Bioinformatics up to Date

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

Inside…..

Advisor:
Dr. D. Ramaiah

Editors:
Mr Robin Das
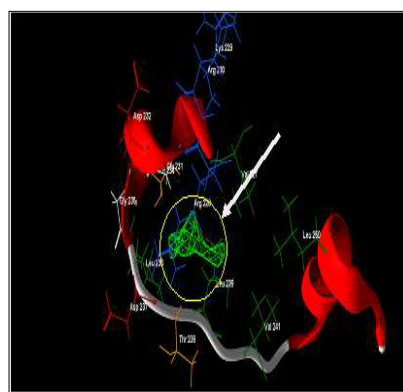**Dr R.L. Bezbaruah**
Dr H P Deka Baruah

About Us

The Bioinformatics Infrastructure Facility (BIF) at Biotechnology division, CSIR NEIST, Jorhat runs under the Biotechnology Information System Network (BTISnet) programme of DBT, Ministry of Science & Technology, and Government of India. The Centre was established on 2nd February, 2008 to promote innovation in Biological research and education through Bioinformatics accomplishment. The main goal is to facilitate and expose students and researchers from different academic institutions of North East India in Bioinformatics. The center conduct training and workshops for enlightening the use of bioinformatics applications in biological research and development. The Centre has access to global information through 24 hour high speed internet facility, and also e-journal facilities with DeLCON, Science Direct etc. To date the Centre has profoundly extended support in R & D work with a great intensity to different biological discipline including medicinal chemistry, computer aided drug design, genomics and proteomic data analysis etc.

Our Focus

Structure prediction and Active site analysis of Coagulase Protein from *Staphylococcus* strain

Staphylocoagulase an extracellular protein which reacts with prothrombin in the blood and enables further conversion of fibrinogen to fibrin. The interesting properties of Staphylocoagulase has put forth to predict the theoretical 3D structure using comparative modeling. The process was carried out in three steps: template identification, template alignment, and model building. Finally, the generated model was assessed for quality, error handling with bioinformatics online server SAVES. The Ramachandran plot analysis showing almost 100% atoms within the allowed regions. The prothrombin binding sites or the active site within the coagulase were found with volume 8.704 Å and surface volume 39.68Å.

[Published in: Preliminary Characterization of Blood Coagulase Protein From a Staphylococcus Strain, Ijp&Ps; Vol 6(5),2014]

Barley Genome

A team led by scientists at the University of California, Riverside has reached a new milestone in its work, begun in 2000, on sequencing the barley genome. The new information, published in The Plant Journal, will not only expand geneticists' knowledge of barley's DNA but will also help in the understanding, at the genetic level, of wheat and other sources of food.

Barley, a widely grown cereal grain commonly used to make beer and other alcoholic beverages, possesses a large and highly repetitive genome that is difficult to fully sequence. The researchers have sequenced large portions of the genome that together contain nearly two-thirds of all barley genes.

It also has applications in plant breeding by increasing the precision of markers for traits such as malting quality or stem rust.

"What we have now is much finer resolution of genetic information throughout the barley genome," said Timothy J. Close, a professor of genetics at UC Riverside and the corresponding author on the research paper. "This is an improved resource used throughout the world. Prior to this work, a long-held view was that the distribution of genes in the genomes of barley, wheat and their relatives is such that the gene-dense regions are only out near the ends of chromosomes where there is also a high rate of recombination. Our work revealed clear exceptions, identifying deviant regions that are gene-rich but low recombination."

New method for Protein Function

SISSA a research group in collaboration with Temple University in Philadelphia has developed a new method for the computerized analysis of the internal dynamics of molecules, demonstrating its efficiency and versatility. The study has been chosen as the cover story of the journal Structure.

In their study, Ponzoni and his colleagues developed a method called SPECTRUS, which allows researchers to
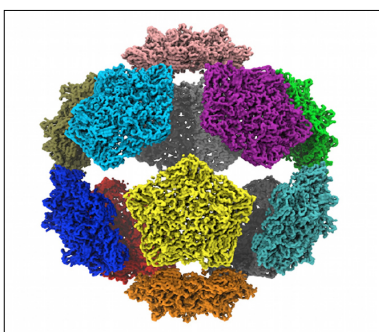
identify the "quasi--rigid" subunits of proteins and reconstruct their movement starting from a limited amount of information. "By knowing the internal dynamics we can understand important things", explains Cristian Micheletti, the SISSA professor who coordinated the study. "Take an enzyme, for example: where two subunits join is usually where we can find the active sites that bind other molecules to the enzyme, which is therefore able to carry out its function as a catalyst".

"The method we used proved to be efficient and reliable", continues Ponzoni. "For instance, it can correctly identify the mechanical modules of a protein starting from only a few frames – to use the same metaphor -providing predictions consistent with those based on far more complex methods that require whole film. Besides, our results were also consistent with those of experimental observations".

Source: SPECTRUS: A Dimensionality Reduction Approach for Identifying Dynamical Domains in Protein Complexes from Limited Structural Datasets. Luca Ponzoni etal.; Structure (2015).

