



Bioinformatics up to Date

(Bioinformatics Center, Biotechnology Division)
North-East Institute of Science & Technology
Jorhat - 785 006, Assam

CONTENTS

Cover Story	1
Bioinfy Quiz	1
Computational Chemistry	2
Proteomics	2
Bioserver	3
Computers for Biologists	3
Genomics	4
Upcoming Events	4
Bioinfy Animator	5
Molecule of the month	5
Contact Us	5
Answers to quiz	5

Adviser:
Dr.P.G.Rao

Editors:
Dr. R.L.Bezbaruah
Pompi Sharma
Dhrubjyoti Gogoi

COVER STORY

National Technology Day was observed at NEIST, Jorhat on 11th of May, 2011. Dr. J. Mahanta, Scientist G and Director, ICMR Dibrugarh, was invited as the chief guest. The event was initiated by presenting a phoolon *gamucha* and bouquet to the guest of Honour and Director, NEIST. It was then followed by a lecture by Dr. J. Mahanta. He spoke on “Challenges for technology in medical diagnostics & personalized drug therapy” considering various technology related to the medical science. How being the technology is used since the ancient times till the most recent period. He also talked about advances in pharmacological science like Tailored drug therapy and genetic profiling and nanobiotechnology in medical sciences.



BIOINFY QUIZ

- The field of study involving the sequencing of the genomes of organisms is
 - bioinformatics.
 - genomics
 - molecular genetics
- Microarray gene chips will eventually be used to?
 - screen for mutations leading to cancer.
 - identify carriers of genetic diseases.
 - identify probable behavioral traits.
 - A, B, and C are Correct.
- The identification of the function of a gene in a genome can be accomplished using
 - gel electrophoresis.
 - functional genomics.
 - bioinformatics.
- Labeling a stretch of DNA according to its function is called
 - functional analysis.
 - annotation.
 - screening.
- The new area of science that seeks to catalog and analyze every protein in the human body in order to help understand the human genome is called
 - bioinformatics.
 - proteomics.
 - molecular genetics.

Answers on page 5

COMPUTATIONAL CHEMISTRY

Computational Chemistry at uBe industries - Tools Used in CUTting-EdgE NaNotEchNol ogy applications

Modelling and Simulation Tools Used In New Materials R&D. The R&D department of UBE Industries Ltd. designs new 'specialty chemicals' that deliver high value through inclusion in pharmaceutical products or creating materials with specific functional properties. The manipulation of materials to create functional properties is an important goal of the emerging area of nanotechnology. This technology that controls material structures on the atomistic or molecular level enables the development of unique innovative functions as well as incremental improvements to existing materials and processes. Computational chemistry is exploited by UBE Industries Ltd. as a dominant approach to creating novel nanomaterials.

Modelling and simulation can be used to study material structures at the atomistic and molecular level, allowing existing materials to be refined and new materials designed, often with unique properties. These tools are particularly well suited to the study of nanomaterials. In looking back almost 15 years, Dr Shigeru Yao, who doubles as Manager of the Materials Design Department and Group Leader of the Nanotechnology Promotion Group in the Polymer Research Laboratory of UBE Industries, reflects on the circumstances leading to adoption of computational chemistry. "Around 1990, while there was a trend towards new polymeric materials development, it appeared that property-estimation methods based on empirical rules had reached their limit. At that point we adopted software that deals with molecular mechanics and molecular dynamics. This was our entry point into the use of computational chemistry." UBE Industries now uses Accelrys' modeling and simulation tools. Accelrys is a leading supplier of research-support software for life and materials science. UBE uses Materials Studio®, a modeling, simulation, and informatics environment for the chemicals and materials industries. The Materials Studio tools that UBE uses include Discover that calculates molecular mechanics and molecular dynamics, and MesoDyn that calculates coarse-grained dynamics. UBE uses these tools to design materials ranging from polymers to catalysts.

