DIMACS Center
Rutgers University

# Special Focus on Information Processing in Biology

## Annual Report

May 2005

**Participants who spent 160 hours or more**

PI: Fred Roberts, DIMACS

**Other Participants**

Ron Levy, BioMaPS, Rutgers University, Special Focus Co-Organizer

Wilma Olson, Center for Molecular Biophysics and Biophysical Chemistry, Rutgers University, Special Focus Co-Organizer

Eduardo Sontag, BioMaPS, DIMACS, Rutgers University, Special Focus Co-Organizer

**Short Course: A Field Guide to GenBank and NCBI Molecular Biology Resources**
March 1 - 2, 2005

Organizers:
Tamar Barkay, Rutgers
Paul Ehrlich, BIOMAPS Institute
Mel Janowitz, DIMACS
Tara Matise, Rutgers

**Workshop on Biomolecular Networks: Topological Properties and Evolution**
May 11 - 13, 2005

Organizers:
Cenk Sahinalp, Case Western
Petra Berenbrink, Simon Fraser

**BioMaPS/DIMACS/MBBC/PMMB/SYCON Short Course: Molecular Mechanisms and Models of Bacterial Signal Transduction**
June 6 - 10, 2005

Organizers:
Eduardo Sontag, Rutgers University
Ann Stock, UMDNJ/HHMI

**Workshop on Information Processing by Protein Structures in Molecular Recognition**
June 13 - 14, 2005

Organizers:
Bhaskar DasGupta, University of Illinois at Chicago
Jie Liang, University of Illinois at Chicago

**Workshop on Detecting and Processing Regularities in High Throughput Biological Data**
June 20 - 22, 2005

Organizer:
Laxmi Parida, IBM T J Watson Research

**Workshop on Machine Learning Approaches for Understanding Gene Regulation**
August 15 - 17, 2005

Organizers:
Christina Lesliem Columbia University
Chris Wiggins, Columbia University

Graduate students who have undertaken small research projects under support of the project.

Tiberius Bonates, RU RUTCOR, winter 04/05:
Maximum Patterns and Binarization in Logical Analysis of Data (LAD)

Miguel Mosteiro, RU CS, winter 04/05:
Sensor Networks

Bin Tian, RU Mathematics, winter 04/05:
Computer-aided Drug Discovery

Akshaya Kumar Vashist, RU CS, winter 04/05:
Rapid Automatic Extraction Groups of Orthologous Genes

**Other Collaborators**

Tamar Barkay, Rutgers, Co-Organizer, Short Course: A Field Guide to GenBank and NCBI Molecular Biology Resources

Petra Berenbrink, Simon Fraser, Co-Organizer, Workshop on Biomolecular Networks: Topological Properties and Evolution

Bhaskar DasGupta, University of Illinois at Chicago, Co-Organizer, Workshop on Information Processing by Protein Structures in Molecular Recognition

Paul Ehrlich, BIOMAPS Institute, Co-Organizer, Short Course: A Field Guide to GenBank and NCBI Molecular Biology Resources

Mel Janowitz, DIMACS, Co-Organizer, Short Course: A Field Guide to GenBank and NCBI Molecular Biology Resources

Christina Leslie, Columbia University, Co-Organizer, Workshop on Machine Learning Approaches for Understanding Gene Regulation

Jie Liang, University of Illinois at Chicago, Co-Organizer, Workshop on Information Processing by Protein Structures in Molecular Recognition

Tara Matise, Rutgers, Co-Organizer, Short Course: A Field Guide to GenBank and NCBI Molecular Biology Resources

Laxmi Parida, IBM T J Watson Research, Co-Organizer, Workshop on Detecting and Processing Regularities in High Throughput Biological Data

Cenk Sahinalp, Case Western, Co-Organizer, Workshop on Biomolecular Networks: Topological Properties and Evolution

Eduardo Sontag, Rutgers University, Co-Organizer, BioMaPS/DIMACS/MBBC/PMMB/SYCON Short Course: Molecular Mechanisms and Models of Bacterial Signal Transduction

Ann Stock, UMDNJ/HHMI, Co-Organizer, BioMaPS/DIMACS/MBBC/PMMB/SYCON Short Course: Molecular Mechanisms and Models of Bacterial Signal Transduction

Chris Wiggins, Columbia University, Co-Organizer, Workshop on Machine Learning Approaches for Understanding Gene Regulation

**Partner Organizations**

Telcordia Technologies: Collaborative Research
Partner organization of DIMACS. Individuals from the organization participated in the program planning.