Pipeline for oscillatory genes

A new statistical pipeline for identifying oscillatory genes has developed by scientists at the Morgridge Institute for Research and the University of Wisconsin-Madison. The new approach published in **Nature Methods** called "**Oscope,"** which helps in identifying oscillating genes in single-cell RNA-sequencing experiments.

Oscope gives scientists a new way to identify the dynamics of oscillatory genes, which play an essential role in development functions like cell division, circadian rhythms and limb formation. The key to Oscope is examining cells from an unsynchronized population, where the cells are in different developmental states. Studying oscillatory genes using traditional RNA-seq technology requires investigators to "pick" only one known oscillatory system and perform synchronization toward this system. This step masks the oscillatory signals of any other systems.

"We study unsynchronized single cell data here so none of the oscillatory systems are disturbed," says Ning Leng, a computational biologist with the Morgridge Institute and co-author of the paper. Oscope identifies independent groups of cyclic genes and captures one "base cycle" of each group, offering a practical way to profile distinct groups of genes that play a cyclical role. This technology could open new research roads into genes that are at the heart of basic development.

The Oscope software package is offered free to scientists at:https://www.biostat.wisc.edu/~kendzior/OSCOPE/.

Source: Oscope identifies oscillatory genes in unsynchronized single-cell RNA-seq experiments.Leng, N et al. Nature Methods (August 24, 2015)
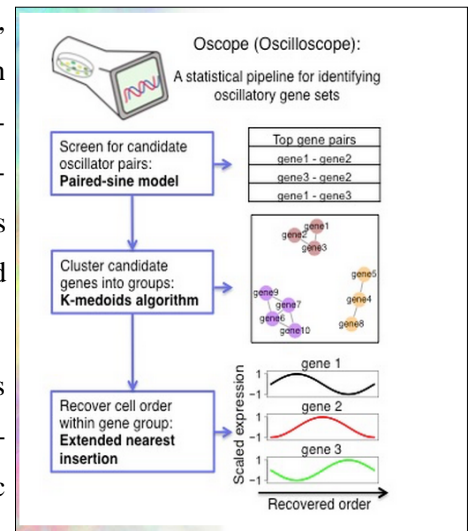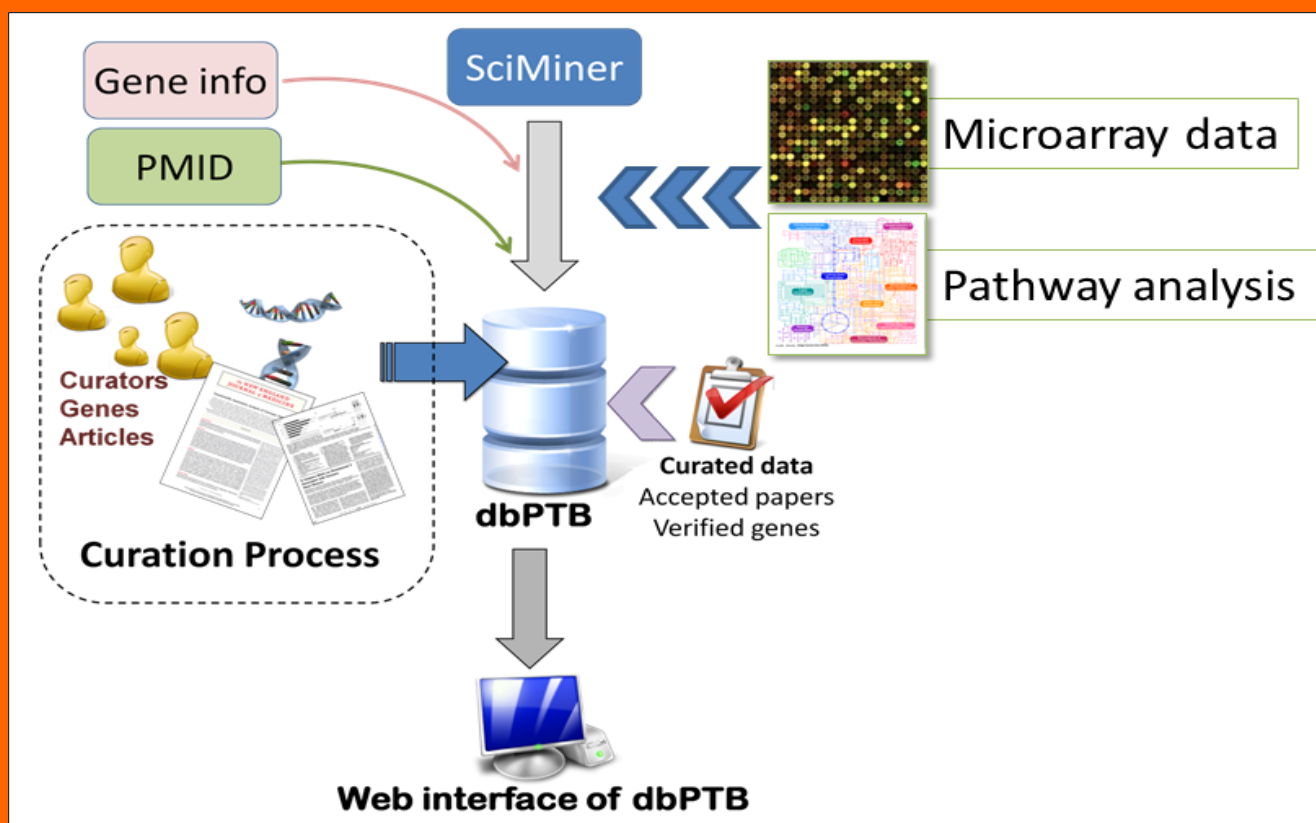
Toolkit for exploring datasets