PROTEOMICS

Quantitative proteomics based on stable isotope tagging

A key aspect to the comprehensive characterization of protein products is the quantitative analysis of protein profiles. For this, two alternative approaches have been developed. The first and most widely used method is based on high resolution two-dimensional electrophoresis (2DE) and mass spectrometry, the second on quantitative mass spectrometry via stable isotope tagging of proteins and peptides. In the 2DE-based method, complex protein mixtures are initially separated electrophoretically and stained. Specific proteins are then selected for mass spectrometric identification based on quantitative comparison of the 2DE staining patterns of suitable experimental and control protein samples. Whereas the technique is mature and robust, several conceptual and technical considerations limit its general utility. Most significantly, a study using unfractionated soluble proteins from a wholecell yeast (*Saccharomyces cerevisiae*) lysate demonstrated that even with maximal sample loading and extended electrophoretic separation, low-abundance proteins, which constitute nearly half the yeast proteome, were systematically excluded. The most significant recent advances in quantitative proteomics have been catalyzed by quantitative mass spectrometry, the subject of this review. This method consists of the following four steps: (1) differential isotopic labeling of separate protein mixtures; (2) digestion of the combined, labeled protein mixtures followed by separation of the resulting peptides by multidimensional liquid chromatography (LC); (3) analysis of the separated peptides by automated tandem mass spectrometry (MS-MS); and (4) automated database searching to identify the peptidesequences (and hence the proteins from which they were derived) and determination of relative protein abundance from the mass spectral data.

BIOSERVER

MAFFT 6.849 – Multiple Alignment Programm

MAFFT (Multiple sequence Alignment based on Fast Fourier Transform) is a multiple sequence alignment program for amino acid or nucleotide sequences which runs on unix-like operating systems. It offers a range of multiple alignment methods, **L-INS-i** (accurate; for alignment of $< \sim 200$ sequences), **FFT-NS-2** (fast; for alignment of $< \sim 10,000$ sequences), etc.

Multiple sequence alignment (MSA) is an important step in comparative sequence analysis. Parallelization is a key technique for reducing the time required for large-scale sequence analyses. The three calculation stages, all-to-all comparison, progressive alignment and iterative refinement, of the MAFFT MSA program were parallelized using the POSIX Threads library. Two natural parallelization strategies (best-first and simple hill-climbing) were implemented for the iterative refinement stage

The screenshot shows the MAFFT 6.849 web interface. At the top, it says "mafft 6.849" and "Multiple alignment program for amino acid or nucleotide sequences." There are "Run" and "Reset" buttons. Below this is the "Input Options" section, which includes a "Sequences File (a file containing several sequences)" field with "paste", "db", and "upload" buttons. A text area contains the following sequences:

```
FNIREVILFFDRRRRLTF
>1ASZB
EDTAKDNYGKLFLLIQSRDSRDTGQRVVKFVLDKAKDSDEVLFRARVHNRQQGATLAF
LTLRQQALIQGLVFNKRETTISNNVWASLNLESIVLVIRGIVKVVDFPKSATVQNL
EIHITKIVTISETPEALPILLEDSRSEAEAEAGLPPVNLDRLDYRVIDLRTVTNGAI
FRIQAGVCELFRYLATKKFTEVHTPKLLGAPSEGGSSVFEVYFRGRAYLAQSQFNKQ
```

Below the text area are "Sequences type" (set to "Automatic") and "Allow unusual symbols" (set to "No"). The "Output Options" section shows "Output format" set to "fasta" and "Output order" set to "Aligned". The "Advanced settings" section includes "Strategy" (set to "Auto (FFT-NS-2, FFT-NS-i or L-INS-i; depends on data size)"), "Scoring matrix for amino acid sequences" (set to "BLOSUM62"), "Scoring matrix for nucleotide sequences" (set to "200PAM/k-2"), "Gap opening penalty" (set to "1.53"), and "Offset value" (set to "0.0").

COMPUTERS FOR BIOLOGISTS

SARpRed

SARpred, a neural network based method predicts the real value of surface accessibility (SA) by using multiple sequence alignment. In this method, two feed forward, back-propagation networks are used. The first sequence-to-structure network is trained with PSI-BLAST generated position specific scoring matrices. Further, the initial predictions from the first network and PSIPRED predicted secondary structure are used as input to the second structure-to-structure network. The input is a single letter-code amino acid sequence in free format and output is a real value of surface accessibility corresponding to the amino acid sequence.

