



# Bioinformation up to Date

## (Bioinformatics Center, Biotechnology Division)

North-East Institute of Science & Technology  
Jorhat - 785 006, Assam

### Contents

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

### Adviser:

Dr. P.G. Rao

### Editors:

Salam Pradeep

Dr. R.L. Bezbaruah

### BIF Upcoming Events

1. 4 Days Lectures, demos & hands on training on "Molecular Techniques & Bioinformatics tools in Biological Research" @ College of Veterinary Sciences, AAU, Guwahati from December 2<sup>nd</sup> - 5<sup>th</sup>, 2008.

2. 4 Days Workshop on "Sequence Analysis & Molecular Simulations" @ Anna University, Chennai from December 10<sup>th</sup> - 13<sup>th</sup>, 2008.

### Cover Story

#### A Report from North-East State's Bioinformatics Centers Two Day's Meet

A 2 days interactive meeting for Bioinformatics Centre of North-East states was held on 12-13<sup>th</sup> November 2008 at Chintan Bhavan, Gangtok, Sikkim sponsored by the Department of Biotechnology (DBT), Government of India, New Delhi. The interactive meeting was organized by Bioinformatics centre, Sikkim and Sikkim State Council of Science & Technology, Government of Sikkim.

The inaugural session begins with the lighting of lamp by Dr. Chamling, Honorable Chief Minister of Sikkim, Shri S.B. Subedi, Honorable Minister Department of Science & Technology, Sikkim, Shri M.L. Arrawatia Secretary, Department. of Science & Technology, Sikkim and Dr. T. Madan Mohan, Adviser, Department of Biotechnology, Government of India, New Delhi.

There are a total of twenty four DBT Bioinformatics Centers in North-East with nine Centers in Assam including one Sub Distributed Information Center (DIC) , four Centers in Manipur including one Sub DIC, three Centers in Meghalaya with one Sub DIC, two Centers each in Arunachal Pradesh & Mizoram and one each in Sikkim - Sub DIC, Tripura, Nagaland and North Bengal.

In the first technical session there was PowerPoint presentation about the activity progress of the twenty Bioinformatics Infrastructure Facility (BIF) centers of North-East states from their respective Coordinators.

On the first day at the interactive session DBT Adviser advised the BIF centers to strictly follow the guidelines of the DBT i.e. the purpose of setting up these facility is to support the teaching activities of biology and its allied areas of the host institute in particular and the neighboring institutions in general or in short Biology Teaching through Bioinformatics (BTBI). The 20 lacs fund is not for carrying out research or for buying bioinformatics software.

He also added that Rs. 90 crores are exclusively for North-east region for supporting biology teaching and research in bioinformatics and biotechnology for this current financial year. Out of which the North-east states have utilized only Rs. 20 crores in the initial stages. Rs. 70 crores still remains to be utilized. They are waiting for good proposals from the Bioinformatics centers of North-East states for financial support.

On the second day there was presentation form the DBT Sub DICs followed by lectures form the subject experts like Dr. D. Mohanty form National Institute of Immunology, New Delhi, Dr. Prasad form Amravati University and Dr. P. Ghosh from Bose Institute, Kolkata.

On the second day at the interactive session, DBT adviser announced Bioinformatics Center, North Eastern Hill University - a Sub DIC as the nodal Bioinformatics Center of North-East.

He further added that an additional grant of 7 lacs will be given to all the BIF centers plus a grant of Rs. 10 lacs for every year to all the Centers for running the Center.

He also told the coordinators that a commercial Bioinformatics software with multiple user license may also be taken for the North-East BIF Centers.

..... Salam Pradeep

## Special Interests

### FightAIDS@Home

FightAIDS@Home ("Fight AIDS at home") is a distributed computing project for internet-connected home computers, operated by the Olson Laboratory at The Scripps Research Institute. It aims to use biomedical software simulation techniques to search for ways to cure or prevent the spread of AIDS and HIV. It was originally implemented using a distributed computing software infrastructure provided by Entropia. However, since May 2003 FightAIDS@Home has not been associated with Entropia, and on November 21, 2005 the project moved to World Community Grid and the Entropia software was abandoned.

