



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
Proteomics	2
Genomics	2
Software Mania	3
Bio Server	3
Bioinfo Quiz	3
Computers for Biologists	4
Molecule of the Month	4
Bioinfo Animator	4
Contact Us	4

Adviser:

Dr. P.G. Rao

Editors:

Salam Pradeep Singh

Dr. R.L. Bezbaruah

Important Events

1. DST Research Workshop on "North-East Wild Life Initiative" @ Gauhati University invites participation of young investigators from North-Eastern region universities & institutes.

2. 9th edition of **Bangalore Bio 2009** is scheduled between 18th and 20th June at Hotel Lalit Ashok, Bangalore, Karnataka

Cover Story

MoU between North-East Institute of Science & Technology, Jorhat and Centre for Bioinformatics Studies, Dibrugarh University, Dibrugarh

A Memorandum of Understanding (MoU) have been made between North-East Institute of Science & Technology, Jorhat, Assam and Centre for Bioinformatics Studies, Dibrugarh University, Dibrugarh on 17th April, 2009.

Review of MoU:

- The MoU shall be effective for a period of five years commencing from the date of signing and its continuance shall be subject to review of expiry of agreed period of five years.
- Any dispute arising out of the MoU shall be settled through mutual negotiation between the parties of this MoU.

Purpose and Objective:

The purpose of this MoU is to create a framework for a platform of collaboration with the following specific objectives:

- 1) To establish a close linkage and functional coordination between DUBioinformatics, Dibrugarh University and NEIST, Jorhat
- 2) To recognize NEIST, Jorhat as an accredited research centre and directed study by DuBioinformatics, Dibrugarh University
- 3) To facilitate NEIST, Jorhat professionals to provide education for higher studies, research and need-based training programmes in Bioinformatics to students of DUBioinformatics, Dibrugarh University through the academic programmers of Dibrugarh University.
- 4) To assess and accredit the skill-based training/professional programmes in Bioinformatics Studies offered by NEIST, Jorhat leading to the award of degree and diplomas by the Dibrugarh University.
- 5) To facilitate the qualified CSIR/NEIST, Jorhat professionals as accredited instructors or guides or examiners for the teaching and research programmes of DUBioinformatics, Dibrugarh University.
- 6) 5 percent of the total course fee per student to be paid to NEIST, Jorhat.
- 7) TA/DA, local hospitality and honourarium per class will be paid to guest faculties from NEIST, Jorhat directly by DUBioinformatics, Dibrugarh University.
- 8) Project works will be carried out by the students at NEIST, Jorhat
- 9) To create or develop new experimental or theoretical facilities through mutual cooperation of NEIST, Jorhat and DUBioinformatics and exchange of experts.

Computational Chemistry

CHARMM (Chemistry at HARvard Macromolecular Mechanics)

CHARMM (Chemistry at HARvard Macromolecular Mechanics) is a versatile and widely used molecular simulation program with broad application to many-particle systems. It has been developed with a primary focus on the study of molecules of biological interest, including peptides, proteins, prosthetic groups, small molecule ligands, nucleic acids, lipids, and carbohydrates, as they occur in solution, crystals, and membrane environments.

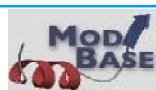
It provides a large suite of computational tools that encompass numerous conformational and path sampling methods, free energy estimates, molecular minimization, dynamics, and analysis techniques, and model-building capabilities and is useful for a much broader class of many-particle systems

CHARMM can be utilized with various energy functions and models, from mixed quantum mechanical-molecular mechanical force fields, to all-atom classical potentials with explicit solvent and various boundary conditions, to implicit solvent and membrane models and it has been ported to numerous platforms in both serial and parallel architectures

CHARMM remains essentially a command line program, i.e. a command is read from the input stream (typed, or from a file) and acted upon.

Proteomics

MODBASE



MODBASE is a relational database of annotated comparative protein structure models for all available protein sequences matched to at least one known protein structure. The models are calculated by MODPIPE, an automated modeling pipeline that relies on the MODELLER package for fold assignment, sequence-structure alignment, model building and model assessment. MODBASE uses the MySQL relational database management system for flexible querying and CHIMERA for viewing the sequences and structures. MODBASE is updated regularly to reflect the growth in protein sequence and structure databases, as well as improvements in the software for calculating the models. For ease of access, MODBASE is organized into different data sets. The largest data set contains 1 262 629 models for domains in 659 495 out of 1 182 126 unique protein sequences in the complete Swiss-Prot/TrEMBL database; only models based on alignments with significant similarity scores and models assessed to have the correct fold despite insignificant alignments are included. Another model data set supports target selection and structure-based annotation by the New York Structural Genomics Research Consortium. MODBASE also contains binding site predictions for small ligands and a set of predicted interactions between pairs of modeled sequences from the same genome.