AT&T Labs - Research: Collaborative Research; Personnel Exchanges
Partner organization of DIMACS. Individuals from the organization participated in the program planning and research.

NEC Laboratories America: Collaborative Research; Personnel Exchanges
Partner organization of DIMACS. Individuals from the organization participated in the program planning and research.

Lucent Technologies, Bell Labs: Collaborative Research
Partner organization of DIMACS. Individuals from the organization participated in the program planning.

Princeton University: Collaborative Research
Partner organization of DIMACS. Individuals from the organization participated in the program planning.

Avaya Labs: Collaborative Research
Partner organization of DIMACS. Individuals from the organization participated in the program planning.

HP Labs: Collaborative Research; Personnel Exchanges
Partner organization of DIMACS. Individuals from the organization participated in the program planning and research and workshop/working group organization.

IBM Research: Collaborative Research; Personnel Exchanges
Partner organization of DIMACS. Individuals from the organization participated in the program planning and research and workshop/working group organization.

Microsoft Research: Collaborative Research; Personnel Exchanges

Partner organization of DIMACS. Individuals from the organization participated in the program planning and research and workshop/working group organization.

Stevens Institute of Technology: Collaborative Research; Personnel Exchanges
Partner organization of DIMACS. Individuals from the organization participated in the program planning and research.

**Activities and Findings**

Overview

This special focus is jointly sponsored by the Center for Discrete Mathematics and Theoretical Computer Science (DIMACS), The Biological, Mathematical, and Physical Sciences Interfaces Institute for Quantitative Biology (BioMaPS), and the Rutgers Center for Molecular Biophysics and Biophysical Chemistry (MB Center). It is a follow up on DIMACS' highly successful special foci on “Computational Molecular Biology” and “Mathematical Support for Molecular Biology.”

Increasingly, many aspects of biology can be viewed as involving the processing of information. Modern information and computer science have played an important role in such major biological accomplishments as the sequencing of the human genome. On the other hand, biological ideas can inspire new concepts and methods in information science. This special focus is motivated by these two observations. The special focus activities are organized around a series of workshops with four themes:

- Algorithmic Approaches to Biological Information Processing
- Computer Science, Engineering and Biology: Applications and Analogies
- Biological Circuits and Cellular Signaling
- Proteomics.

Two of these themes represent approaches and two represent areas of application of these approaches.

**Theme 1: Algorithmic Approaches to Biological Information Processing.**

A major theme of the special focus revolves around algorithms for biological information processing. We take two points of view here. One involves how biological organisms use “algorithms” to process information and another involves how we use algorithmic methods to understand how organisms process information. The two points of view are interrelated and are reflected in three workshops.

Understanding information processing in the biological organism involves dealing with huge data sets. Modern algorithmic methods for dealing with such data sets, especially algorithms involved in pattern recognition, learning, cluster analysis, and, generally speaking, data mining, are especially relevant. Biological information processing takes advantage of regularities such as repetition, structural motifs and patterns, clustering, etc. Understanding such biological processes might, by analogy, lead us to new data mining algorithms and, in turn, methods of data mining might be useful in understanding how organisms process such regularities. One workshop, Detecting and Processing Regularities in High Throughput Biological Data, was devoted to this topic.

The massive amounts of information gathered in recent years has made it possible to study complex cellular networks using algorithmic methods of data analysis and information science. Predictions about the structure and behavior of gene regulatory networks provide a major challenge for this kind of approach. One of our workshops, Machine Learning Approaches for Understanding Gene Regulation, examined machine-learning approaches to understanding gene regulation. Modern methods of machine learning are especially appropriate given the nature of the data -- copious but noisy and incomplete -- and also provide tools that have been a major area of research at DIMACS.

