necessary.

The GRC only maintains genomes that have been generated using a hierarchical (clone) based assembly method. Typically, these projects are considered complete in that most of the genome is well represented and typically the funding for the main genome project has come to an end. Currently, the only genomes supported are mouse and human.

The GRC focuses on the following:

- identifying and correcting assembly errors
- identifying regions of allelic complexity that require the addition of a partial assembly for that locus.
- working with the research community to address questions and concerns
- producing updated assemblies on a regular cycle

A set of TPF files are maintained for each assembled chromosome and partial assembly. These files are stored in a central database that manages TPF tracking and validation. A series of checks and analyses are performed and regions of the genome needing manual curation are identified and cataloged for further review.

Sequences (also known as components) which are adjacent on the TPF are expected to have a specific type of sequence alignment known as a full dovetail. A program call 'find_overlaps' assesses all adjacent component sequences to determine if they have an appropriate overlap. If an alignment can be identified, it is evaluated to ensure it meets our criteria for joining two sequences in a contig. GRC use this evaluation to prioritize overlaps that need further investigation. Alignments can fall into the following categories:

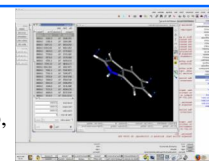
can fall into the following categories:

- Excellent alignment, meets all defined criteria.
- Minor alignment problem.
- Serious alignment problem requiring review.
- An alignment certificate providing external evidence for accepting the join has been submitted, but not approved.
- An alignment certificate has been approved for this join.

Software Mania

Gabedit

Gabedit is a graphical user interface to computational chemistry packages like Gamess-US, Gaussian, Molcas, Molpro, MPQC, OpenMopac, PCGameess and Q-Chem



It can display a variety of calculation results including support for most major molecular file formats. The advanced "Molecule Builder" allows to rapidly sketch in molecules and examine them in 3D. Graphics can be exported to various formats, including animations.

Major features

Gabedit can Create input file for GAMESS(US), GAUSSIAN, MOLCAS, MOLPRO, MPQC, OpenMopac, PCGameess and Q-Chem.

Gabedit can graphically display a variety of Gamess-US, Gaussian, Molcas, Molpro, MPQC, OpenMopac, PCGameess, Q-Chem and (partially) ADF calculation results, including the following

- Molecular orbitals
- Surfaces from the electron density, electrostatic potential, NMR shielding density, and other properties.
- Surfaces may be displayed in solid, translucent and wire mesh modes. they are can be colorcoded by a separate property.
- Contours (colorcoded), Planes colorcoded, Dipole. XYZ axes and the principal axes of the molecule.
- Animation of the normal modes corresponding to vibrational frequencies.
- Animation of the rotation of geometry, surfaces, contours, planes colorcoded, xyz and the principal axes of the molecule.
- Animation of contours, Animation of planes colorcoded.

Gabedit can display UV-Vis, IR and Raman computed spectra.

Gabedit can generate a povray file for geometry surfaces (including colorcoded surfaces), contours, planes colorcoded.

Gabedit can save picture in BMP, JPEG, PNG, PPM and PS format.

Gabedit can generate automatically a series of pictures for animation (vibration, geometry convergence, rotation,, planes colorcoded).

Bio Servers

Aldente - Peptide Mass Fingerprinting Tool

Aldente is a tool to identify proteins from peptide mass fingerprinting data. It is a new, fast and powerful tool takes advantage of the Hough transform for spectra recalibration and outlier exclusion. The Aldente search form can be used in two modes:

- a) for a global view of all the search parameters on one page
- b) to have search parameters grouped into smaller logical sections

Peptide mass fingerprinting involves the digestion of an unknown protein with a proteinase of known cleavage specificity and the measurement of the resulting peptides by mass spectrometry. Those experimentally measured peptide masses are then compared with the theoretical peptides calculated for all proteins in a protein sequence database. One of the main ideas of Aldente is to avoid calibration problems by taking into account the mass spectrometer deviation. It implements a number of rules, empirical observations and user knowledge that approach the expert human interpretation of results in various steps of the identification procedure.