Database of Comparative Protein Structure Models

Welcome to ModBase, a database of three-dimensional protein models calculated by comparative modeling.

ModBase search form

Search

Search type

Display type

All available datasets are selected

[Select specific dataset\(s\)](#)

To include the academic (comprehensive) dataset, go to 'User Login'

Search by properties

Property

Organism

or

[Advanced search](#)

Genomics

The Arabidopsis Information Resource - TAIR



The Arabidopsis Information Resource (TAIR) is a database of genetic and molecular biology data for the model higher plant Arabidopsis thaliana. Data available from TAIR includes the complete genome sequence along with gene structure, gene product information, metabolism, gene expression, DNA and seed stocks, genome maps, genetic and physical markers, publications, and information about the Arabidopsis research community. Gene product function data is updated every two weeks from the latest published research literature and community data submissions. Gene structures are updated 1-2 times per year using computational and manual methods as well as community submissions of new and updated genes. TAIR also provides extensive link outs to other Arabidopsis resources.

Some important features provided by TAIR through search option are listed below:

DNA/Clones: Search for any DNA, which includes clones, libraries, genomic DNA, and filters.

Ecotypes: Identify natural variants of Arabidopsis and closely related species using various parameters.

Genes: Genes may be searched by name, keywords, features, and/or location.

GO Annotations: Retrieve Gene Ontology annotations and draw gene function pie charts for your list of AGI locus

Markers:

The TAIR Marker Search window provides three ways of searching for a marker: simple search by name only, feature search using more limits, and search by position.

Microarray Elements: Find information about the microarray elements (AFGC clones and amplicons, Affymetrix probe sets, CATMA GSTs, and Agilent oligos) contained on the AFGC, Affymetrix AG (8K) and ATH1 (25K) GeneChip, CATMA and Agilent arrays

Microarray Experiments: Search microarray experiments by name, description, experimenter's last name, array manufacturer and keywords.

Microarray Expression: Search for microarray gene expression profiles.

Seeds/Germplasm: Search for ABRC seed stocks or other mutant lines.

Sequences: Download a variety of sequences from the Arabidopsis Genome Initiative (AGI) in FASTA or tab-delimited formats.

Software Mania

SCHRÖDINGER

Schrödinger - CombiGlide

CombiGlide is a structure-based virtual screening program for the design of optimal, focused combinatorial libraries. CombiGlide significantly accelerates lead discovery and streamlines lead optimization efforts. Schrödinger's CombiGlide is a program specifically created to design focused libraries. Its key features are listed below:

- **Library enrichment:** CombiGlide identifies the most effective reagent combinations to produce focused libraries that have the highest likelihood of binding well to the target protein. CombiGlide dramatically reduces the overwhelming combinatorial space down to manageable library sizes by selecting and ranking reagents.
- **Unmatched accuracy:** CombiGlide takes advantage of the incomparable XP (extra precision) scoring function of Glide; thus ensuring that ranking of compounds and reagents is performed with the highest possible accuracy.
- **ADME properties:** CombiGlide can optionally filter compounds using predicted ADME properties, eliminating from consideration compounds that may bind well but exhibit undesirable pharmacokinetic profiles.
- **Flexible library selection:** CombiGlide provides an efficient and powerful library selection tool that allows the user full flexibility in configuring the focused library based on docking results and ADME properties, as well as user-input parameters.
- **Easy-to-use interface:** CombiGlide guides the user through the workflow with an intuitive interface that begins with preparing the protein and reagents, and proceeds to docking calculations and selecting the final compound library. The Maestro interface provides helpful structural visualization and analysis tools.
- **Cross-platform support:** CombiGlide supports Linux and SGI. Calculations may run across multiple CPUs to maximize throughput.

Bio Servers

MHC Pred

MHC Pred 2.0 Server

MHC Pred 2.0 Server is an additive method to predict the binding affinity of major histocompatibility complex (MHC) class I and II molecules and also to the Transporter associated with Processing (TAP). Allele specific Quantitative Structure Activity Relationship (QSAR) models were generated using partial least squares (PLS). Sequences are limited to a maximum of 1000 residues. Longer sequences are not be accepted, for reasons of CPU usage.