**Theme 2: Computer Science, Engineering, and Biology: Applications and Analogies.**

The study of analogies between information processing in biology and information processing in computer science and engineering offers promise for the understanding of both and we investigate these analogies. More generally, we investigate applications of ideas from the biological sciences in computer science and engineering and vice versa. Such analogies and applications are a second major theme of the special focus.

**Theme 3: Biological Circuits and Cellular Signaling.**

Biochemical networks in the cell are responsible for processing environmental signals, inducing appropriate cellular responses, and sequencing internal events such as gene expression. Through elaborate mechanisms, they allow cells and entire organisms to perform their basic functions. A third theme of the special focus is the elucidation of the function and role of biological circuits and cellular signaling, with an eye to how non-biological networks can be applied to biological ones and vice versa.

**Theme 4: Proteomics.**

The fourth theme of the special focus revolves around proteomics. We seek to build on the knowledge gained from genomics to understand the activities and interactions of proteins in the cell. Studying the complete set of proteins expressed by the genome of an organism, cell or tissue type during its lifetime is a complex problem because the number of proteins is so large compared to the number of genes, because proteins can undergo numerous modifications, and because the makeup of the proteome changes frequently in response to the environment.

Understanding how information encoded in the three-dimensional structures that underlie complex protein-DNA and protein-protein network interaction is one of the fundamental challenges of biology. The workshop on Information Processing by Protein Structures in Molecular Recognition emphasized algorithms for discovery of spatial patterns, uncovering of relationships of proteins preceding the emergence of folds, and for simulating the protein-protein and protein-DNA recognition process.

**Tutorials, Workshops, and Working Groups During This Reporting Period**

*Short Course: A Field Guide to GenBank and NCBI Molecular Biology Resources*
Dates: March 1 - 2, 2005
Location: DIMACS Center, CoRE Building, Rutgers University
Organizer: Tamar Barkay, Rutgers; Paul Ehrlich, BIOMAPS Institute; Mel Janowitz, DIMACS; Tara Matise, Rutgers
Attendance: 123

The National Center for Biotechnology Information (NCBI) presented **A Field Guide to GenBank and NCBI Molecular Biology Resources**, a lecture and hands-on computer workshop on GenBank and

related databases covering effective use of the Entrez databases and search service, the BLAST similarity search engine, genome data and related resources. Further information about NCBI may be found at http://www.ncbi.nlm.nih.gov.

Now featuring the NCBI assembly and annotation of human and mouse genomes, the updated map viewer genome displays, the new genome-specific BLAST pages, the new NCBI curated conserved domains, and Cn3D 4.1.

Topics covered included:

- GenBank Database: description and scope
- The NCBI Derivative Databases: RefSeqs
- Database Searching using Entrez
    - Neighboring and Links
    - Entrez searching
- The NCBI Structures Database
    - The Molecular Modeling Database (MMDB)
    - Structural Alignments
    - Viewing Structures and Structural Alignments with Cn3D
- Similarity Searching using NCBI BLAST
    - Local Alignment Statistics
    - Scoring Systems
    - Using BLAST web services
    - PSI-BLAST
    - RPS-BLAST (CDD Search)
    - Specialized BLAST pages
- Genomic Resources at NCBI
    - Complete Microbial Genomes in Entrez
    - Higher Genome Resources
        - RefSeq and LocusLink
        - UniGene
        - Variation Data (SNPs)
        - The Human, Mouse and Rat Genomes
        - The Map Viewer
        - Other Genomes

*Workshop on Biomolecular Networks: Topological Properties and Evolution*

Dates: May 11 - 13, 2005
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: Cenk Sahinalp, Case Western and Petra Berenbrink, Simon Fraser
Attendance: 102

The functioning of a biological system largely depends on the mutual interactions among its constituent components such as proteins. It is a common practice to represent such a system by a network, within which objects are represented as nodes and relations are represented as edges linking related pairs of nodes. A biological network is broadly defined as any network (graph) where the nodes are identified with some biologically relevant entities and edges define a relation over these entities. For example, in a protein-protein interaction (PPI) network nodes correspond to proteins and edges to interactions between

them; in a metabolic network nodes usually correspond to metabolites and edges to reactions; yet another biological network may be used to describe co-occurrence of protein domains within proteins.

