International Symposium
on
Computational Biology & Bioinformatics
(ISBB 06)

15-17 December 2006, Bhubaneswar
http://isbb06.googlepages.com

Organized by

Silicon Institute of Technology, Bhubaneswar

Co-sponsors

Science & Tech Department (Biotechnology Wing),
Govt of Orissa, Bhubaneswar
AICTE, Govt of India

Orissa Computer Application Centre, Bhubaneswar
Dept of Biotechnology, Govt of India

Venue

Swosti Plaza, Bhubaneswar

Message from Program Chair

We live in very interesting times: the culturally and linguistically diverse world is speaking the common language of mathematics and science. As we are rediscovering, with Thomas Friedman's assurance, that the world is indeed flat, it seems only befitting that an international congregation of scientists and mathematicians meet at the temple city of Bhubaneswar: a place with a glorious past and a vibrant scientific life with the promise of a bright future.

Bioinformatics and Computational Biology is a burgeoning field at the confluence of disciplines such as computer science, mathematics, and statistics. The aim of the symposium is to bring together researchers of different fields to synergistically tackle the difficult problems that biology presents to us. We received forty-one contributions from twelve countries: India, US, Canada, France, Germany, Italy, Taiwan, Ireland, Spain, Singapore, Israel and Tunisia. The program committee had to make some tough choices. I take this opportunity to thank the committee members, the reviewers, the invited speakers and the contributors for their valuable time. In the end, I truly hope that Silicon Institute of Technology will continue to organize the symposium as a regular annual event.

Laxmi Parida,
Program Chair
IBM Thomas J Watson Research Center, Yorktown Heights, &
Courant Institute of Mathematical Sciences, New York University.

Message from Organising Chair

On behalf of the Organizing Committee I heartily welcome you to the 1st International Symposium on Computational Biology & Bioinformatics (ISBB06) to be held from 15 to 17 December 2006 in the beautiful historical city of Bhubaneswar, Orissa in India.

This is organised by Silicon Institute of Technology with support from Department of Science & Technology, Government of Orissa, Orissa Computer Application Center, Bhubaneswar and AICTE New Delhi. The aim of ISBB06 is to enable eminent researchers from different parts of the world to share their insights, which will lead to new collaborations in the field of Computational Biology and Bioinformatics in our country.

The response to the call for papers was overwhelming due to the untiring efforts of Laxmi Parida and Laxman Mohanty. The chairs and members of Programme Committee are to be praised for their exceptional work in processing the submitted papers, conducting reviews and selecting high quality technical papers and posters. We deeply appreciate their commitment for the success of ISSBB06. Special thanks to Programme Chair, Laxmi Parida for arranging an array of outstanding keynote and invited speakers.

We are grateful to the officers of Government of Orissa, who selflessly volunteered to help in planning and organizing this symposium. In particular we express our deep appreciation to Ajit K Nayak, Prayag Prasad Mishra, Bhagawat Prasad Choudhury, Sushree Samita Rout and Umasankar Das and the entire organizing team of Silicon. We thank the Finance Chair, Ramananda Mishra for taking care of financial matters, Chitta Baral for being the brain child behind this whole programme.

We have put together a program that is rich in quality research papers, keynote address and invited talks by well-known researchers. Hope you will enjoy ISBB06 as much as your stay, sightseeing of the city and surroundings. Finally we thank every one of you for your participation, without you this Programme could not have been a success. Thank you. Once again welcome to Bhubaneswar .

Nitai Gaur Dhall
Silicon Institute of Technology
Bhubaneswar
nitai@silicon.ac.in

Objective of ISBB 06

The twenty-first century is undoubtedly the century of biology. The subject has undergone a paradigm shift, and it is now a truly multidisciplinary pursuit involving biology, computer science, mathematics and statistics. The goal of this symposium is to bring together distinguished researchers from different disciplines working on biologically motivated problems in computer and information sciences. Enabling conversations between these groups would be of benefit to everyone and, hopefully, would lead to new collaborations and insights for all. This is the objective of International Symposium on Computational Biology and Bioinformatics (ISBB 06).

Various topics on which ISBB 06 sought papers and posters included-

- Pattern discovery, classification & applications
- Software tools/environments for bioinformatics
- Sequence analysis & alignment
- Gene networks: theory & applications
- Metabolic modeling & pathways
- RNA & DNA structures
- Protein structures & protein-protein interactions
- Evolution & phylogenetics
- Computational intelligence (artificial neural networks, fuzzy logic, and evolutionary computing) in biological sciences

Symposium Chairs

- Laxmi Parida, Computational Biology Center & Deep Computing Institute, IBM T. J. Watson Research Center (Program Chair)
- Nitai G Dhall, Silicon Institute of Technology, Bhubaneswar (Organising Chair)
- Chitta Baral, Arizona State University, USA (Co-Origanising Chair)
- Ramananda Mishra, Silicon Institute of Technology, Bhubaneswar (Finance Chair)
- Laxman Mohanty, Silicon Institute of Technology, Bhubaneswar (Publicity Chair)

Program Committee

- Laxmi Parida, IBM T. J. Watson Research Center, USA (PC)
- Limsoon Wong, Institute of Infocomm Research, Singapore
- Jens Stoye, Universitaet Bielefeld, Germany
- Naren Ramakrishnan, IBAB, India & Virginia Tech, USA
- Pavel Pevzner, UC San Diego, USA
- Saturo Miyano,University of Tokyo, Japan
- Bubhaneswar Mishra, New York University, USA
- Dan Gusfield, UC Davis, USA
- Chitta Baral, Arizona State University, USA

Organising Committee

- Bhagawat P Choudhury, Coordinator-Registration
- Prayag P Mishra, Coordinator- Accommodation & Conveyance
- Ajit Nayak , Coordinator-Conference arrangement
- Sushri S Rout , Coordinator-Poster session
- Umasankar Das, Coordinator- Technology support

Invited Speakers

- David Sankoff, University of Ottawa, Canada
- See-Kiong Ng, Institute of Infocomm Research, Singapore
- Gene Myers, Howard Hughes Medical Institute & UC Berkeley, USA
- Gad M Landau, University of Haifa, Israel & Polytechnic University., NY, USA
- Ramesh Hariharan, Strand Life Sciences, Bangalore, India
- Terry Gaasterland, UC San Diego, USA & Scripps Institution of Oceanography, USA
- Alberto Apostolico, Georgia Institute of Technology, USA & University of Padova, Italy
- Saugata Basu, Georgia Institute of Technology, USA
- USN Murty, Biology Divn IICT Hyderabad

Some Useful Phone Nos

Nitai G Dhall	-	9437024448
Laxman Mohanty	-	9437015448
Prayag Mishra	-	9861092623
Ajit Nayak	-	9861049992
Bhagwat P Chaudhary	-	9861067397

PROGRAM SCHEDULE

Friday, Dec 15

8:00-8:45	Registration
8:45-9:30	**Inauguration**
9.45-10.15	*Tea/Coffee*
10:15-11:15	Gad M Landau, University of Haifa, Israel & Polytechnic University, NY, USAIndexing Genotypes for Haplotype Search
11:15-11:45	Ravindra Pushker, UCD Conway Institute, University College Dublin, IrelandRenu Gautam, Manipal Institute of Technology, IndiaAlex Mira, Universidad Miguel Hernandez, Alicante, SpainRNAGeneTracer: a Tool to Trace Evolutionary History of the Ribosomal Genes
11:45-12:00	*Tea/Coffee*
12:00-12:30	Xiao-Dong Yang, Debiprosad Roy Mahapatra and Roderick V. N. Melnik,Mathematical Modeling & Computational Sciences,Wilfrid Laurier University, Waterloo, Ontario CanadaSimulation of RNA Silencing Pathway Based on a Generalized Amplification Model
12:30 -1:00	Ofer Gill, New York University, USANaren Ramakrishnan, Virginia Tech, USABud Mishra, New York University, USAPLANAR: RNA Sequence Alignment using Non-Affine Gap Penalty and Secondary Structure
1:00 -2:00	*Lunch*
2:00 -3:00	See-Kiong Ng, Institute of Infocomm Research, SingaporeUnraveling the Common Denominators of Protein-Protein Interaction Networks
3:00 -3:30	Simone Scalabrin, Alberto Policriti,Dipartimento di Matematica e Informatica - Universita' di Udine, ItalyMichele Morgante, Dipartimento di Scienze Agrarie ed Ambientali, ItalyStem-Loop Structure Search for Helitron Discovery
3:30 -4:00	*Tea/Coffee*
4:00 -5:00	Saugata Basu, Georgia Institute of Technology, USAAlgebraic Geometry and Computational Biology
5:00 -5:30	Samantha Kleinberg, New York University, Bioinformatics Group, USABud Mishra, New York University, USACLARITY: Algorithms for Semantic Comparison of Time-course Transcriptomic Data
5.30-6.00	Saumyadipta Pyne, Broad Institute of MIT and Harvard, USAMohammad Sajjad Hossain, SUNY at Stony Brook, USADiscriminative Meta-Analysis of Microarray Data: Combining Fission Yeast Cell Cycle Studies
6.00-7.00	*Cultural Program*
7.00	*Symposium Banquet*

