Volume 3, Issue 1

30th January 2012

# TezuBioin

**(A Quarterly e-Newsletter on Bioinformatics and Biotechnology)**

**DBT Funded-Bioinformatics Infrastructure Facility**

**Department of Molecular Biology and Biotechnology**

**Tezpur University, Tezpur**

## In this issue



## BTISnet News

Annual All India BTISnet Coordinators meeting will be held on 3rd & 4th February, 2012 at Shri Mata Vaishno Devi University, Katra, Jammu. Preceding the Coordinators meeting the University of Jammu would be organizing a one day national Symposium on Bioinformatics on 2nd February 2012 at Jammu.

## Front News: Report of the IV NEBINET Meeting

The IV NEBINET Coordinator's Meeting was held at North Eastern Regional Institute of Science and Technology (NERIST), Nirjuli, Arunachal Pradesh on 10-11th November 2011. This was also the silver jubilee year of the BTISnet.

Release of the proceedings of the IV NEBINET Coordinators Meeting

The meeting started with the lighting of the lamp by the Chief Guest Prof. Mihir K. Chaudhuri, Honorable, Vice Chancellor, Tezpur University. Dr. C.L. Sharma, Head, Department of Forestry, NERIST welcomed the chief guest and the participants. This was followed by the key-note address by Dr. T. Madhan Mohan, Adviser, DBT. In his address, he informed about "Role of North Eastern Bioinformatics (NEBINET) for the promotion of Biotechnology in North Eastern States" wherein he informed about the various programmes implemented by DBT for the North East Region like twining R&D programme, star colleges, biotech hubs, DBT-AAU Centre and many other important programmes. The proceedings of the meeting were released followed by the address by Dr. Dipankar Pal, Director, NERIST appreciating the programme and motivating the coordinators for active participation in NEBINET. Prof. P. Tandon, Former VC, NEHU also addressed the participants. The chief guest Prof. Mihir K. Chaudhuri, VC, Tezpur University, in his inaugural address suggested to evolve the thrust areas for the NEBINET. He also informed that most of the bioinformaticians are from biological background which is of course important and we have to master various bioinformatics softwares for which we have to engage more and more interdisciplinary scientists and job opportunities for bioinformatics. The Inaugural session ended with vote of thanks by Dr. Suresh Kumar, Coordinator, Bioinformatics Center NERIST. Some of the important recommendations of the IV NEBINET Coordinators Meeting are listed below :


A group photo of the participants of the IV NEBINET Coordinators Meeting

**i)** An amount of Rs 1.00 lac to be increased under the contingency head to meet the expenses for AMC of computers, software & communication equipments and North East being the remote area, the committee recommended the enhancement of the travel grant to Rs. 1.00 lakhs and also enhancement of training component of the grant to Rs 1.00 lac.

**ii)** Barcoding networking programme as well as whole genome sequencing programme to be initiated by the NER Bioinformatics Centers

**iii)** A two weeks training programme on bioinformatics and wet lab to be conducted by Dr. A Verma, ACTREC, Mumbai for NER researchers and Dr. S. K Ray, Tezpur university, Tezpur, will coordinate for organizing this training programme.

**iv)** Dr. Sureshkumar, NERIST will coordinate for the procurement of need based application software packages for the North East Bioinformatics Centres.

**v)** Dr. M. K. Modi, AAU, Jorhat will coordinate for the networking of NEBInet centres through NKN.

**vi)** Only licensed and/or free bioinformatics softwares should be mentioned as part of the report.

**vii)** Inventory of Bioresources of NER to be carried out through remote sensing technology.

**viii)** The committee recommended that the next meeting of the NEBINet will be held at the BIF, Mizoram University, Mizoram.

## Bioinformatics for Students: Prosite

PROSITE is a method of determining the function of uncharacterized proteins translated from genomic or cDNA sequences. It consists of a database of biologically significant sites and patterns formulated in such a way that with appropriate computational tools it can rapidly and reliably identify to which known family of protein the new sequence belongs.

In some cases the sequence of an unknown protein is too distantly related to any protein of known structure to detect its resemblance by overall sequence alignment, but it can be identified by the occurrence in its sequence of a particular cluster of residue types which is variously known as a pattern, motif, signature, or fingerprint. These motifs arise because of particular requirements on the structure of specific region of a protein which may be important, for example, for their binding properties or for their enzymatic activity. These requirements impose very tight constraints on the evolution of those limited in size but important portion of a protein sequence.