The structure of these biological networks resemble that of many other natural networks such as the world wide web (WWW) graph, where each vertex is a web page, and each edge is a hyperlink from one web page to another. The PPI and the WWW networks both exhibit a power law degree distribution and a small diameter. This is very different from networks generated by standard random graph models which are static and do not have power law degree distribution. Furthermore, the growth of both networks can be attributed to mechanisms of node duplication. Thus recent work on structural properties and evolution of the PPI network in conjunction with that of the WWW network has attracted considerable attention.

This workshop brought together researchers from diverse backgrounds who work on evolution and the structural properties of biological networks and how these properties relate to those observed in other natural networks. The workshop included talks on state of the art of and open questions in the following aspects of biological networks:

- Computational and experimental methods for discovering biological networks
- Evolutionary models for biological networks and their relationship to the models for WWW graph and other natural networks
- Combinatorial and statistical properties of biological network and their implications
- The structure of motifs and motif finding in biological networks

Leading specialists in the field gave invited presentations. There was a poster session and we invited poster contributions.

*BioMaPS/DIMACS/MBBC/PMMB/SYCON Short Course: Molecular Mechanisms and Models of Bacterial Signal Transduction*

Dates: June 6 - 10, 2005
Location: Busch Campus, Rutgers University
Organizers: Eduardo Sontag, Rutgers University, and Ann Stock, UMDNJ/HHMI
Attendance: 49+ (Registration still open for this workshop)

The course is a five-day intensive investigation of signal transduction divided into two related parts:

1. Basic introduction to signal transduction in bacteria for participants with extensive training in the mathematical, computational, and physical sciences but with a more limited background in molecular biology
2. Advanced reviews of current contributions to the understanding of bacterial signal transduction with an emphasis on computational approaches to modeling biological systems by leading scientists and their group members

The foundation required by non-expert researchers for the understanding of signal transduction will be provided by five leaders in the field in a series of 2-hour presentations during the first half of the course. Stanislav Shvartsman of Princeton will start the course with a basic description of signal transduction including an outline of chemical pathways and their biological significance. Igor Zhulin of the Georgia Institute of Technology will followwith a more detailed description of signal transduction in bacteria. Ann Stock of the University of Medicine and Dentistry of NJ will focus on bacterial "two-component" proteins involved in phosphotransfer signaling systems, a general mechanism of signal transduction that is widespread throughout nature. Various aspects of bacterial motility and chemotaxis are to be covered by

Robert Bourret of the University of North Carolina and Ned Wingreen of Princeton University. These introductory lectures lay the groundwork for seminars on current research in signal transduction. The lectures are designed to provide participants with a limited knowledge of molecular biology a smooth transition to the understanding and appreciation of cutting-edge research.

In the remaining presentations, participants will gain an in-depth view of signal transduction both from the content of the presentations and the perspectives provided by both molecular biologists and physicists. Speakers include Bonnie Bassler (Princeton University), William Bialek (Princeton University), Mark Goulian (University of Pennsylvania), Tom Silhavy (Princeton University), and Alexander van Oudenaarden (MIT). In addition to his introductory lecture, Robert Bourret (University of North Carolina) offer advice on the important issue of achieving effective collaborations between experimentalists and modelers.

*Workshop on Information Processing by Protein Structures in Molecular Recognition*

Dates: June 13 - 14, 2005
Location: DIMACS Center, CoRE Building, Rutgers University
Organizer: Bhaskar DasGupta, University of Illinois at Chicago, Jie Liang, University of Illinois at Chicago
Attendance: 24+ (Registration still open for this workshop)

Biological processes in cells are based on specific molecular recognitions, which triggers cascade of biological responses. The physical basis of complex network interaction is the three-dimensional structure of proteins and their functional regions. Understanding how information encoded in these biomolecules is recognized and processed by the interacting partners is a fundamental problem of biology.