Nanopore sequencing may be the next disruptive technology in genomics, owing to its ability to detect single DNA molecules without prior amplification, lack of reliance on expensive optical components, and the ability to sequence long fragments. The MinION™ from Oxford Nanopore Technologies (ONT) is the first nanopore sequencer to be commercialized and is now available to early-access users. The MinION™ is a USB-connected, portable nanopore sequencer that permits real-time analysis of streaming event data. Currently, the research community lacks a standardized toolkit for the analysis of nanopore datasets.

Poretools, a flexible toolkit for exploring datasets generated by nanopore sequencing devices from MinION™ for the purposes of quality control and downstream analysis. Poretools operates directly on the native FAST5 (an application of the HDF5 standard) file format produced by ONT and provides a wealth of format conversion utilities and data exploration and visualization tools. Poretools is an open-source software and is written in Python as both a suite of command line utilities and a Python application programming interface. Source code is freely available in Github at https://www.github.com/arq5x/poretools

Source: http://poretools.readthedocs.org/en/latest/

Web interface of dbPTB

Patent News

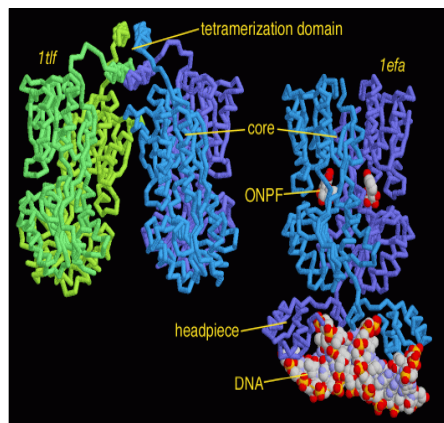# Database system for predictive cellular bioinformatics

US6615141B1

Inventor: James H. Sabry *et al.*

Abstract

A system for acquiring knowledge from cellular information. The system has a database comprising a database management module ("DBMS"). The system also has a variety of modules, including a population module coupled to the DBMS for categorizing and storing a plurality of features (e.g., cell size, distance between cells, cell population, cell type) from an image acquisition device into the database. The system has a translation module coupled to the DBMS for defining a descriptor from a set of selected features from the plurality of features. In a specific embodiment, the descriptor is for a known or unknown compound, e.g., drug. A prediction module is coupled to the DBMS for selecting one of a plurality of a descriptors from known and unknown compounds from the database based upon a selected descriptor from a selected compound. The selected compound may be one that is useful for treatment of human beings or the like.

Lac Repressor

The lac repressor is part of the first regulatory network--the lac operon--that was discovered. It is found in bacteria, where it controls the production of three proteins that are involved in the metabolism of lactose. Its action is very simple. It is a tetramer of four identical subunits that normally binds tightly to a specific region in the bacterial DNA, termed the operator, that is next to a region that encodes three lactose-metabolizing proteins. When bound there, it blocks production of the proteins. But when lac repressor binds to lactose and similar sugars, it changes shape and no longer can bind to the DNA. Then, RNA polymerase is free to transcribe the gene, and the proteins are made.

Lac repressor controls the synthesis of three proteins. Beta-galactosidase, shown here from PDB entry 1bgl, is the enzyme that performs the first step in the metabolism of lactose, breaking it in half into the simple sugars glucose and galactose. Galactoside acetyltransferase, shown here from PDB entry 1krv, is another enzyme that acts on sugars, but its role in lactose metabolism is not as clear. Lactose permease is a protein found in the membrane of the bacterium, through which it transports lactose to the interior.

Each lac repressor subunit folds into three functional regions. The first is a tetramerization domain that links four subunits together into the functional complex. The second is a core domain that binds to lactose and other similar molecules, like the sugar mimic ONPF. The third is the headpiece, which binds to DNA.

Upcoming Events

Workshop on Molecular Biotechnology and Bioinformatics

10th – 14th September 2015

ICSCCB, R.H. 2, Pune – 411045, India

Pharmaceutical Summit and Expo

October 08-10, 2015 New Delhi, India

Kindly send us your feedback to

Robin Das
Research Fellow; BIF, Biotech Division.
CSIR-North East Institute of Science and Technology, Jorhat, Assam
E-mail: robindas460@gmail.com
Ph No-07399923578