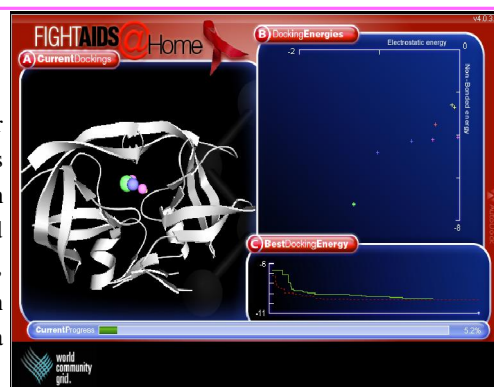
FightAIDS@Home makes use of the AutoDock UD agent software, which tests how well a particular molecule binds to the HIV-1 protease. The minimum system requirements to run FightAIDS@home are:

Processor: 550 MHz | Memory: 250 MB | Hard drive: 600 MB (50 MB free)

Screen resolution: 8-bit graphics at 800 x 600 | Internet connection 40 kbit/s.

Scripps Research Institute published its first peer-reviewed scientific paper about the results of FightAIDS@Home on April 21, 2007. This paper explains that the results up to that point will primarily be used to improve the efficiency of future FightAIDS@Home calculations.

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



## Proteomics

Swiss 2DPAGE



### Two-Dimensional Polyacrylamide Gel Electrophoresis Database

SWISS-2DPAGE is an annotated two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) and SDS-PAGE database established in 1993 and maintained collaboratively by the Biomedical Proteomics Research Group (BPRG) of the Geneva University and the Proteome Informatics Group of the Swiss Institute of Bioinformatics (SIB).

The SWISS-2DPAGE database assembles data on proteins identified on various 2-D PAGE and SDS-PAGE maps. Each SWISS-2DPAGE entry contains textual data on one protein, including mapping procedures, physiological and pathological information, experimental data (isoelectric point, molecular weight, amino acid composition, peptide masses) and bibliographical references. In addition to this textual data, SWISS-2DPAGE provides several 2-D PAGE and SDS-PAGE images showing the experimentally determined location of the protein, as well as a theoretical region computed from the sequence protein, indicating where the protein might be found in the gel.

Cross-references are provided to Medline and other federated 2-DE databases (Cornea-2DPAGE, COMPLUYEAST-2DPAGE, DOSAC-COBS 2D Page, ECO2DBASE, HSC-2DPAGE, LENS-2DPAGE, OGP-WWW, PHCI-2DPAGE, PMMA-2DPAGE, Siena-2DPAGE, YEPD) and to UniProtKB / Swiss-Prot, which provides many links to other molecular databases like EMBL, Genbank, PROSITE, OMIM, etc.

*Courtesy: Swiss Institute of Bioinformatics & Geneva University, Geneva, Switzerland.*

## Genomics

NCBI dbEST



### Expressed Sequence Tag Database

dbEST is a division of GenBank that contains sequence data and other information on "single-pass" cDNA sequences, or "Expressed Sequence Tags", from a number of organisms.

An Expressed Sequence Tag or EST is a short sub-sequence of a transcribed cDNA sequence. They may be used to identify gene transcripts, and are instrumental in gene discovery and gene sequence determination. The identification of ESTs has proceeded rapidly, with approximately 52 million ESTs now available in public databases like GenBank 5/2008, all species.

An EST is produced by one-shot sequencing of a cloned mRNA i.e. sequencing several hundred base pairs from an end of a cDNA clone taken from a cDNA library. The resulting sequence is a relatively low quality fragment whose length is limited by current technology to approximately 500 to 800 nucleotides. Because these clones consist of DNA that is complementary to mRNA, the ESTs represent portions of expressed genes. They may be present in the database as either cDNA / mRNA sequence or as the reverse complement of the mRNA, the template strand.

ESTs can be mapped to specific chromosome locations using physical mapping techniques, such as radiation hybrid mapping or FISH. Alternatively, if the genome of the organism that originated the EST has been sequenced one can align the EST sequence to that genome. The current understanding of the human set of genes (2006) includes the existence of thousands of genes based solely on EST evidence. In this respect, ESTs become a tool to refine the predicted transcripts for those genes, which leads to prediction of their protein products, and eventually of their function.

*Courtesy: National Center for Biotechnology Information, USA*

**Software Mania**



**Accelrys GCG (Group for Computer Genetics)**

The Accelrys GCG Package (Accelrys Inc., San Diego, CA) is an integrated comprehensive sequence analysis package consisting of over 130 programs. These programs can be categorized within the following functional groups:

- (i) Comparison (ii) Database searching and retrieval (iii) DNA/RNA secondary structure (iv) Evolution (v) Fragment assembly (vi) Gene finding and Exporting (vii) Mapping (viii) Primer selection (ix) Protein analysis (x) Translation.