	Saturday, Dec 16
9:00 -10:00	Gene Myers, Howard Hughes Medical Institute, Janelia Farms Research Campus, USAImaging-Based Systems Biology
10:00 -10:30	Rahul Siddharthan, The Institute of Mathematical Sciences, Chennai, IndiaDescribing the DNA \background"
10:30 -11:00	*Tea/Coffee*
11:00 -11:30	Marco Zantoni, Alberto Policriti, University of Udine, ItalyEmiliano Dalla, Claudio Schneider, LNCIB, ItalyFinding regulatory elements fixing error layouts
11:30 -12:00	Preetam Ghosh, Samik Ghosh, Kalyan Basu, Sajal DasUniversity of Texas at Arlington, USASimon Daefler, Mount Sinai School of Medicine, USAModeling the Diffusion process in Stochastic Event based Simulation of the PhoPQ system
12:00 -1:30	*Lunch*
1:30 -2:30	David Sankoff, University of Ottawa, CanadaUse of phylogenetic context in inferring ancestral tetraploid gene order
2:30 -3:00	Gobind Bisht, Indian Institute of Technology, Kharagpur, IndiaManju Bansal, Indian Institute of Science, Bangalore, IndiaA computational investigation of Mg-ADP inhibition of F1-ATPase and its mechanism
3:00 -3:30	*Tea/Coffee*
3:30 -4:30	Ramesh Hariharan, Strand Life Sciences, Bangalore, IndiaExtracting Information from Literature via Language Processing
4:30 -5:00	Francois-Joseph Lapointe and Claudine Levasseur,Universite de Montreal, CanadaAn Efficient and Stable Consensus Method for Combining Weighted Trees in Phylogenetic Analysis
	Sunday, Dec 17
9:00 -10:00	USN Murty, Indian Institute of Chemical Technology (CSIR), Hyderabad, Data mining applications for the control of Vector Borne Diseases
10:00 -12:00	Poster Sessions 1 & 2 (details given below)
12:00 -12:15	Closing Remarks- Nitai Gaur Dhall, Silicon Institute of Technology, Bhubaneshwar, India
12.30	*Lunch*

Poster Session-1 (Sunday, Dec 17, 10:00-12:00)

Shrishail Patil, BVDUCOE, Pune, India
Dattatraya Bormane, BVCOEW, Pune, India
Sandeep Narote, SINHGAD ENGG COLLEGE, Pune, India
Electroencephalograph Signal Analysis During Basarika Using wavelet

Bishnu Mukhopadhyay, Bratati Ghosh and Hridoy Bairagya
National Institute Of Technology, Durgapur, India
Invariant water molecular interaction in the acidophilic redox protein Rysticyanin of Thiobacillus ferrooxidans

Sumeet Dua, Pradeep Chowriappa and Ramakrishnan Rajagopalan
College of Engineering and Science, Louisiana Tech University, USA
Computational Prediction of Protein Structure Using Supervised Classification

Dharmendra Kumar, RV College of Engineering, Bangalore, India
Shweta Gupta, IOB, Bangalore, India
Sushil Middha, MACW, Bangalore, India
Pushpa Agrawal, RVCE, Bangalore, India
Sequence Analysis and Evolutionary Studies of Aquaporins

Poongavanam Sardar Maran, Vinita Vishwakarma and Manoharan N
Sathyabama Institute of Science and Technology, Chennai, India
Development of Asthma Drug Database? A Nanotechnology approach

Gadi Gayatri and Swastik Choudhury
National Institute of Science and Technology, BPUT, Behrampur, India
Use of Artificial Neural Network in Protein Structure Prediction

Rajan Mishra, Yogesh Joshi and Sujay singh
IMGENEX INDIA Pvt. Ltd., Bhubaneswar, India
Bioinformatics Strategies for Better Understanding of Antibody Development

Syamal Kumar Dana, Indian Institute of Chemical Biology, Kolkata, India
Chin-Kun Hu, Academia Sinia, Taipei, Taiwan
Neuron-like Spiking and Bursting in Josephson junction

Madhumita Patra and Chhabinath MandalIndian
Institute of Chemical Biology, Kolkata, India
Search for binding site using different computational tools

Poster Session-2 (Sunday, Dec 17, 10:00-12:00)

Bishnu Mukhopadhyay, Bratati Ghosh and Hridoy Bairagya
National Institute of Technology, Durgapur, India
Modeling Studies of the Rysticyanin Cytochrome C4 Complex Involving in the Thiobacillus ferooxidans During the Electron Transfer Reaction

Dharmendra Kumar, RV College of Engineering, Bangalore, India
Sushil Middha, MACW, Bangalore, India
Biplab Bhattacharjee, Kavita Iyer, BTM, Bangalore, India
Shweta Gupta, IOB, Bangalore, India
Pushpa Agrawal, RVCE Banaglore, India
Sequence Analysis of H5N1 & H1N1 for Ascertaining Test Systems

---

Virendra Gomase and Karbhari Kale
Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, India
In silico sequence analysis of coat protein from garlic mosaic virus

---

Sainty C . Thomas, Raunak Guha, Anand Poddar and Ankit Sehgal
Sathyabama Institute of Science and Technology, Chennai, India
A comprehensive web portal on bioinformatics

Pallavi Somvanshi, Biotech Park, Lucknow, India
Genome Wide Identi¯cation of Vaccine Candidates in Antigenically Distinct Variants of Denguevirus: A Bioinformatics Approach

---

Kinshuk Nayak and Subrata Kumar Das
Institute of Life Sciences, Dept of Biotechnology, Bhubaneswar, India
Analysis of codon usage variation in a hyperthermophilic bacterium Nanoarchaeum equitans

---

Santhosh Srinivasan and Venkatakrishnan A.J.
Vellore institute of technology, Vellore, India
Perlpred - A Novel Approach for Protein Structure Identification

---

Anjali Mohapatra, ITER, Bhubaneswar, India
P M Mishra, OMC Ltd., India
S Padhy, Utkal University, Bhubaneswar, India
Motif Search using Dynamic Hybrid Tree

Kaouther NOUIRA and Abdelwahed Trabelsi
Business & Economic Statistics Modelling Laboratory, Tunis University, Tunisia
Temporal Multi-Agent System in Intensive care Monitoring

Virendra Gomase, Karbhari Kale,
Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, India
Satish Hatture, Department of Biotechnology, SCB, Solapur, India
Bioinformatics based antigenic prediction of viral capsid protein

---

D Ashok Reddy and C K Mitra
Department of Biochemistry, University of Hyderabad, Hyderabad-500046
Analysis of gene regulatory elements using mutual information

---

Suma Jaini
Indian Institute of Technology Kharagpur.
Are Serotonin Receptor Genes under Circadian Control?

Abstracts
(Invited Talks)

Indexing Genotypes for Haplotype Search

Gad M Landau, *University of Haifa, Israel & Polytechnic University, NY, USA*

We suggest an efficient solution to the following subset matching problem. Given a dictionary D of genotypes, our goal is to preprocess all the genotypes in D such that given a query haplotype H, we answer if H can be inferred from one (or more) of the genotypes in D. Furthermore, for every such haplotype we would like to know how many genotypes it resolves.

Indexing the possible haplotypes as well as gathering statistical information about them can speed up various algorithms for haplotyping such as “Clark's rule" based algorithms, “pure parsimony", and inferring haplotypes from a set of Trios.

Joint work with Dekel Tsur and Oren Weimann

Algebraic Geometry and Computatioal Biology

Saugata Basu, *Georgia Institute of Technology, USA*

The talk will survey some recent work done by Bernd Sturmfels and his co-authors on applications of techniques from algebraic geometry in computational biology.
In particular, we discuss the connections between the traditional statistical techniques used in biological applications such as sequence analysis, and the corresponding notions in algebraic geometry.