GENOMICS

TCF7L2 gene and Type 2 diabetes

Transcription factor-7-like 2 (*TCF7L2*) is the most important type 2 diabetes susceptibility gene identified to date, with common intronic variants strongly associated with diabetes in all major racial groups. This gene encoding protein TCF7L2 is a transcription factor expressed in several tissues, including the gut and the pancreas, involved in the WNT signalling pathway, which plays an important role in β -cell proliferation, insulin secretion and glucagon-like peptide (GLP)-1 synthesis from intestinal L cells. TCF7L2 is a ubiquitous protein that belongs to a family of TCF/LEF transcription factors.

Several hypotheses have been proposed to explain how the genetic variants in *TCF7L2* could increase T2D risk. At the start of 2006, *TCF7L2* was revealed as an unexpected suspect for a type 2 diabetes gene by the DECODE group in Iceland.

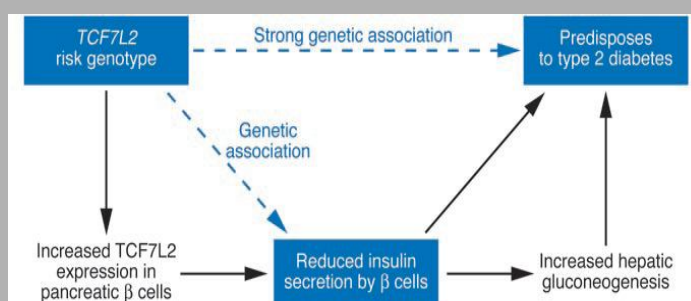


Figure: Pathophysiological pathway of TCF7L2

The most significant genetic T2D association within *TCF7L2* was detected for two intronic single nucleotide polymorphisms (SNPs), rs7903146 and rs12255372, located 50 kb from each other within a 92 kb linkage disequilibrium (LD) block spanning exon 4 and parts of introns 3 and 4. Both *TCF7L2*

SNPs were found to be significantly associated with plasma proinsulin when adjusting for insulin sensitivity. *TCF7L2* appears to play a role in the pancreatic insulin secretory response to incretins, potentially by regulating GLP-1 receptor and GIP receptor expression, and to be particularly involved in the distal regulation of insulin secretion.

Robin Das
MSc Bioinformatics
CBS, Dibrugarh University

UPCOMING EVENTS

Modern Drug Target Crystallography and Structure Based Drug Discovery

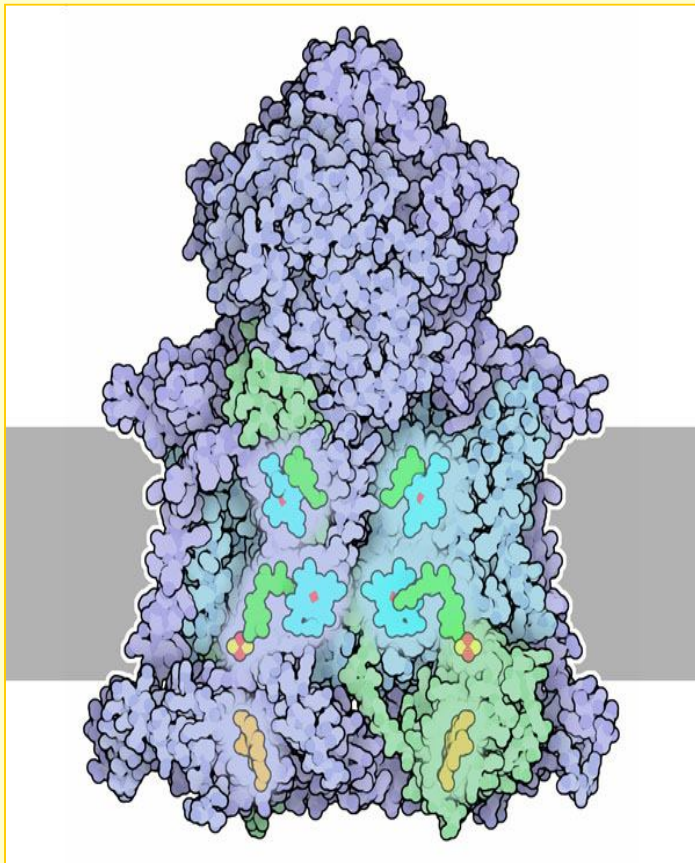
June 16-17, 2011
Molsoft LLC facility
San Diego, California