The package allows users to conduct intensive sequence database retrieval and search, including BLAST, FASTA, HmmerSearch, that uses a profile hidden Markov Model as a query to search a sequence database and MotifSearch, that uses a set of MEME profiles to search a database to find new sequences similar to the original family. The databases that GCG supports include most major public ones of nucleic acid and protein sequences. The package provides means for paired and multiple sequences alignments, evolutionary phylogeny study, fragment assembly, restriction enzyme mapping, secondary structure prediction, protein analysis, primer selection etc. Using the Wisconsin Package makes it easy to work with different data formats, such as EMBL, GenBank, PIR and FastA. Furthermore, GCG allows users to create new scoring matrices for sequence comparison and even to make their own databases for database search. In addition to the command line interface, most programs of GCG can be accessed through the graphic user's interface (GUI) by which user can run a program by clicking on and selecting from pull-down menus.

*Courtesy: Accelrys Inc., San Diego, California,*

**Bio Servers**

**EBI ClustalW**

**Multiple Sequence Alignment Server**

ClustalW is a general purpose multiple sequence alignment program for DNA or proteins. It produces biologically meaningful multiple sequence alignments of divergent sequences. It calculates the best match for the selected sequences, and lines them up so that the identities, similarities and differences can be seen. Evolutionary relationships can be seen via viewing Cladograms or Phylograms.

The ClustalW web form is available at: <http://www.ebi.ac.uk/Tools/clustalw/>.

There are two ways to use this service at the EBI. The first is interactively and the second is by email. Using it interactively, the user must wait for the results to be displayed in the browser window. The email option means that the results will not be displayed in the browser window but will be sent by email. The email option is the better one to take when submitting large amounts of data.

The program ClustalW can be used for two purposes:

1. It can be used to produce a multiple sequence alignment.
2. It can be used to produce a true phylogenetic tree.

*Courtesy: European Bioinformatics Institute, U.K.*

YOUR EMAIL <input type="text"/>	ALIGNMENT TITLE Sequence	RESULTS interactive	ALIGNMENT full
KTUP (WORD SIZE) def	WINDOW LENGTH def	SCORE TYPE percent	TOPDIAG def
MATRIX def	GAP OPEN def	END GAPS def	PAIRGAP def
OUTPUT OUTPUT FORMAT: aln w/numbers		PHYLOGENETIC TREE CORRECT DIST.: off	
OUTPUT ORDER: aligned		TREE TYPE: none	IGNORE GAPS: off

Enter or paste a set of sequences in any supported format:

Paste a Multiple Sequence

Click Run

Upload a file:

**Bioinfy Quiz - 006**

1. The chloroplasts found in photosynthetic cells are descendants of

- a) cyanobacteria.
- b) purple non sulfur bacteria.
- c) green sulfur bacteria.
- d) green non sulfur bacteria

2. Transmission of the yeast L and M viruses occurs

- a) by transduction.
- b) when infected yeast cells burst to release infectious virus particles.
- c) when the yeast mate.
- d) all of the above

3. The agent cycloheximide inhibits the function of

- a) cytoplasmic ribosome
- b) mitochondrial ribosome
- c) both cytoplasmic and mitochondrial
- d) neither cytoplasmic or mitochondrial

4. The virus responsible for acquired immunodeficiency syndrome is a(n)

- a) DNA tumor virus.
- b) RNA tumor virus.
- c) retrovirus.
- d) both b and c are correct

## Computers for Biologist

### A Simple PERL Program for Reverse Complement of DNA Sequence

```
print "Program for Reverse Complement of DNA\n";
print "Enter the DNA sequence:";
$DNA = <STDIN>;
$DNA =~ s/\s//g;
print "The original DNA sequence:\n$DNA\n";
@DNA = split( ' ', $DNA );
Print "Reverse complement of the DNA sequence:\n";
foreach $nucleotide(reverse(@DNA)) {
    if $nucleotide =~ /a/i) {
        print "T";
        print WRITE "T";
    } elsif ($nucleotide =~ /t/i) {
        print "A";
        print WRITE "A";
    } elsif ($nucleotide =~ /g/i) {
        print "C";
        print WRITE "C";
    } elsif ($nucleotide =~ /c/i) {
        print "G";
        print WRITE "G";
    } else {
        die "$0: Bad nucleotide! [$nucleotide]\n";
    }
}
```

#### Program Output:

```
C:\Perl\bin\sideshow>perl comp.pl
Program for Reverse Complement of DNA
Enter the DNA sequence:
ATCGTAGCTCGTA
The original DNA sequence:
ATCGTAGCTCGTA
Reverse complement of the DNA sequence:
TACGAGCTACGAT
C:\Perl\bin\sideshow>
```

Answer to  
Bioinfy Quiz  
006

1- a ; 2 - c ; 3 - a ; 4 - d.