[Enter the query sequence \(plain format\)](#)
[Select the allele](#)

Sequences are limited to a maximum of 1000 residues

Currently only sequences in plain format are accepted

FASTA or other format are not accepted by MHC Pred

HLA-A*0201

H-2Db

H-2Kb (8-mer)

H-2Kk (8-mer)

HLA-A*0101

[Select the model](#)
[Set absent value](#)
[List the results in the order of](#)
[The cut-off values of IC₅₀](#)

Only amino acids

Amino acids and interactions

0

Input sequence

Predicted -logIC₅₀

0

[Select the anchor positions](#)

position(1-9)	position (1-9)	position (1-9)	position (1-9)
<input checked="" type="checkbox"/> A	<input type="checkbox"/> M	<input checked="" type="checkbox"/> A	<input type="checkbox"/> M
<input type="checkbox"/> C	<input type="checkbox"/> N	<input type="checkbox"/> C	<input type="checkbox"/> N
<input type="checkbox"/> D	<input type="checkbox"/> P	<input type="checkbox"/> D	<input type="checkbox"/> P
<input type="checkbox"/> E	<input type="checkbox"/> Q	<input type="checkbox"/> E	<input type="checkbox"/> Q
<input type="checkbox"/> F	<input type="checkbox"/> R	<input type="checkbox"/> F	<input type="checkbox"/> R
<input type="checkbox"/> G	<input type="checkbox"/> S	<input type="checkbox"/> G	<input type="checkbox"/> S
<input type="checkbox"/> H	<input type="checkbox"/> T	<input type="checkbox"/> H	<input type="checkbox"/> T
<input type="checkbox"/> I	<input type="checkbox"/> V	<input type="checkbox"/> I	<input type="checkbox"/> V
<input type="checkbox"/> K	<input type="checkbox"/> W	<input type="checkbox"/> K	<input type="checkbox"/> W
<input type="checkbox"/> L	<input type="checkbox"/> Y	<input type="checkbox"/> L	<input type="checkbox"/> Y

Bioinfo Quiz - 012

- Which virus was the causative agent of small pox?
 - Cowpox virus
 - Vaccinia virus
 - Variola virus
- Which of these drugs cause inhibition of translocation in prokaryotes?
 - Erythromycin
 - Streptomycin
 - Tetracycline
- When was the first t-RNA sequence published?
 - In 1964
 - In 1966
 - In 1968
- Jaundice is due to an accumulation of which heme degradation product?
 - Porphobilinogen
 - Bilirubin
 - Biliverdin
- Which was the first protein to be completely sequenced?
 - Cytochrome C
 - Insulin
 - Lysozyme

Answers on Page 4

DAS - BIODAS

The Distributed Annotation System (DAS) defines a communication protocol used to exchange annotations on genomic or protein sequences. It is motivated by the idea that such annotations should not be provided by single centralized databases, but should instead be spread over multiple sites. Data distribution, performed by DAS servers, is separated from visualization, which is done by DAS clients.

DAS is a client-server system in which a single client integrates information from multiple servers.

It allows a single machine to gather up sequence annotation information from multiple distant web sites, collate the information, and display it to the user in a single view. Little coordination is needed among the various information providers.

DAS is heavily used in the genome bioinformatics community. Over the past years DAS have also seen growing acceptance in the protein sequence and structure communities.

The original DAS specification was written by Lincoln Stein, Sean Eddy, and Robin Dowell. It is widely adopted and well supported, particularly throughout Europe, and is the basis for a large number of existing clients and servers.

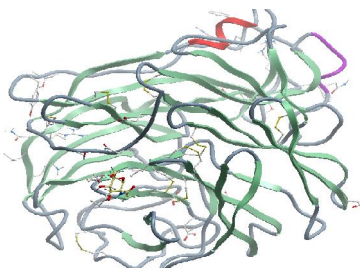
Hundreds of DAS servers are currently running worldwide including WormBase, FlyBase, Ensembl, TIGR, UCSC, and UniProt.

Molecule of the Month

Influenza Neuraminidase

Influenza virus is continually changing and every decade or so, a dangerous new strain appears and poses a threat to public health. This year, there has been an outbreak of a new strain of H1N1 flu, commonly known as swine flu. The H1N1 designation refers to the 2 molecules that cover the surface of the virus: hemagglutinin and neuraminidase. These 2 molecules control the infectivity of the virus. Hemagglutinin plays the starring role as the virus approaches a cell, binding to polysaccharide chains on the cell surface and then injecting the viral genome into the cell. Neuraminidase, plays its major role after the virus leaves an infected cell. It ensures that the virus doesn't get stuck on the cell surface by clipping off the ends of these polysaccharide chains.