The use of protein sequence patterns or motifs to determine the function of proteins is becoming very rapidly one of the essential tools of sequence analysis.

There are many short sequences that are often but not always diagnostics of certain binding properties or active sites. These can be set into a small sub-collection and searched against our sequence.

In some cases, the structure and function of an unknown protein which is too distantly related to any protein of known structure to detect its affinity by overall sequence alignment may be identified by its possession of a particular cluster of residues types classified as a motifs. The motifs, or templates, or fingerprints, arise because of particular requirements of binding sites that impose very tight constraint on the evolution of portions of a protein sequence ."

Based on these observations, to actively pursue the development of a database of patterns which would be used to search against sequences of unknown function, Proite was developed in 1988.

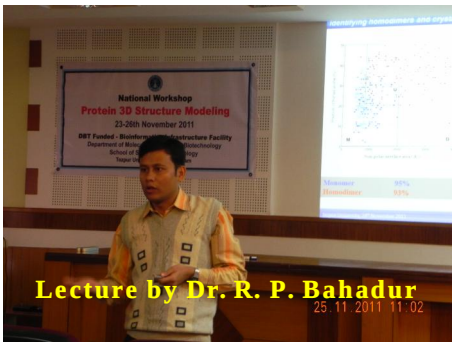
PROSITE contains a few patterns which have been published in the literature by Lesk A.M, but the majority have been developed, in the last ten years by the author. Originally this dictionary was conceived as part of the author's doctoral dissertation as well as an integral part of the PROSITE program in the PC/ Gene sequence analysis software package.

There are a number of protein families as well as functional or structural domains that cannot be detected using patterns due to their extreme sequence divergence; the use of techniques based on weight matrices also known as profiles allows the detection of such proteins or domains. Most of the new PROSITE entries are centered around profiles and are developed by the PROSITE collaborators at the Swiss Institute of Bioinformatics in Geneva and Lausanne. *…. Salam Pradeep*

*Courtesy: expasy.org*

## BIF Workshop: Protein 3D Structure Modeling

The Bioinformatics Infrastructure Facility, Department of Molecular Biology and Biotechnology , has organized a 4 days National Workshop on Protein 3D Structure Modeling from 23-26th November 2011. The workshop started with an introductory talk on Protein to Proteomics by Prof. A.K. Buragohain, Registrar, Tezpur University. Dr. S. K. Ray, Coordinator BIF, discuss the objective of the workshop with the participants. After this, Dr. A. N. Jha deliver a lecture on Protein 3D Structure Modeling and concluded the session of day one. On day two, it started with a hands on session by Dr. A. N. Jha, Dept of MBBT, TU on Protein Databank and biological databases. After this it was followed by Mr. Salam Pradeep on hands on practical session on homology modeling of proteins using Modeller 9.10 program. This was followed by lecture on Ramanchandran Plot including video on G. N. Ramanchandran by Prof. R. C. Deka, Dept of Chem. Sciences, Tezpur University. Prof. R. C. Deka concluded with hands on practical session on molecular docking with Hex software. On day three it was a lecture on 'Macromolecular recognition in the Protein Databank' by Dr. R. P. Bahadur, Indian Institute of Technology, Kharagpur. Dr. Bahadur's lecture was followed by hands on session on macro-





molecular recognition. On day four, the last day of the workshop Prof. R. Swaminathan, Indian Institute of Technology Guwahati delivered a lecture on protein folding . His lecture was followed by a tea cum group discussion on protein folding on which the workshop concludes. A total of 22 participants belong to different institute of North-East India participated in the workshop.


# Editorial

In 2012, tezubioin is entering in its 2nd year. Having the opportunity to write my first editorial for this e-news letter, I would like to thank Prof. B. K. Konwar, the founder editor of this e-news letter and the then coordinator of DBT-funded BIF, at present honourable Vice-Chancellor of Nagaland University. I am also grateful to Prof. M. K. Chaudhuri, honourable Vice Chancellor, Tezpur University for entrusting upon me the responsibility of DBT-funded BIF.

Science has taken root in human civilization ever since human started expressing his thinking into action. The insatiable desire to learn in human mind resulted in unfolding the secret of nature one after the other. One such nature's secret which is getting unfolded at unprecedented rate is genome sequence of different organisms, ever since the method for sequencing DNA was invented. The living world which has always surprised any intelligent mind in the Earth, the whole genome sequence has given hope to scientists to get an understating of the mysterious world.