## Molecule of the Month

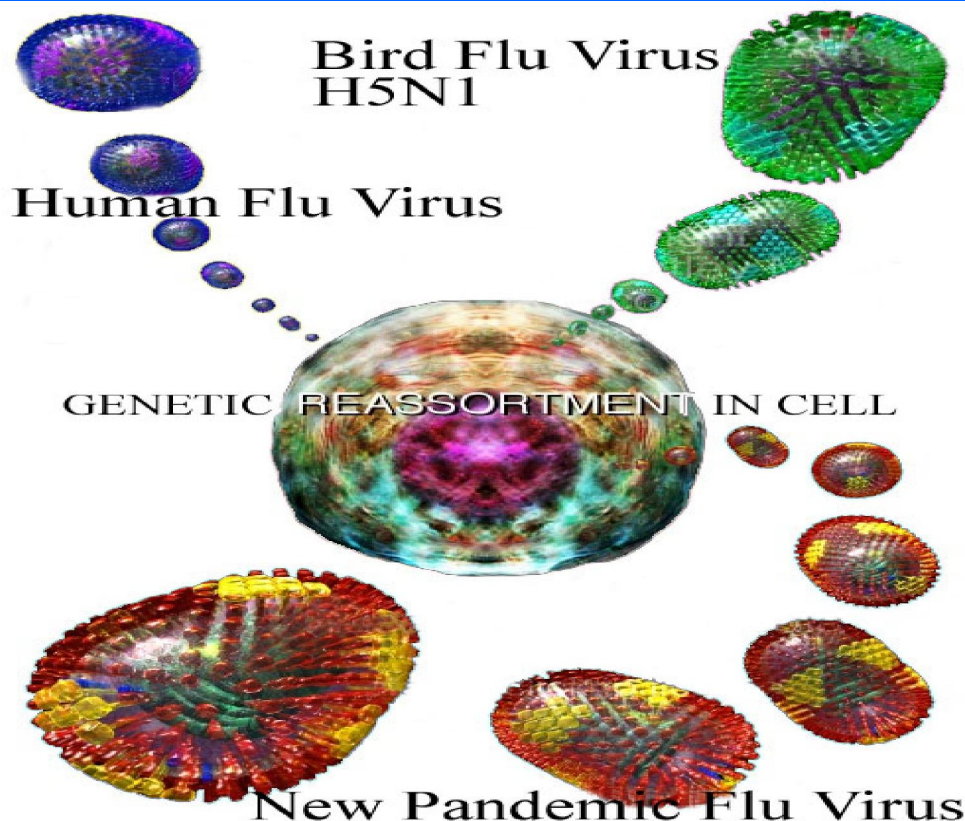
### ANTHRAX PROTECTIVE ANTIGEN

Protective antigen (PA) is the central component of the three-part protein toxin secreted by *Bacillus anthracis*, the organism responsible for anthrax. After proteolytic activation on the host cell surface, PA forms a membrane-inserting heptamer that translocates the toxic enzymes, oedema factor and lethal factor, into the cytosol.



#### Molecular Data

PDB ID : 1ACC  
Amino acids : 735  
Atoms : 5283  
Exp. Method : X-Ray Diff.  
Chains : A (1)  
Release : 1998-02-11  
Classification : Toxin



*Bioinfy Animator: Reassortment of viral RNA segments in a cell infected by two strains of influenza virus (human and bird flu) leading to a new and potentially dangerous strain that could spread easily from human to human and so trigger a deadly worldwide epidemic. Such genetic mixing might occur in pigs, since a pig might be infected by both strains and then pass the new virus on to humans. Alternatively, a person might become infected with bird flu and human flu and start an epidemic of the novel virus.*

*Courtesy: Ruseell Kightley Media, Australia*

#### Please contribute to this bulletin, please contact:

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