In this workshop we will discuss the development of algorithms for discovery of spatial patterns important for recognition, for uncovering deep evolutionary relationship of proteins, for predicting binding partners, and for simulating the protein-protein and protein-DNA recognition process. Specific topics of interest include protein-ligand and protein-protein binding site prediction, functional prediction of proteins with known structures but unknown functions, protein-protein interactions and docking, prediction of immune epitope, design of peptide modulators of protein-protein interactons, protein substructure matching, and evolution of structural biopattern. We hope further development in these areas will formulate new research problems and motivate new algorithms in combinatorics, optimization, discrete mathematics, mathematical programming, and additional areas.

*Workshop on Detecting and Processing Regularities in High Throughput Biological Data*

Dates: June 20 - 22, 2005
Location: DIMACS Center, CoRE Building, Rutgers University
Organizer: Laxmi Parida, IBM T J Watson Research
Attendance: 21+ (Registration still open for this workshop)

The biological community is being inundated with a large amount of data and understanding this data is lagging behind the process of acquiring it. It is believed nature has left vital clues hidden in this data and there is a need for techniques and methodologies to work effectively in detecting these. Biological information processing exploits these regularities to gain understanding of the underlying model or phenomenon. For example, in its simplest form regularity could be repetition of functional or structural domains in a protein sequences or co-expression of genes in microarrays. When the data is in terms of networks, either representing protein-protein interactions or metabolic pathways, topological motifs tell a tale that will be fundamental in understanding the working of a biological system. The workshop aims to

contribute significantly to the research effort by bringing together researchers from the many different groups engaged in biological projects having the study of regularities in the data as an underlying theme.

List of Keynote Speakers:

- Alberto Apostolico, Purdue University and University of Padova
- Andrea Califano, Columbia University
- Bud Mishra, New York University
- David Mount, University of Arizona
- Andrey Rzhetsky, Columbia University
- David Sankoff, University of Ottawa

*Workshop on Machine Learning Approaches for Understanding Gene Regulation*

Dates: August 15 - 17, 2005
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: Christina Leslie and Chris Wiggins, Columbia University
Attendance: 9+ (Registration still open for this workshop)

Over the last decade, biology has been transformed into a data-driven science. Through innovations in sequencing, high-throughput microscopy, mRNA expression arrays, protein-protein and protein-DNA binding assays, and numerous other high-throughput methods, it is now possible to query simultaneously the activities of thousands of genes and their products under a wide variety of experimental conditions.

The resulting data pose an exciting challenge for the field of machine learning. Many of the model organisms (most notably S. cerevisiae) are of sufficient complexity to render detailed mathematical modeling intractable. However, it is still possible to try to learn quantitative models that are rich enough to fit data, yet simple enough to generalize and to be interpretable. Work by numerous groups suggests a promising future for more complex eukaryotes (e.g., C. elegans, S. pombe, or D. melanogaster).

Qualitatively new challenges to the machine learning community include the integration of heterogeneous datasets, such as sequence, binding, and expression data; the creation of models which are interpretable even to those not trained in probabilistic reasoning or statistical learning theory; and the presentation of the resulting models in a way useful to bench biologists as well as computational biologists.

This three-day workshop will encourage interaction among innovators in computational biology and innovators in machine learning; illuminated recent successes as well as pressing challenges; and hopefully will inspire the development of novel, biologically relevant, and biologically interpretable machine learning approaches to the current problems in biology.

**Findings**

*Screening for ortholog clusters using multipartite graph clustering*

Akshay Vashist, and Casimir Kulikowski and Ilya Muchnik

Genes related through evolution are called homologous genes. They provide a good basis for extrapolating our knowledge from well-studied organisms to new ones, as in functional annotation of their genes. An important class of homologous sequences is that of orthologous genes, or gene sequences