Genome sequence is like a huge basket having components of all games you name it. Unless you separate different components from each other and arrange it well, player can't play. For separating different components a person must have knowledge which part belongs to which game. This is how the job of a Bioinformatician was defined when bioinformatics field was born in the genomics era. Therefore Bioinformatics has been rightly defined as an interdisciplinary science. The players are biochemists, geneticists, evolutionary biologists, developmental biologists, molecular biologists etc. Interestingly, after genomics, technological revolution has given rise to new research fields such transcriptomics, proteomics and metabolomics. Eventually bioinformatics has also evolved from its role from separation and arrangement of playing items to become good players also. Bioinformatician role in the game is that of an astrologer, who does all predictions from his calculations using some defined rules.

DBT funded BIF-TU is trying implement DBT objectives of making BIF facility available to students of the University and also encouraging students to take up computational biology projects, and organizing workshops on computational biology research. At the same time we are also trying to keep our foot on modern computational biology research by asking evolutionary questions on molecular biology and protein structure.

I hope two years down the line, BIF will be taking up its research on some of the frontier areas of research in computational biology.

*…… S. K. Ray, Dept. of MBBT*

# Bio-Quiz - 06

1. The identification of drugs through genomic study
A) Genomics
B) Proteomics
C) Phrmacogenetics
D) Pharmagenomics

2. SWISSPROT protein sequence database started in the year
A) 1985
B) 1986
C) 1987
D) 1988

3. Which of the following is a crystallographic database?
A) DDBJ
B) CCDC
C) DALI
D) EMBL

4. The factor influencing bioavailability
A) LogP
B) pKa
C) LogS
D) Vd

5. Which of the following is a not chemical file format?
A) CIF
B) CDX
C) SDF
D)MDL

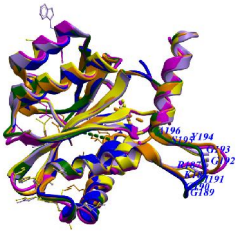
***Prepared by Salam Pradeep Singh***

**Answers: on Page 4**

# Research in BIF Tezpur University

***In silico* structure assessment analysis of core domain of six protein data bank entries of HIV-1 Integrase**

HIV integrase is a 32 kDa protein produced from the C-terminal portion of the Pol gene product, and is an attractive target for new anti-HIV drugs. Its main function is to insert the viral DNA into the host chromosomal DNA, a step that is essential for its replication. However there are six different Protein Data Bank (PDB) entries of the same protein with the same amino acids with PDB IDs 1BIS, 1BIU, 1BIZ, 1HYV, 1HYZ and 1QS4. The present work focuses on the structure assessment analysis of chain A of the different PDB entries of the same protein using *in silico* approaches via the Swiss Model structure assessment server, ANOLEA assessment server and Ramachandran plot analysis. The structure assessment analysis reveals that there is a major difference among these PDB entries based on the energy assessment and structural analysis. The authors also bring out the deviation in the residues form GLY189 to ALA196 of the six PDB entries of HIV-1 Integrase which is shown in the figure. Based on the analysis performed by the authors such as stability, energy analysis and geometrical errors using various assessment servers and tools the authors conclude that the Chain A of 1BIU is more reliable and accurate as compared to chain A's of 1BIS, 1BIZ, 1HYV, 1HYZ and 1QS4 and this finding would aid for assisting computer aided drug designing.



*Salam Pradeep Singh and B. K. Konwar in the Journal of Computational Biology and Bioinformatics Research Vol. 4(1), pp. 1 - 7, January 2012.*

# Guest Column: Symmetry Breaking in Fruits

Identical arrangement of atoms/particles, about a principal axis is generally termed as symmetry. Symmetry can be translational as well rotational. Symmetry is the law of nature. Most of the time, beauty and architectural planning are inadvertantly defined by symmetry. Whether it is the artist's intention or just an aesthetic look, symmetry has always been praised as it not only provides completeness but also draws a balanced overall view of the object.

In modern times, any departure from symmetry, especially in arts and painting, is described loosely as '*smart and stylish*'. Nevertheless, the quest for any sort of unusual property has geared up only when profound attention was given to assess the nature of symmetry as well as asymmetry.