Unraveling the Common Denominators of Protein-Protein Interaction Networks

See-Kiong Ng, *Institute of Infocomm Research, Singapore*

Cellular processes are largely mediated by complex interactions between various proteins. As such, protein-protein interactions have been routinely determined in the laboratories for decades, culminating with the recent advent of high throughput experimental techniques that have enabled the detection of protein interactions en masse. The availability of such large sets of interaction data, readily retrievable from public databases for an increasing number of genomes, can facilitate the discovery of new biological knowledge using novel data mining methods. In this talk, we will present our recent attempts to uncover the potential building blocks of protein-protein interactions by unraveling the possible common denominators of protein interaction networks at the submolecular level (domain-domain interactions) as well as the subgraph level (network motifs). We will show how the protein-protein interaction network can be a rich source of computationally challenging problems for the computer scientists and a gold mine of biologically interesting knowledge for the biologists.

Imaging-Based Systems Biology

Gene Myers, Ph.D.
Howard Hughes Medical Institute, Janelia Farms

Arguably the most significant contribution of the human genome project is that we can now build a recombinant construct of every gene and every promotor in C. elegans (worm), D. melanogaster (fly), M. musculus (mouse), and H. sapiens (human). These include fluorescent proteins and other markers that can be induced at controlled time points via a change in temperature, light, or chemistry. Combined with tremendous advances in light and electron microscopy in recent years, I believe we are now poised to visualize the meso-scale of the cell, and development and small organs (e.g. a fly's brain) at the resolution of individual cells.

Toward this end, my group is working on a number of preliminary imaging projects along these lines. These include (a) studies of development and gene expression in worms and flies, (b) the biophysics of mitosis, (c) neural patterning in flies and mice, and (d) the interpretation of signals from a new sub-wavelength resolution light microscope. We describe preliminary results on limited data sets and extrapolate on what we might be able to infer from such data. We further speculate on the potential implications of such work for the future of molecular biology.

Use of phylogenetic context in inferring ancestral tetraploid gene order

David Sankoff, *University of Ottawa, Canada*

Some genomes are known to have incurred a genome doubling (tetraploidization) event in their evolutionary history, and this is reflected today in patterns of duplicated segments scattered throughout their chromosomes. These duplications may be used as data to ``halve" the genome, i.e., to reconstruct the ancestral genome at the moment of tetraploidization, but the solution is often highly non-unique. To resolve this problem, we adapt the genome halving algorithm of El-Mabrouk and Sankoff to take account of the phylogenetic context of the genome. This involves integrating the halving algorithm with the rearrangement median algorithm. We apply this to reconstruct the tetraploid ancestor of maize, where the phylogenetic context is provided by the rice and sorghum genomes.

Extracting Information from Literature via Language Processing

Ramesh Hariharan, *Strand Life Sciences, Bangalore, India*

The talk will describe experiences at STrand Life Sciences in building a Natural Language Processing Engine to extract information (for instance, interactions between proteins) from literature. The talk will describe issues and challenges encountered.

Data mining applications for the control of
Vector Borne Diseases

Dr U.S.N. Murty
Deputy Director/Head Biology Division
Indian Institute of Chemical Technology (CSIR)
Hyderabad 500007
murty_usn@yahoo.com

Gaining a better understanding of the distribution of vector diseases within endemic regions is increasingly recognized as being critical to the rational design and monitoring of mosquito control programs.

Information on vector borne disease and use of computer applications are easier with the advancement of Information Technology (IT). Standard modern computer information systems and communication technology has an important role because of the abilities for the direct operative verification of the results. Different data mining applications have been successfully employed to study the different areas in vector borne diseases. Filaria, Malaria and Japanese Encephalitis are important vector borne diseases mostly prevalent in this country. The following areas are taken up as case studies in different vector borne diseases.

- Classification and identification
- Prioritization
- Prediction of out break/PMHD

Classification of principal vector is first and foremost in control operations. A menu-driven software was developed to identify 54 Indian Anopheline mosquitoes by using expert system. ID3 algorithm was employed which works based on if..then rules. Similar algorithm was used to identify the Culex species from South-East India. A powerful datamining tool Self Organizing Map was applied to prioritize the endemic zones of filariasis at village level in Andhra Pradesh. By this application several thousands of villages can be clustered into different grids and prioritized the endemic villages with reference to parameters responsible for promoting the disease. Advanced forecasting to predict the outbreak of disease is an emerging area which has great applied value in control operations.

JEBNET an innovative software was developed to predict the possible outbreak of Japanese Encephalitis in Andhra Pradesh. Details are discussed.

Abstracts
(Papers)

RNAGeneTracer: a Tool to Trace Evolutionary History of the Ribosomal Genes

Ravindra Pushker
Clinical Bioinformatics Lab, UCD Conway Institute, University College Dublin
Belfield, Dublin 4, Dublin, Ireland

Renu Gautam
Department of Computer Science & Engineering, Manipal Institute of Technology
Manipal - 576104, Karnataka, India

Alex Mira
Evolutionary Genomics Group, Universidad Miguel Hern´andez, Apartado 18, 03550
San Juan de Alicante, Alicante, Spain

The genes in a bacterium can be divided into proteincoding genes and ribosomal genes. The latter are involved in the translation of the former into proteins and are vital for the survival of the cell. Thus, all bacteria contain genes that code for the different subunits of the ribosome, named 5S, 16S and 23S, the number being related to the molecular weight of the ARN product. These 3 genes are normally integrated into the same operon (they are physically linked to each other in the genome) and different bacterial species contain between 1 and over 10 copies per genome. These copies from the same species are almost identical (generally over 98% identical in DNA sequence identity) and must have been originated by duplication from a common ancestor.

Therefore, the 16S-23S-5S gene system is a very good study case to explore the process of gene duplication. We have developed RNAGeneTracer, a tool to trace evolutionary history of the ribosomal genes. We concentrated on two cases where many genomes are fully sequenced, and where the relationships among species or phylogeny is well known, namely the Gammaproteobacteria (to which the model organism Escherichia coli belongs) and the Firmicutes or Low-GC Gram positives (to which the model organism Bacillus subtilis belongs). In these two groups, RNAGeneTracer identifies homologous ribosomal operons by looking at the flanking genes. Those ribosomal operons with the same flanking genes in two species will be considered homologous. Homologous flanking genes are identified by the BLAST search. Those operons present only in one species are considered to have arisen by duplication. RNAGeneTracer is available on line at http://egg.umh.es/rnagenetracer/ .

Simulation of RNA Silencing Pathway Based on a Generalized Amplification Model

Xiao-Dong Yang[1,2], Debiprosad Roy Mahapatra[1] and Roderick V. N. Melnik[1]

[1]Mathematical Modeling & Computational Sciences, Wilfrid Laurier University, Waterloo, Ontario Canada N2L 3C5
[2]Department of Engineering Mechanics, Shenyang Institute of Aeronautical Engineering, Shenyang 110034, China

The synthesis of dsRNA is analyzed using a generalized model of pathway with amplifications. The dynamics of the transgene induced RNA silencing is investigated based on a system of coupled ordinary nonlinear differential equations which describes the model phenomenologically. The transgene influx is assumed constant or an exponential function of time. The equilibrium solutions of the system are presented to study the contribution of amplification parameters and transgene influx rate. The bifurcation diagrams are presented to demonstrate nonlinear dynamics of the system.

PLANAR: RNA Sequence Alignment using Non-Affine Gap Penalty and Secondary Structure

O. GILL
Courant Inst., NYU;
E-mail: gill@cs.nyu.edu.

N. RAMAKRISHNAN
Virginia Tech.;
Email: naren@cs.vt.edu

B. MISHRA
Courant Inst., NYU.;
Email: mishra@nyu.edu

An important component of bioinformatics research is aimed at finding evolutionary relationships among species, since it allows us to better understand various important biological functions as they emerged in these species. These tools simultaneously trace the associated evolutionary history. In this context, sequence alignment is commonly used to understand similarities among the species by comparing their genomic, transcriptomic and proteomic sequence data. In our earlier research, we devised PLAINS (Piecewise- Linear Alignment with Important Nucleotide Seeker) to align genomic DNA sequences using a general class of gap distribution model. This paper proposes PLANAR (Piecewise-Linear Alignment for Nucleotides Arranged as RNA), a variant of PLAINS designed for aligning RNA sequences, while properly accounting for their secondary structures. This paper presents an overview of the PLANAR algorithm, and compares it to other competing RNA alignment tools, while emphasizing many interesting correlations discovered in the process.