The procedure can be divided into 3 steps:

- The selection of experimental MS peaks to be matched with theoretical peptide masses within a user defined tolerance space.
- The interpretation of the experimental calibration deviation of the instrument: exclusion of false matches and reduction of the tolerance space to the internal deviation of the instrument.
- The entirely tuneable scoring and ranking system of the protein entries that allow correct interpretation of real identifications, presence of protein mixtures, etc.

Aldente proposes a realignment of experimental masses using the Hough transform. The Hough Transform is standard method used in image analysis for finding straight lines hidden in larger amounts of other data. It is a robust method, therefore not sensitive to noise. It has no difficulties to work with very crowded spectra and with a lot of theoretical masses (that can be generated when considering the combinatory related with heterogeneously modified or missed-cleaved peptides).

The output provides a significant amount of useful information. It includes identified proteins and peptides, as well as the deviation function used, and the worked out ambiguities.

Bioinfy Quiz - 019

1. Who first discovered that DNA has genetic specificity?

- a) Oswald Avery
- b) Arthur Kornberg
- c) James Watson

2. What is the defective enzyme in the glycogen storage Von Gierke's disease?

- a) Phosphorylase
- b) Phosphofruktokinase
- c) Glucose-6-phosphatase

3. In C4 plants what molecule is produced from CO₂ fixation ?

- a) Malate
- b) Oxaloacetate
- c) Pyruvate

4. Which method is preferred to stain human chromosomes so as to show their bands?

- a) Allopuridol staining
- b) Modified Giemsa staining
- c) Quinacrine staining

5. Which of these cytokines is also known under the name cachectin?

- a) Interferon gamma
- b) Interleukin 2
- c) Tumor necrosis factor (TNF)

Answers on Page 4

Proteomics:

Protein Clusters: A Collection of Proteins Grouped by Sequence Similarity and Function

The NCBI Entrez Protein Clusters database is a collection of Reference Sequence (RefSeq) proteins from the complete genomes of prokaryotes, plasmids, viruses (including viruses), organelles, and complete and incomplete genomes of protozoa grouped and annotated based on sequence similarity and protein function (1). Proteins are automatically grouped into clusters based on reciprocal best-hit BLAST scores. The protein clusters database is updated at quarterly intervals whereupon validation and quality assessment processes occur. Therefore, there is a 3-month delay for the incorporation of proteins from new genomes into protein clusters because of the processing time necessary.

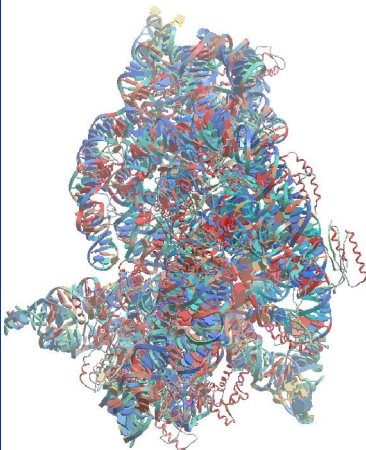
Clusters in the protein clusters database (protclustdb) are named and functional descriptions are assigned by manual curation. Alignments, information on genome neighborhood, and links to NCBI and external databases are provided for each protein cluster. Specific query and search terms can be found under Querying and Searching.

Proteins encoded by prokaryotes, plasmids, viruses, organelles, and complete and incomplete genomes of protozoa are contained within separate cluster groups. Each cluster in the database has a unique identifying number (UID). Each cluster also has an accession number consisting of a three or four-letter code followed by five numbers. Accession numbers are somewhat stable between releases unless the curated cluster is removed, split into subclusters, or joined with another curated cluster (this information is not tracked between releases). The clusters are divided into curated and non-curated sets. Non-curated clusters are automatically generated and have not yet been manually annotated. Manual