MOLECULE OF THE MONTH

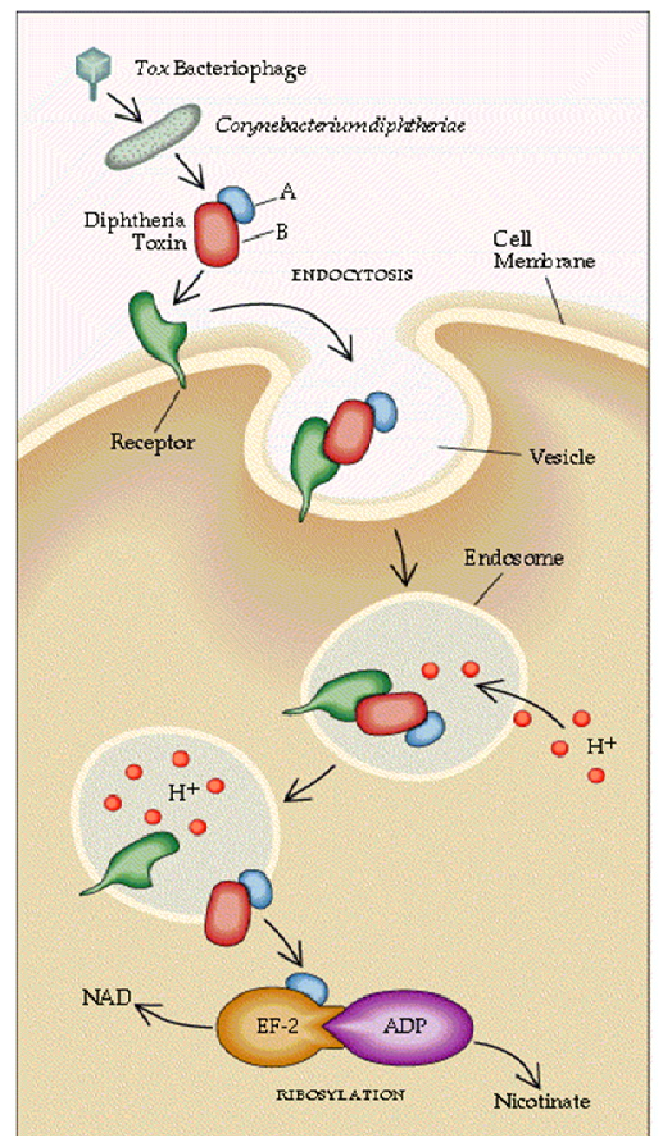
Cytochrome bc1

Cytochrome bc1 binds to ubiquinol, a carrier of hydrogen atoms that is found in the mitochondrial membrane, and removes two protons and two electrons. It is a dimeric protein. Each half is composed of 11 protein chains and a complex collection of cofactor including several hemes and an iron-sulfur cluster.

PDB ID:3H1J.pdb
METHOD: X-Ray Crystallography
RESOLUTION: 3.00



UPTAKE AND ACTIVITY OF DIPHTHERIA TOXIN IN EUKARYOTIC CELLS



CONTACT US:

Pompi Sharma, Project Assistant (Level – II)

E-mail: pompi.sharma86@gmail.com

Dhrubajyoti Gogoi, Project Assistant (Level – II)

E-mail: dhruba.bio.du@gmail.com

Answers of Bioinfy Quiz

1) B 2)D 3)A 4)B 5)B