Stem-Loop Structure Search for Helitron Discovery

Simone Scalabrin and **Alberto Policriti**
{scalabrin,policriti}@dimi.uniud.it, Department of Mathematics and Computer Science
via delle Scienze 206, 33100 Udine (Italy)}

Michele Morgante
michele.morgante@uniud.it, Department of Crop Science and Agricultural Engineering
via delle Scienze 208, 33100 Udine (Italy)

The base pairing of a nucleic acid secondary structure is a sort of biological palindrome. Typical secondary structures found in DNA sequences are those that form so called hairpins or stem-loop structures. The base pairs of nucleic acid stems nest in a palindromic fashion with complementary base pairings rather than identical letters. An interesting biological example of such a structure can be found in hairpins at the 30 end of helitrons, a recently discovered kind of transposable elements. At the moment, the problem of finding stemloops is solved through dynamic programming (Hofacker et al. 1994; Zuker 2003) or in linear time with the use of pattern matching and suffix trees (Gusfield 1997; Horesh et al. 2003). We propose an algorithm for stem-loop structure search which exploits matching statistics, is linear on the size of the input string S, and does not suffer for space inef-ficiecies since it builds the suffix tree only for S and not for SS R (where S R is the complement reversed string of S), resulting in a great saving of memory, highly valuable in long sequence scan.

CLARITY: Algorithms for Semantic Comparison of Time-course Transcriptomic Data

Samantha Kleinberg and **Bud Mishra**
New York University Bioinformatics Group
715 Broadway, New York, NY, USA

Imagine a personalized medical application, built upon timecourse gene expression microarray data for a group of patients. Here, by mining shared process-level patterns, implicit in these data, one could derive a better understanding of the etiology and taxonomy of the disease, disease prognosis, and choice of the therapy. For instance, in breast cancer patients, processes involved in ductal-tree reconstruction, or lumen formation, and the variability in the processes involved in their execution, could not only elucidate how breast cancer evolves, but also help in choosing the intervention to the disease. Other non-biomedical applications of similar nature abound. In this context, rather than looking at each dataset one by one, or continually comparing data sets in pair (e.g., case and control), we propose a general form of comparison among groups of data, and its efficient algorithmic implementation as well as efficient representation for the storage of these data and the results of comparisons for future use. The results are automatically updated as new work is added, without the need for manual re-evaluation. With this method, available as the software package CLARITY, researchers can update their assumptions and findings as new datasets become available. Instead of reevaluating manually, we save our results, and then automatically adapt them to new findings as they arrive. This paper formulates a key algorithmic problem at the core of this comparison process, and suggests an efficient implementation using dynamic programming.

Discriminative Meta-Analysis of Microarray Data: Combining Fission Yeast Cell Cycle Studies

Saumyadipta Pyne1, Mohammad Sajjad Hossain2

[1]Computational Biology and Bioinformatics, Broad Institute of MIT and Harvard
[7]Cambridge Center, Cambridge, MA 02141. USA. Email: spyne@broad.mit.edu
[2]Department of Computer Science, SUNY at Stony Brook, NY 11794. USA.

High-throughput microarray technology has revolutionized the way gene expression studies are done and have greatly influenced fields like genomics, medicine and bioinformatics. As large number of genome-wide data sets from similar or related microarray experiments accumulate in the databases, integrative analysis of the published data becomes increasingly essential and useful for identifying features of potential biological interest. Given the existing heterogeneity among microarray experiments with respect to various factors (e.g., the underlying platforms) and the resulting differences in their output, we propose two techniques for discriminative meta-analysis of multiple microarray data sets measuring gene expression under similar or related conditions. Ten recently published microarray data sets from three labs studying genome-wide cell cycle expression in fission yeast were combined with the techniques and a final list of 1,000 cyclic transcripts were validated by enrichment analysis with annotated data.

Describing the DNA "background"

Rahul Siddharthan

The Institute of Mathematical Sciences, CIT Campus, Taramani, Chennai 600 113, India

DNA consists of protein-coding genes, RNA-coding genes, regulatory regions, and miscellaneous non-coding DNA. It is widely assumed that the statistical properties of most non-coding DNA can be described by a generic "background model", and that these statistical properties differ significantly from the properties of genes and regulatory sites. Here we describe the DNA background and its statistical properties in different regions of the genome and in different genomes, review the effect of using non-trivial Markovian backgrounds in sequence analysis programs (such as motif-finders and sequence-alignment programs), and argue that a better understanding of the statistical properties of DNA will greatly aid the task of sequence analysis.

Finding regulatory elements fixing error layouts

M. Zantoni and **A. Policriti**
University of Udine
Dipartimento di Matematica e Informatica
{zantoni,policriti}@dimi.uniud.it
E. Dalla and **C. Schneider**
LNCIB - Trieste
Area Science Park
{dalla,schneider}@lncib.it

Motif discovery abstracts many problems encountered during the analysis of biological sequence data, where the sequences can be made by nucleotides or protein molecules and motifs represent short functionally important patterns. In this work we focus on a new computational approach to the problem of Transcription Factor Binding Sites (TFBS) identification. The task is the search for genomic motifs responsible for the binding of transcription factors to Promoters and other regulative elements: the major event underlying gene expression control. We tackled the problem designing data structures and algorithms for solving a consensus problem under an error fixed layout hypothesis, that seems to be biologically plausible and makes the task computationally treatable.

Modeling the Diffusion process in Stochastic Event based Simulation of the PhoPQ system

Preetam Ghosh[1], Samik Ghosh[1], Kalyan Basu[1], Sajal K Das[1] and Simon Daefler[2]
[1]University of Texas at Arlington, USA
[2]Simon Daefler, Mount Sinai School of Medicine, USA

The challenge today is to develop a modeling paradigm that integrates structural, molecular and genetic data for a quantitative understanding of physiology and behavior of biological processes at multiple scales - starting from cell, to tissues and finally to the whole organism. This modeling requires techniques that maintain a reasonable accuracy of the biological process and also reduces the computational overhead. This objective motivates the use of new methods that can transform the problem from energy and affinity based modeling to information based modeling. To achieve this we transform all dynamics within the cell into a random event time, which is specified through an information domain measure like probability distribution. This allows us to use the "in silico" stochastic event based modeling approach to find the molecular dynamics of the system. We present the discrete event framework concept using the example of the signal transduction cascade triggered by extra-cellular Mg2+ concentration in the two component PhoPQ regulatory system of Salmonella Enterica serovar Typhimurium. We also present a model to compute the information domain measure of the molecular transport process by estimating the statistical parameters of inter-arrival time between molecules/ions coming to a cell receptor as external signal. This model transforms the diffusion process into the information domain measure of stochastic event time to get the distribution of the Mg2+ departure events.

A computational investigation of Mg-ADP inhibition of F1-ATPase and its mechanism

Gobind Singh Bisht1, Manju Bansal[2]

[1]Department of Biotechnology, Indian Institute of Technology, Kharagpur
[2]Molecular Biophysics Unit, Indian Institute of Science, Bangalore

Nature employs several biomolecular motors to carry out some of its very essential functions of life, out of which some have shown a potential for applications in nanotechnology. One of such proteins is F1-ATPase which has lately been hybridized with mechanical structures to rotate them using ATP as the fuel. F1 ATPase contains five different subunits with the stoichiometry alpha3-beta3-gamma-delta-epsilon, the central gamma rotor is surrounded by a hexagonal "stator "in which the three alpha and there beta subunits are arranged alternately. The catalytic sites are located at the interfaces between the beta and alpha subunits, mainly on the beta subunit, and are arranged 120 degrees apart around the gamma subunit. However its direct application in nanotechnology has seen some major hurdles, in spite of its promising abilities. Mg-ADP inhibition of F1-ATPase is one drawback that causes these motors to choke in the presence of ADP, a byproduct of ATP hydrolysis, and makes its consistent rotation difficult. The present paper discusses some computational techniques that can be used to investigate the Mg-ADP inhibition state of F1-ATPase and the mechanisms behind it. The models yet available for the operation of F1-ATPase in hydrolysis mode have been surveyed and focus has been zeroed in, onto the half closed conformation of beta subunit that F1-ATPase attains before ultimately falling into the inhibition state in presence of ADP. The inhibition state has been hypothesized to be a thermodynamically stable conformation, with support from existing models. Minimizations were run on the half closed conformation, the precursor to the inhibition state, with ADP docked into the appropriate binding site and minor conformational changes favoring stabilization of ADP docking was observed. It is thereby proposed to carry out MD simulations of this minimized structure to allow a conformational sampling of the final inhibition state. A binding energetics study has been planned to establish the thermodynamic stability of the final inhibition structure concluded from MD simulations. Also TMD simulations have been proposed to investigate some of the hypothesized mechanisms of ADP inhibition. TMD fitting masks have also been generated by analysing the two available crystal structures, PDB ID: 1BMF and IE1Q.