Most of the trees, including fruit bearing ones, and plants adopt symmetric growth all around us. We shall limit our query only to fruit kingdom. A number of fruits generally possess symmetric shape. They are, either spherical or distorted but maintain a balanced shape dictated by their genetic code. For instance, look at the shape of an orange or, a guava; the shape of an apple or, a pomoegranate. In all these cases, though shapes are non-identical all experience an overall spherical symmetry. Surprisingly, *mango* is an exception! It violates symmetry about the principal axis and thus, in overall shape!! Though the extent varies from variety to variety, a kink is found in all of the mangoes. It has not yet been known if the appearance of such a kink or, 'nose' is genetically dictated or thermodynamically favored; or because of both. From physics point of view, a round shape is associated with the minimum energy configuration. A system having minimum energy would realize maximum stability. Accordingly, fruits like orange, apple, palm all experience a totality in shape. From biological perspectives, nature has its own selection which may not be formulated by a direct solution to some mathematical equations**.**

***..... Dr. D. Mahanta, Dept. of Physics, TU***



## Upcoming Event

National workshop on Chemical Informatics form 21-24th March 2012 organized by BIF, Dept. of Molecular Biology and Biotechnology, Tezpur University, Tezpur.

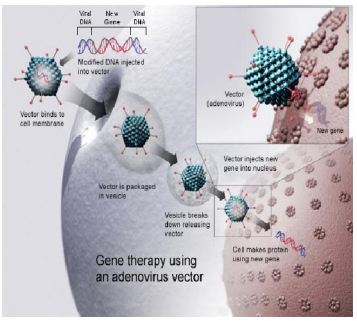
## Information for Students

JNU New Delhi, on behalf of Department of Biotechnology, Government of India, will conduct the BINC examination on February 25-26, 2012.

## Answer to Bio-Quiz-06

**1-D ; 2-C ; 3-B ; 4-B ; 5–B**

## Students' Corner: Gene Therapy




Gene therapy is the use of DNA as a pharmaceutical agent to treat disease. It derives its name from the idea that DNA can be used to supplement or alter genes within an individual's cells as a therapy to treat disease. The most common form of gene therapy involves using DNA that encodes a functional, therapeutic gene in order to replace a mutated gene. Other forms involve directly correcting a mutation, or using DNA that encodes a therapeutic protein drug rather than a natural human gene to provide treatment.

In gene therapy, DNA that encodes a therapeutic protein is packaged within a "vector", which is used to get the DNA inside cells within the body. Once inside, the DNA becomes expressed by the cell machinery, resulting in the production of therapeutic protein, which in turn treats the patient's disease.

Gene therapy was first conceptualized in 1972, with the authors urging caution before commencing gene therapy studies in humans. The first FDA-approved gene therapy experiment in the United States occurred in 1990, when Ashanti DeSilva was treated for X-linked SCID. Since then, over 1,700 clinical trials have been conducted using a number of techniques for gene therapy.

Although early clinical failures led many to dismiss gene therapy as over-hyped, clinical successes in 2009-2011 have bolstered new optimism in the promise of gene therapy. These include successful treatment of patients with the retinal disease Leber's Congenital Amaurosis, X-linked SCID, ADA-SCID, adrenoleukodystrophy, and Parkinson's disease. These recent clinical successes have led to a renewed interest in gene therapy, with several articles in scientific and popular publications calling for continued investment in the field.

Scientists have taken the logical step of trying to introduce genes directly into human cells, focusing on diseases caused by single-gene defects, such as cystic fibrosis, haemophilia, muscular dystrophy and sickle cell anemia. However, this has proven more difficult than modifying bacteria, primarily because of the problems involved in carrying large sections of DNA and delivering them to the correct site on the gene.

Today, most gene therapy studies are aimed at cancer and hereditary diseases linked to a genetic defect. Antisense therapy is not strictly a form of gene therapy, but is a related, genetically-mediated therapy.

The most common form of genetic engineering involves the insertion of a functional gene at an unspecified location in the host genome. This is accomplished by isolating and copying the gene of interest, generating a construct containing all the genetic elements for correct expression, and then inserting this construct into a random location in the host organism. Other forms of genetic engineering include gene targeting and knocking out specific genes via engineered nucleases such as zinc finger nucleases, engineered I-CreI homing endonucleases, or nucleases generated from TAL effectors. An example of gene-knockout mediated gene therapy is the knockout of the human CCR5 gene in T-cells in order to control HIV infection. This approach is currently being used in several human clinical trials.

***..... Compiled by Salam Pradeep Singh***