Molecule of the Month

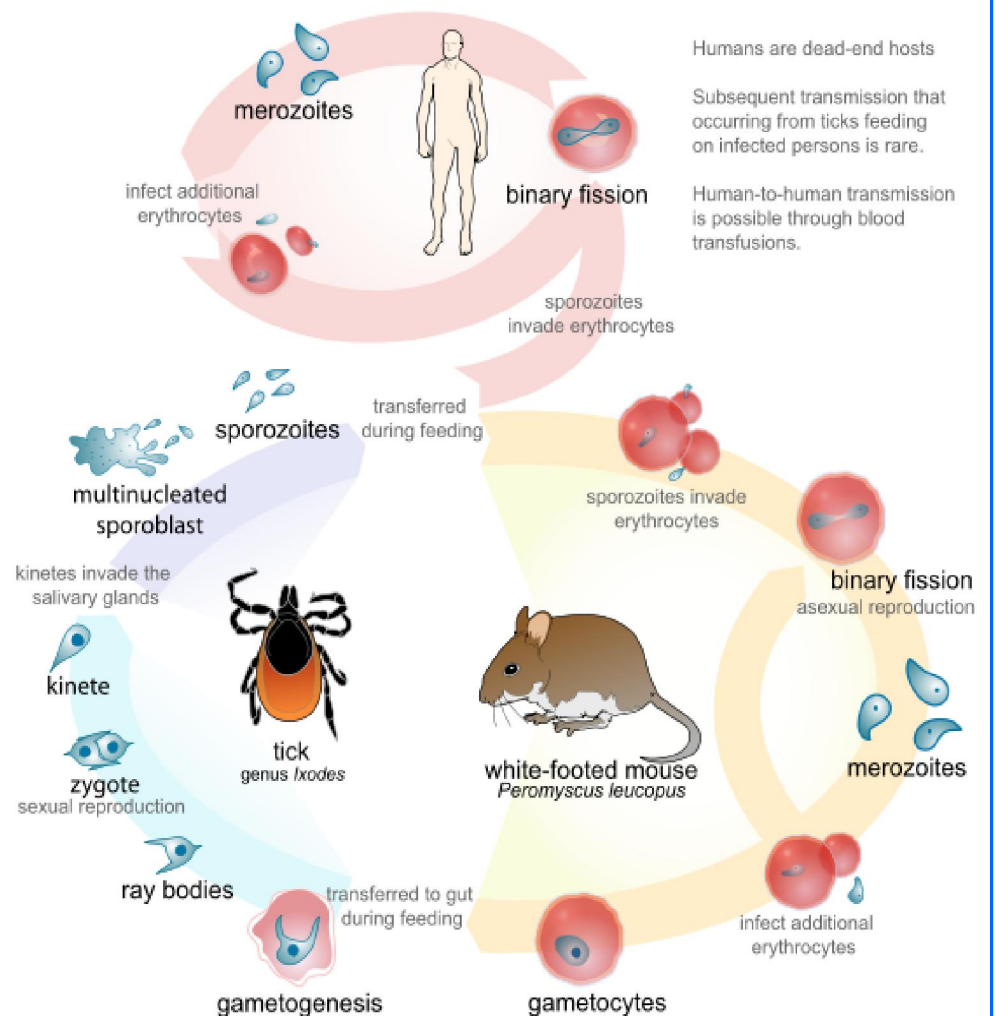
70 S Ribosomes

Ribosomes are one of the wonders of the cellular world, and one of the many wonders you can explore yourself at the RCSB PDB. In 2000, structural biologists Venkatraman Ramakrishnan, Thomas A. Steitz and Ada E. Yonath made the first structures of ribosomal subunits available in the PDB, and in 2009, they each received a Nobel Prize for this work. Structures are also available for many of the other players in protein synthesis, including transfer RNA and elongation factors. Building on these structures, there are now hundreds of structures of entire ribosomes in the PDB, revealing the atomic details of many important steps in protein synthesis.



Molecular Data

PDB ID : 2WDK
 Deposited Date : 25-03-2009
 Release Date : 14-4-2009
 Exp. Method : X-Ray Diff
 Chains : A - Y(25)
 Institute : MRC Lab, U.K



Bioinfy Animator:- Life Cycle of Babesia

Babesia is a protozoan parasite of the blood that causes a hemolytic disease known as Babesiosis. There are over 100 species of Babesia identified; however only a handful have been documented as pathogenic in humans. generated from human, bird and pig flu viruses. Mexican Swine flu is now a pandemic

Courtesy: Wikimedia Foundation, Inc.

For suggestions & contributions contact:

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

Bioinfy Quiz

017

Answers

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