An Efficient and Stable Consensus Method for Combining Weighted Trees in Phylogenetic Analysis

François-Joseph Lapointe[1] and Claudine Levasseur[2]

Université de Montréal
Département de sciences biologiques
CP 6218, Succ. Centre-ville, Montréal (Québec) H3C 3J7, Canada
[1]Francois-Joseph.Lapointe@UMontreal.CA
[2]Claudine.Levasseur@UMontreal.CA

There exist a wide array of consensus methods for representing the agreement among a set of input trees, but only few methods take branch lengths into account when combining such weighted trees. Consensus techniques for weighted trees are based on optimization criteria that do not always satisfy the simplest consensus axioms. In addition, the computational cost of estimating consensus trees with branch lengths is very high, compared to other methods that only take topological relationships into account. In this paper, we introduce the maximum consensus method for weighted trees that satisfies some desirable axiomatic properties, as well as being cost-effective in terms of computational efficiency. This new method based on the star decomposition of additive trees is then extended to the case of supertrees.

Abstracts
(Posters)

Electroencephalograph Signal Analysis During Basarika

S. T. Patil, BVDUCOE, Pune, India
D. S. Bormane, BVCOEW, Pune, India
Sandeep Narote, Sinhgad College of Engineering, Pune, India

Basarika is one part of the Pranayama, as traditionally conceived, involves much more than merely breathing for relaxation. Basarika is a term with a wide range of meanings. "The regulation of the incoming and outgoing flow of breath with retention." It is to be practiced only after perfection in asana is attained. Basarika also denotes cosmic power, or the power of the entire universe which manifests itself as conscious living being in us through the phenomenon of breathing. Because of this connection between breath and consciousness. Pranayama has devised basarika to stabilize energy and consciousness.

A wavelet transformation is applied to electroencephalograph (EEG) records from persons under basarika. Correlation dimension, largest lyapunov exponent, approximate entropy and coherence values are analyzed. In this project we have extensively used MATLAB, its signal processing and wavelet toolbox. It can be used by the doctors, electroencephalographers and yoga teachers for keeping track on the improvement of the persons mind, aging, balance, flexibility, personnel values, mental values, social values, love, sex, knowledge, weight reduction and body fitness.

Keywords- basarika, approximate entropy, EEG, coherence, largest lyapunov exponent, correlation dimension, wavelets.

Invariant water molecular interaction in the acidophilic redox protein Rysticyanin of Thiobacillus ferrooxidans

B. Ghosh[1], B. P. Mukhopadhyay[1,2] and H. R. Bairagya[1,2].
[1]Department of Chemistry & [2]Department of Biotechnology
National Institute of Technology, Durgapur-713209, W.B.
E-mail: bpmk2@yahoo.com , Fax: 0343-2547375

The isolation of copper from pyrite ore by ***Thiobacillus ferrooxidans*** is currently getting importance in the biotechnological area. The Cu-containing redox protein rusticyanin is thought to play a role in the electron transport chain (involving cytochrome C4 and cytochrome oxidase) and acts as primary electron acceptor following the oxidation of Fe+2 in solution of the leaching process.

Here we study the water molecular dynamics of acidophilic protein by bioinformatics & modeling using the seven mutated & nonmutated varieties of rysticyanin x-ray crystal structure from PDB.

The detail comparative analysis of the x-ray, energy minimized and auto solvated structures revealed the presence of five specific invariant water molecules (among the ~ 200 water molecules in the crystals) and their interactional scenario. Moreover, the direct water molecular coordination to the tetrahedrally bonded Cu (by His 85, 143, Met 148, & Cys 138) and their structural change induced by an invariant water molecule(to trigonal bipyramidal)seems to be interesting and the respective structural dynamics at the active metal centre could play a role in the electron transport reaction and protein-protein complexation.

Computational Prediction of Protein Structure Using Supervised Classification

Sumeet Dua, Pradeep Chowriappa and Ramakrishnan Rajagopalan

Data Mining Research Laboratory, College of Engineering and Science,
Louisiana Tech University, Ruston. USA.
E-mail: sdua@coes.latech.edu

The Protein Folding Problem revolves around the different chemical, physical, and environmental factors that enable protein conversion from a primary sequence of amino acids to secondary structures and also govern how information is encoded in them. Based on the sheer amount of data, this is a computational problem. Experimental methods for determining protein structure are difficult and time consuming, but knowing the structure of a protein can give immense amount of information about the function of a protein sequence. These impediments have given credence to Machine Learning techniques which have exhibited substantial success in this research area. We propose a Computational Data mining-based multiclass protein fold recognition system that employs Detrended Fluctuation Analysis (DFA) and four protein descriptors: Accessible

Residues, Alpha Helices, Beta Sheets, and Hydrophobic pattern frequency information. The results of the experiments conducted are encouraging and have created an interest in the way different physicochemical properties can be analyzed and used for the purpose of classification.

Keywords: multi-class protein fold classification, Hydrophobicity, learning from multiple data types, Detrended Fluctuation Analysis.

Sequence Analysis & Evolution & Studies of Aquaporins

Dharmendra Kumar, RV College of Engineering, Bangalore, India
Sushil Middha, MACW, Bangalore, India
Biplab Bhattacharjee, Kavita Iyer, BTM, Bangalore, India
Shweta Gupta, IOB, Bangalore, India
Pushpa Agrawal, RVCE Banaglore, India

The Aquaporins (AQPs) are a family of water-selective membrane channels found in animals, plants and microorganisms. Until now ten Aquaporins have been identified, including AQP0-AQP9. They are integral membrane proteins belonging to the major intrinsic proteins (MIP) family. The result of analysis has now provided new insight into the sequential & structural similarity between AQP0-AQP9. AQP4 has been taken as the key molecule since it has remarkable higher intrinsic water permeability than other water selective channels. As per the result of sequence alignment, and structural similarity, the AQP4 and AQP5 are highly similar at both gene and protein level. Phylogenetic analysis shows, the AQP4 of Homo sapiens is highly analogous with Orictolagus.

Development of Asthma Drug Database - A Nanotechnology approach

P.Sardar Maran, Vinita Vishwakarma, N.Manoharan
Sathyabama Institute of Science and Technology, Chennai, India

Data management system development for Asthma research study began with the requirements for drug delivery with nanotechnology. Special methods and advanced software's has to be used to establish an accelerated software development cycle. The development of a fully integrated data management system prior to the start of the study was not possible. Therefore an early analysis focused on identifying discrete groupings of data management functions that would allow development of distinct database modules to provide specific functionality such as subject randomization, subject registration, and data entry. The modules were categorized as either being associated with pharmaceutical companies so that the modules could be developed and delivered to meet the start date of the study. The database can be used for nanotechnology based drug delivery. Many Anti-asthma drugs are stored in the database and that can be extracted using data mining techniques.

The drug properties can be easily for this nanomedicine can be easily carried out. This paper describes the methodologies used to develop a web based, with client and server operations, and develop a data management system. It describes the circumstances that contributed to the development of these systems and the special methodologies developed. The technical aspects of the data management system and nanotechnology are presented as well as a description of the requirements and constraints analysis used to develop the hardware and software systems.

Use of Artificial Neural Network in Protein Structure Prediction

G. Gayatri* & Swastik Choudhury**

*Dept. of Computer Science & Engg. (4th Year) / **Dept. of Information Technolgy (4th Year)
National Institute of Science & Technology, Palur Hills,
Berhampur - 761008, Orissa, INDIA

What is ANN?
ANN is a constituent of Soft Computing. It is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information.

What is Protein Structure?
The graphical visualizations of protein backbones primarily show two types of secondary structures. One region corresponds to alpha helices, where backbone hydrogen bonds link residues i and i+4; another, to beta sheets, where hydrogen bonds link two sequence segments in either a parallel or anti-parallel fashion.