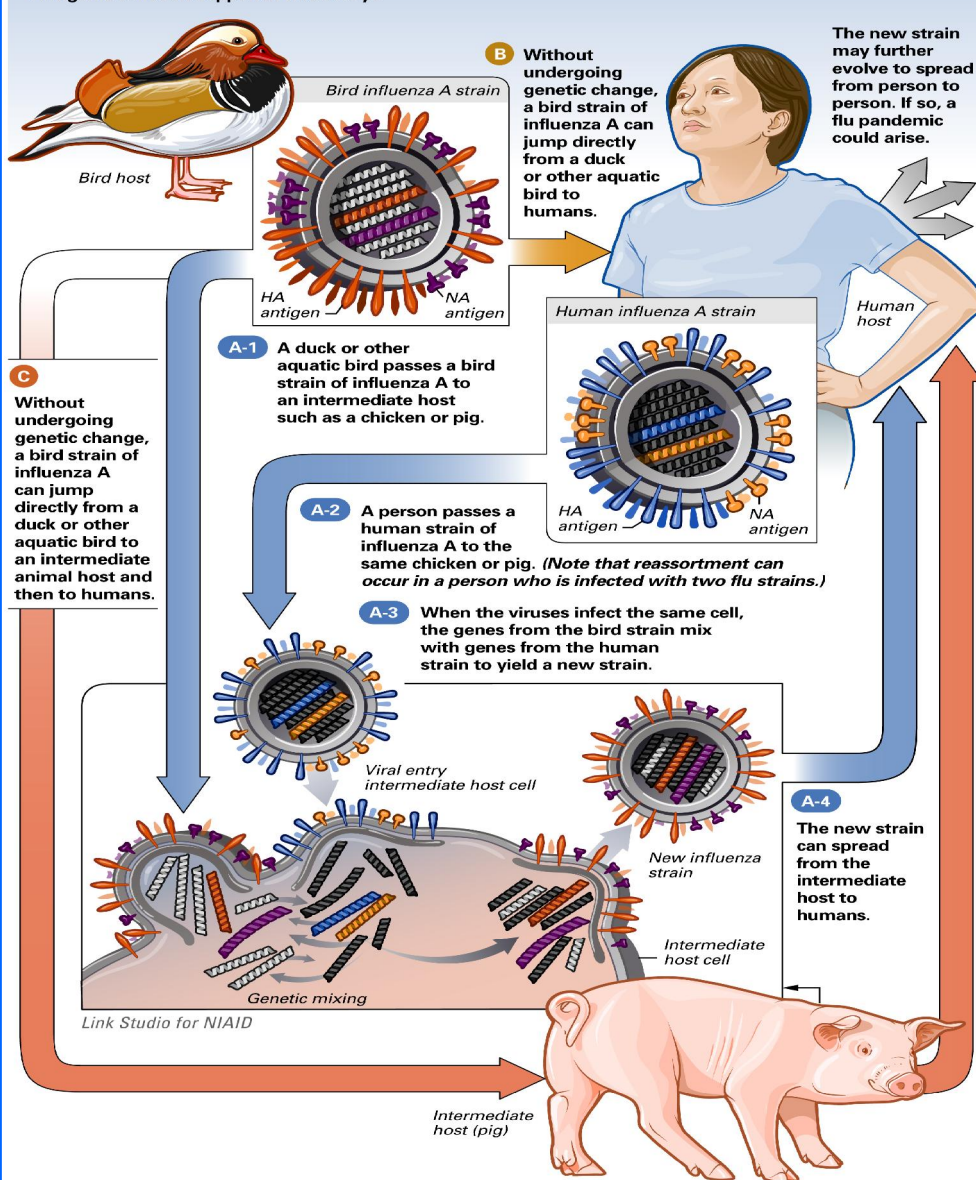
Neuraminidase, shown below is composed of 4 identical subunits arranged in a square. It is normally attached to the virus surface through a long protein stalk. The active sites are in a deep depression on the upper surface. They bind to polysaccharide chains and clip off the sugars at the end.



Molecular Data

PDB ID	: 1NN2
Amino acids	: 388
Atoms	: 3747
Exp. Method	: X-ray
Chains	: A

The genetic change that enables a flu strain to jump from one animal species to another, including humans, is called "ANTIGENIC SHIFT."
Antigenic shift can happen in three ways:



Bioinfy Animator:- Swine can be infected by both avian and human influenza strains of influenza, and therefore are hosts where the antigenic shifts can occur that create new influenza strains.

For suggestions & contributions contact:

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

Bioinfy Quiz

012

Answers

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