Why Predict Protein Structure?

The understanding of the mechanisms of disease, the identification of drug targets, and the understanding of protein interactions all require the three-dimensional structure of proteins. The challenge of determining the connection between protein structure and protein sequence is particularly critical because of our advances in the methodologies to elucidate all the genes of entire organisms, to identify the partnering of these genes in controlling cellular processes; and the well-established link between a protein's 3-D structure and its biochemical function.

Use of ANN in PSP

The prediction of protein structure is done by combining only three neural networks; an average Q3 accuracy prediction by residues of 75.93% is achieved. The networks are trained with RPROP, an efficient variation of the back-propagation algorithm. The backpropagation algorithm consists of the Feed Forward Step (the output of the nodes are computed from the input layer to the output layer) and the Back Propagation Step (the weights are updated so as to get better agreement between the computed and desired output). The application of ANN in fast, is found to be quite affective and will be dealt with at length in final paper.

Bioinformatics Strategies for Better Understanding of Antibody Development

Rajan K. Mishra[1], Yogesh Joshi[1], and Sujay Singh[1,2]

[1]IMGENEX India Pvt. Ltd., E-5, Infocity, Bhubaneswar, Orissa
[2]IMGENEX Corporation, San Diego, USA. Email - ssingh@imgenex.com

Antibodies are powerful tool to profile protein functions, interactions, cellular expression and locations and as such indispensable to modern biological research. Monoclonal antibodies are now also used for treatment of cancers. Antibodies are developed by immunizing animals either with recombinant proteins, peptides, virus, bacteria or other immunogens. Antibody production against recombinant proteins is time consuming and expensive. The use of viral and bacterial antigens to immunize animals is prohibited in many laboratories. For the development of antibodies, synthetic peptides may be used instead of recombinant proteins, viruses or bacteria. However, since the peptides are usually smaller than the native protein antigens, proper design of immunogen is important for development of effective antibodies.

A new paradigm of antibody design is now emerging, following essential discoveries in immunology and the development of bioinformatics tools for the development of B-cell epitope prediction. Computational immunology methods dramatically reduce the time and effort involved in screening potential epitopes and antigenic determinants, ranging between a reduction of 10 - 20- folds. And also the availability of protein sequences information has increased the need for accurate methods of predicting both antigenic and immunogenic sites directly from such sequences. We have developed software for epitope prediction using immuno-informatics approach.

Neuron-like Spiking and Bursting in Josephson junction

Syamal Kumar Dana[1], Chin -Kun Hu[2]
[1]Instrument Division, Indian Institute of Chemical Biology, Kolkata 700032, India
[2]Institute of Physics, Academia Sinica, Taipei 11529, Taiwan

Bursting is reported in Josephson junction. Both resistive-capacitive-inductive shunted junction (RCLSJ) and resistivecapacitive shunted junction (RCSJ) models of the superconductor junction showed bursting dynamics under the interaction of external slow motion. Most interestingly, neuron-like regular spiking (RS), intrinsic bursting (IB) and fast spiking (FS) as usually seen in mammalian neocortex are observed in the RCLSJ model under external DC bias. The origin of bursting oscillations is explained in both the models. The model has potential application in simulation of artificial neural network.

Search for binding site using different computational tools

Madhumita Patra and Chhabinath Mandal

Drug Design, Development and Molecular Modelling Division
Indian Institute of Chemical Biology,
4, Raja S. C. Mullick Road, Jadavpur, Kolkata - 700 032, INDIA
E-mail: madhumita_ucst@yahoo.co.in / cnmandal@iicb.res.in

Sialic acid binding Ig-like lectins (siglecs) are type I membrane proteins characterized by their sequence similarity and ability to bind sialic acid in glycoproteins and glycolipids. All Siglecs appear to be expressed in a cell type specific fashion. Their functions include control of myeloid cell interactions, inhibition of B cell receptor signaling, regulation of neuronal cell growth and maintenance of myelination in the nervous system. Till date there is very little data available describing their binding sites. Prediction of binding site of siglecs by computational tools can help us to understand the function of this class of proteins. In our present study we have predicted the binding site of sialoadhesin using software tools such as insightII and compared the results with the experimental findings. The structure of the functional N-terminal domain of sialoadhesin in complex with the oligosachharide 3' sialyllactose at 1.85 Å resolution was know to us (PDB ID: 1QFO). The theoretical dissociation constant values range from 10 mM - 0.1 ìM. Therefore, different binding sites may exist in solution other than the one found in the crystal structure.

Keywords: Docking, molecular dynamics, siglecs

Modeling Studies of the Rysticyanin - Cytochrome C4 Complex Involving in the Thiobacillus ferooxidans During the Electron Transfer Reaction.

B. Ghosh, B. P. Mukhopadhyay and H. R. Bairagya
National Institute of Technology,
Durgapur-713209, W.B. , India .
E-mail: bpmk2@yahoo.com, Fax: 0343-2547375

To study the probable mode of interaction during e- transfer between rusticyanin (RCy ; PDB id :- 1rcy.pdb) & cytochrome c4 (Cyc4 ; PDB id :- 1h10.pdb) , the two proteins (chain A) were docked to each other followed by energy minimization(by performing 1000 cycles of Steepest Descent and 500 cycles of Conjugate Gradient) . Two probable complexes (A & B) were obtained . The energy of complex A was -14585.98 kJ/mole while that of complex B was -15096.00 kJ / mole. In complex A , one aspartic acid residue (Asp15) of Cyc4 (near the N-terminal haem) was H - bonded to the active site Cu - bonded histidine (His143) of rusticyanin . Similarly in complex B , one glutamic acid residue (Glu121) of Cyc4 (near the C-terminal haem) was found to be H - bonded to the active site histidine residue (His143) of rusticyanin . Analysis of their energy profiles , solvent accessibilities proved that both the docked complexes were conformationally stable and energetically favourable. The DASA study shows that a decrease in accessibility occurs near the binding region of the two redox proteins in both the complexes. To find out the more probable complex among the two , each complex was solvated . After solvation , it was found that six water molecules had entered the interfacial zone in complex A and bridged the two protein molecules (RCy & Cyc4) via H-bonds , thus stabilizing the molecule. However in complex B , only two water molecules were found to be present in the interfacial zone . Complex A was hence considered to be the more probable complex due to its extra stabilization by solvent molecules . Presence of higher number of water molecules at the interfacial zone in complex may be due to the two serine residues of cytochrome c4 (Ser 13 & Ser 14) near the binding region which endow hydrophilic character at the interacting zone.

Sequence Analysis of H5N1 & H1N1 For Ascertaining Test Systems

Dharmendra Kumar

Department of Biotechnology, R V College of Engineering, Mysore Road, Bangalore. (09880533356) dk_bioinfo@yahoo.com
Sushil Kumar Middha IBMB
#31/3, 1 st main, New Gurrupan Pallya, Bennergatta Road, BTM, Bangalore
sushilmiddha@yahoo.com
Biplab Bhattacharjee IBMB
#31/3, 1 st main, New Gurrupan Pallya, Bennergatta Road, BTM, Bangalore
biplab_bhaiyu@rediffmail.com
Kavita Iyer IBMB
#31/3, 1 st main, New Gurrupan Pallya, Bennergatta Road, BTM, Bangalore
anu.i@rediffmail.com
Shweta Gupta IBMB
#31/3, 1 st main, New Gurrupan Pallya, Bennergatta Road, BTM, Bangalore
shwetkan@gmail.com

The H5N1 genome has got eight segments, each segment codes for a protein. Every segment has its own specified functions, which aid virus in infecting the cells of mammals. Segment 1 code for polymerase basic protein 2 (PB2), which plays a significant role in the virus infection. Mutations in segment 1 of H5N1 has led to some major research work. The mutation which changed glutamic acid (E) to lysine (K) at 627 position in PB2 gene produces a protein inside the capsule. What makes this result particularly significant is that an E627K mutation in PB2 has been found consistently in H5N1 avian influenza virus and it seems to play a critical role in making H5N1 so deadly. This work gave a new approach of vaccine treatment for bird flu.

In silico sequence analysis of coat protein from garlic mosaic Virus

Gomase V.S. and **Kale K.V**.

Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, 431004 (MS), India , Email- virusgene1@yahoo.co.in Mobile- +91-9226884829; Email- kvkale91@rediffmail.com Mobile- +91-9422203089, Ph-+91-240-2400431-37, Fax- +91-240-2400986, 2400291

Garlic (Allium sativum L.) is one of the most important Allium plants widely cultivated throughout the world. A significant reduction in yield and quality due to virus infection is now a serious economic problem. Garlic mosaic viruses often occur in mixed infections under field conditions. Susceptible host species of GMV are Allium ampeloprasum var. Holmense, Allium porrum, Allium sativum, Celosia argentea, Chenopodium amaranticolor, Chenopodium murale, Chenopodium quinoa, and Nicotiana occidentalis. A coat protein sequence is 289 residues long, having eight antigenic determinants in sequence. It has a Molecular weight: 32247.2 KD; Aliphatic index is 70.31, grand average of hydropathicity (GRAVY) is -0.599 and Theoretical pI is 6.16. The recombinant coat proteins showed strong immunogenicity and stimulate immune response. Small segments SILNLDHLLQYKPSQLCIS, 85-103; EMSIILNGLMVWC, 129-141; QVEFPLRPVVEHAQPTLRQIMAHFSALAE, 162-190;
of coat protein called the antigenic determinants or the epitopes are sufficient for eliciting the desired immune response, which having average antigenic propensity is 1.0020. Antigenic epitopes of coat protein of garlic mosaic virus are important determinant of protection against garlic plants. These data will be virtually significant to the further investigation of viruses infecting garlic plant, the control of garlic virus diseases and the
production of virus-freed garlic plants.

Keywords- GMV, coat protein, epitopes, immune response, hydrophilicity, hydrophobicity

A comprehensive web portal on bioinformatics

www.bioinfo.in

Anand Poddar
Sainty Thomas
Ankit Sehgal
Raunak Guha

Sathyabama Institute of Science And Technology
Deemed University
Chennai

This project is an initiative to create a comprehensive Web Portal for Bioinformatics. The website http://www.bioinfo.in has been designed so as to cater to all primary needs of a bioinformatics student and researchers. It has umpteen numbers of valuable links to the best of resources available on the web along with two salient features.

Genome Wide Identification of Vaccine Candidates in Antigenically Distinct Variants of Dengue virus: A Bioinformatics Approach

Pallavi Somvanshi[1*], P.K Seth[2] and B.N Mishra[3]
[1,2]Biotech Park, Sector G, Jankipuram Lucknow 226021, Uttar Pradesh, India
[3]Department of Biotechnology, Institute of Engineering & Technology
*Corresponding Author, Email: psomvanshi@gmail.com

The most common disease associated with human dengue virus infection is the debilitating illness dengue fever and less common but more serious is dengue hemorrhagic fever and dengue shock syndrome. Dengue virus is a member of Flaviviridae, a family of viruses containing a positive-sense, singlestranded RNA genome of about 11 kilobases (Kb). Flavivirus gene products are synthesized as a polyprotein and are subsequently proteolytically processed to mature size. Dengue virus genome encodes three structural (capsid, premembrane, and envelope) and seven nonstructural proteins (NS1, NS2a, NS2b, NS3, NS4a, NS4b, NS5) all of which are potential targets of the immune response. Individuals with naturally acquired immunity to dengue viruses have high levels of antibodies directed against NS1 and E and T cell-mediated response to NS3 and E.

Antigenic diversity in dengue virus is mainly generated through mutation and, resulting in emergence of antigenically distinct variants which cause the evasion of immune system. This work concentrates on T cell epitope analysis in dengue virus serotypes 1,2,3,4 and systematic identification of minimal set of sequences for each dengue protein that would be relevant for vaccine design against multiple antigenic variants of dengue virus, which are relevant as targets of cellular immune responses. Bioinformatics analysis is convenient for screening large number of sequences and selection of key experiments. This analysis will improve our understanding of immune response to viral variants and help identify peptide targets for vaccine formulation.

Analysis of codon usage variation in a hyperthermophilic bacterium Nanoarchaeum equitans

Kinshuk C. Nayak and **Subrata K. Das**[1*]

Institute of Life Sciences, Department of Biotechnology, Nalco Square, Bhubaneswar- 751 023, India1* E.Mail: subratkdas@hotmail.com

The fast growing sequencing of new species opens opportunities to analyze its genome evolution. The analysis of codon and amino acid data has theoretical and practical importance for understanding the basics of molecular biology (Moriyama & Hartl, 1993; McInerney, 1998). It is well known that the synonymous codon usage bias is non-random and species specific (Gramtham et al., 1981) and the pattern of codon usage vary significantly among different genes within the same taxa. In extremely AT or GC rich unicellular organisms the compositional constraints play a dominant role in shaping codon usage variation among genes (Ohkubo et al., 1987; Wright & Bibb, 1992; Gupta et al., 2002). The G+C base compositional bias and translational selection have been considered to be the most important factors that affecting the codon usage variation in prokaryotic genome (Sharp et al., 1993). In many cases translational selection determines the codon usage bias of highly expressed genes and subsequently it is advocated that preferred codons in highly expressed genes are recognized by most abundant tRNAs (Ikemura, 1981; Ikemura, 1981; Bennetzen & Hal, 1982). The pattern of codon usage in any gene reflects a complex balance among biases generated by translational selection (Bennetzen and Hall, 1982), mutation (Sueoka and Kawanishi, 2000) and random genetic drift (Sharp and Li, 1986; Bulmer, 1991 and Sharp et al., 1993) etc. Among bacteria, genomic G+C content varies over a wide range, presumably reflecting variation in mutational biases (Sueoka, 1962), with a major impact on codon usage (Muto and Osawa, 1987). Codon usage was first examined in Escherichia coli (Post and Nomura, 1980; Ikemura, 1981) and Saccharomyces cerevisiae (Ikemura, 1982; Bennetzen and Hall, 1982) where these organisms show a significant evidence of natural selection on codon usage. In addition, there are species where codon selection has been detected, but the effect seems relatively minor (Anderson and sharp, 1996 and Romero et al., 2000). Synonymous codon usage in microbial organisms is usually shaped by two major forces, the directional mutation bias (Sueoka, 1988) and the natural selection operative at the level of translation (Sharp and Li, 1987 and Gouy and Gautier, 1982). The relative strength and direction of these forces vary within and across the organisms, depending on their life style and /or genomic G+C content. Further, it has been demonstrated that gene expressivity is one of the key factor in dictating the codon usage variation among the genes of Pseudomonas aeruginosa (Gupta and Ghosh, 2001). Also it has been recently found that the codon usage pattern of Pseudomonas aeruginosa is mainly governed by translational selection rather than mutational bias (Gupta & Ghosh, 2001). Whereas, in Helicobactor pylori which is a moderate AT rich organism, only amino acid usage plays role in codon usage variation (Lafay et al., 2000). In the present study, we describe the different factors causing variation in codon usage in the genome of Nanoarchaeum equitans.

Perlpred - A Novel Approach for Protein Structure Identification

Santhosh Srinivasan and **Venkatakrishnan A.J.**
Vellore institute of technology, Vellore, India

Proteins are linear chains of amino acids which adopt a unique 3-D structure. It is this unique structure that characterizes the biochemical function to be carried out by it. Proteins adopt different levels of structures, they being primary, secondary and tertiary and one of the challenges in the world of computational biology has been the determination of these structures. This problem which represents the progression from genomics to structural genomics has been widely known as the structure prediction-inprotein folding problem. Methods for structure prediction can be divided into 3 groups namely : (1) comparative modeling, (2) fold recognition, (3) Ab initio method. Secondary structure prediction is often an important precursor in tackling protein folding problems. They have practical applications like detecting remote homologs, rational drug design etc. We present PERLPRED, a software tool for the prediction of alpha-helices and beta-sheets based on true ab initio methods and follow ASTRO-FOLD protein folding approach. The main thrusts of PERLPRED are:

(1) Identification of helical segments through partitioning of protein sequence into overlapping sub sequences and performing deterministic global optimization with free energy analysis to determine helical propensities.

(2) Prediction of beta-sheet topology and di-sulfide bridge network through the postulation of a beta strand super structure and modeling of the desolvation forces that govern hydrophobic collapse and thus generating a formulation for the beta-sheet prediction.
The advantage of the method used is that it requires no information other than amino acid sequence for prediction whereas comparative protein modeling uses previously solved structures as starting points or templates.

The first stage of the algorithm involves the prediction of helical segments and is accomplished by partitioning the overall target sequences such that consecutive oligopeptides possess an overlap of n-1 amino acids(n - length of the oligopeptide) inturn generating an emsembl of low energy conformations; calculating free energies that include entropic cavity formation, polarization, ionization contributions for each oligopeptide and thus calculating helical propensities for each residue using equilibrium occupational probabilities of helical clusters.

In the prediction of beta-sheets the hydrophobic interactions between beta strand residues are used to develop several optimization models that can be globally optimized. The first step is the postulation of a super set of possible beta segments in the target sequence. A super structure approach is used to enclose the full spectrum of possible beta strands. In contrast the strandbased formulation relies on identification of contiguous sets of residues to define potential beta strands. This is implemented computationally through the use of perl scripting language and SMMP(Simple Molecular Mechanics for Proteins) which reduces the complexities and improves the accuracy of prediction.

The results obtained from the software developed were compared for accuracy with other secondary structure prediction tools such as nn-predict and GOR. In most cases the accuracy of the structure predicted was grater than 65% and in some cases it was as high as 80%.For somatostatin, the results that were obtained were better than that obtained from nn-predict and had an accuracy of 78% compared to the PDB structure. For neurokinin, PERLPRED correctly predicts the presence of alpha- helices and beta-sheets.

This novel method of protein secondary structure prediction for alpha- helices and beta-sheets can be further extended to provide tertiary structure of the protein also.

Motif Search using Dynamic Hybrid Tree

Anjali Mohapatra[1], P. M. Mishra[2], S. Padhy[3]

Department of Computer Application, ITER, Jagamara[1]
Orissa Mining Corporation Ltd., OMC House.[2]
School of Mathematics Statistics and Computer Science, Utkal University, Vani Vihar[3]
Bhubaneswar, Orissa, India
email: anjalineel@yahoo.com[1], pmmishra@yahoo.com[2], spadhy04@yahoo.co.in[3]

An important problem in computational biology is finding repeating pattern (possibly interspersed with don't care characters) known as motifs, that occur in DNA/protein sequences. In this paper we present a new algorithm for motif search that substantially reduces the search space through the use of several heuristics. A generalized suffix tree is created from a given database of DNA sequences and then it is processed using different parameters such as weight of leaves, edge labels, number of visits of an edge and partial match to generate a dynamic hybrid tree. The dynamic hybrid tree frees the solution process from depending only on the earlier inadequate matching information and we show that the notion of biasing the current influence in the direction of a composite influence is an attractive strategy for motif search.

Temporal Multi-Agent System in Intensive care Monitoring

Kaouther Nouira, Abdelwahed Trabelsi

Business & Economic STatistics MODelling (BESTMOD) Laboratory
Institut Supérieur de Gestion de Tunis, Université de Tunis, Tunisia
kaouthar.nouira@isg.rnu.tn
abdel.trabelsi@isg.rnu.tn

This paper outlines how to improve the alarm systems from fixed threshold based systems to intelligent ones. In fact we aim to combine time series technology which can be helpful to interpret data coming from the clinical information system with multi-agent technology. We can use this system to provide an abstract description of the patient's state and to understand the response to the current therapy. In the other hand we can detect sudden shifts as outliers, level changes, change points or trends and assess the pathologic ones in order to take the appropriate decision and predict future patient state.

Bioinformatics based antigenic prediction of viral capsid protein

Gomase V.S. and Kale K.V.

Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, 431004 (MS)
Email- virusgene1@yahoo.co.in Mobile- +91-9226884829; Email- kvkale91@rediffmail.com Mobile- +91-9422203089, Ph-+91-240-2400431-37, Fax- +91-240-2400986, 2400291

Bioinformatics is being increasingly used to support target validation by providing functionally predictive information mined from databases and experimental datasets using a variety of computational tools. Fragment identified through this approach tend to be high-efficiency binders, which is a lagers percentage of their atoms are directly involved in binding as compared to larger molecules. Antigenic epitopes on Capsid protein of cucumber mosaic virus strain fny are important determinant of protection of many plants diseases. A viral capsid protein sequence is 218 residues long, having eight antigenic determinants are found in sequence. As our knowledge of the immune responses to a protein antigen progressed, it became clear that the whole protein is not necessary for raising the immune response, but small peptide segments DGASPVLVYQYAASG-145-159; RKYAVLVYSK- 181-190; ETDELVLHVDIEH-195-207 called as antigenic epitope are sufficient for eliciting the desired immune response against viral infection. Study focused on computational approach to deciphering the antigenic epitope prediction, molecular modeling and their function of coat protein of cucumber mosaic virus strain fny. In analysis predicted antigenic epitope coat protein are predicted a successful immunization strategy against various diseases.

Keywords- capsid protein, epitopes, immune response, crystal structure, hydrophilicity, hydrophobicity.

Analysis of gene regulatory elements using mutual information

D Ashok Reddy* and C K Mitra
Department of Biochemistry, University of Hyderabad, Hyderabad-500046
*E.mail: ashokmtech@yahoo.com

The present study involves three parts. First, we have studied the core promoter region in five sets (E.coli, Plants, Drosophila, Human and Mouse) of promoter sequences by calculating the average mutual information content. Here we have studied substitution matrices (both neighbor independent and neighbor dependent) for the core promoter region and calculated the information content from these substitution matrices to study the TSS-region, TATA-box, and downstream region. Neighbor independent substitutions will give 4×4 matrix and lack any preferences in other words, adjacent bases are considered independent. Next we looked at the formulae for the base pairs taken together, which correspond to a nearest neighbor preference. As we are considering a pair, there will be 16×16 matrix. These matrices include adjacent pair preferences explicitly. The results show that the TSS-region is likely to be 5-10 bases in size. We also noticed that both in the case of mouse and humans, both TATA-box and TSS-region are likely to play important roles in transcription initiation. However, in case of plant, the results showed the importance of TSS-region for transcriptional initiation compared to the TATA-box. Second, we also analyze the mitochondrial genome sequences for the transcription start sites with the information content. In this study, we concluded that the presence of short-range correlations within the TSS region is species dependent and is not universal. The information content of the mitochondrial genome near the control element computed as a 5-nucleotide overlapping block. We further noticed that there are other variable regions in the mitochondrial control element apart from the TSS. We also observed that effective comparisons can only be made on the blocks and single nucleotide comparisons does not give us any detectable signals. These results imply a similar regulatory structure in almost all organisms and have been conserved during evolution due to functional constraints. As we already know that there are well-established tools to locate conserved regions in DNA but looking for variability is also important. So we have found that information content may be useful to study the variable functional regions in genome in an efficient manner. Third, we present a new way of clustering to classify TFBS. The clustering of TFBS (JASPAR database) with information content suggests that any one of TF can bind to the any one of the corresponding clustered TFBS-class. Thus in JASPAR database, out of the 41 TFBS (in humans), perhaps only 5 -10 TFs may be actually needed and in case of mouse instead of 13 TFs, there are approximately 5 TFs are needed for gene regulation. The experimental data of TFs of specific gene expression from Transcription Regulatory Regions Database (TRRD) also coincides with our computational results. This gives us a new way to look at the protein classification-not based on their structure or function of TFs but by the nature of their TFBS.

1. D Ashok Reddy, B V L S Prasad and C K Mitra. (2006). Comparative analysis of core promoter region: Information content from mono and dinucleotide substitution matrices. Computational Biology and Chemistry, 30, 58-62.

2. D Ashok Reddy, B V L S Prasad and C K Mitra. (2006). Functional classification of transcription factor binding sites: Information content as a metric. Journal of Integrative Bioinformatics, 3(1), 0020.

3. D Ashok Reddy and C K Mitra. (2006). Comparative analysis of transcription start site using mutual information. Genomics, Proteomics & Bioinformatics. (In Press).

Are Serotonin Receptor Genes under Circadian Control?

Suma Jaini
Indian Institute of Technology Kharagpur.
sumajainibk@yahoo.co.in

Genetic analysis has identified numerous mammalian clock genes. The intracellular clock mechanism involves interaction of both positive and negative transcriptional feedback loops that drive recurrent rhythmic expression of RNA and protein levels of key clock components [27, 57]. Numerous conserved and functional E boxes, D boxes and RREs have been identified in clock and clock-related genes which are encoded in the transcriptional circuits underlying mammalian circadian clocks [63]. The present study is to locate E-boxes (CANNTG) and E' box (CACGTT) boxes in the upstream promoter region of 11 serotonin receptor genes of mammals, rats and mice and determine if their expression could be potentially regulated by clock genes.