present in different genomes that have arisen through vertical descent from a single ancestral gene in the last common ancestor. Such genes usually perform the same function(s) in different organisms but the degree of sequence similarity across the organisms varies, and usually depends on the time elapsed since their divergence. Ortholog detection is a fundamental problem in estimating traces of the vertical evolution of genes; its practical uses include gene function annotation and finding targets for experimental studies. While many ortholog detection procedures have been proposed, they suffer from limitations that present real challenges. They may be limited to a pair of genomes, require phylogenetic information, which may not be reliable, or can handle only small sized data. The most widely used and trusted databases of orthologous groups require manual curation step(s) by experts and this is a rate limiting factor in addressing the current demand. Automatic procedures of ortholog screening, grounded on methodologies that can tackle the problem as completely as possible, while ensuring the sensitivity of the orthologous groups produced, could therefore be valuable adjuncts to speedup the process. Akshay Vashist, Rutgers University graduate student, Casimir Kulikowski, Department of Computer Science, Rutgers University, and Ilya Muchnik, DIMACS, developed a model for automatically extracting candidate ortholog clusters in a large set of genomes using a new clustering method for multipartite graphs. They designed a new kind of similarity function (linkage function) to capture the relationship between a gene and a subset of genes. The similarity relationships among genes from multiple genomes are represented as a multipartite graph, where nodes in a partite set correspond to genes in a genome. To this they apply a new clustering method for multipartite graphs. The method is fast and enables them to extract ortholog clusters from a large set of genomes. The key to the efficiency of the procedure is a particular property of the objective function, which is based on the linkage function. They evaluated the performance of their method by applying it to screening for orthologous genes in a large number of prokaryote genomes. The analyses of the results shows that orthologous clusters obtained using their approach show a high degree of correlation with the manually curated ortholog clusters in one of the most trusted databases of ortholog clusters, COG. The multipartite graph clustering method produces smaller but conserved orthologous clusters. This feature ensures that clusters contain sequences that share an orthologous relationship, and is critical to balance for the manual curation of orthologous clusters. On the other hand, related conserved clusters can be merged, using a variant of the proposed method, to obtain a desired level of aggregation. This could be useful to biologists who search genomic data to discover relationships between genes in related organisms.

**Outreach Activities**

This project is closely intertwined with DIMACS efforts to link mathematics and computer science with biology in the high schools. The project organizers were involved in planning a DIMACS conference on this subject in April 2005 (see http://dimacs.rutgers.edu/Workshops/Biomath/). Also, the project organizers will work closely with the Summer 2005 DIMACS Bio-math Connect Institute (BMCI), which is aimed at introducing high school math/CS and Bio teachers to topics at the interface. This project is informing the BMCI effort and specific topics from the project are being adapted for use in BMCI.

**Books**

**Papers**

Akshay Vashist, Casimir Kulikowski, Ilya Muchnik, “Screening for ortholog clusters using multipartite graph clustering by quasi-concave set function optimization,” to appear in *Proceedings of The Tenth International Conference on Rough Sets, Fuzzy Sets,Data Mining, and Granular Computing* (RSFDGrC 2005).

Akshay Vashist, Casimir Kulikowski, Ilya Muchnik, "Ortholog groups as clusters on a multipartite graph," submitted to *Workshop on Algorithms in Bioinformatics* (WABI 05).

Akshay Vashist, Casimir Kulikowski, Ilya Muchnik, "Automating protein function annotation through candidate ortholog clusters from incomplete genomes," in preparation.

**Talks**

Akshay Vashist, Casimir Kulikowski, Ilya Muchnik, "Screening for ortholog clusters using multipartite graph clustering by quasi-concave set function optimization," *Tenth International Conference on Rough Sets, Fuzzy Sets,Data Mining, and Granular Computing* (RSFDGrC 2005).

**Main website**

http://dimacs.rutgers.edu/SpecialYears/2005_IPB/

**Other Specific Products**

**Web pages**

***Short Course: A Field Guide to GenBank and NCBI Molecular Biology Resources***
http://dimacs.rutgers.edu/Workshops/NCBI

***Workshop on Biomolecular Networks: Topological Properties and Evolution***
http://dimacs.rutgers.edu/Workshops/Biomolecular/

***BioMaPS/DIMACS/MBBC/PMMB/SYCON Short Course: Molecular Mechanisms and Models of Bacterial Signal Transduction***
http://dimacs.rutgers.edu/Workshops/Transduction

***Workshop on Information Processing by Protein Structures in Molecular Recognition***
http://dimacs.rutgers.edu/Workshops/InformationProcessing/

***Workshop on Detecting and Processing Regularities in High Throughput Biological Data***
http://dimacs.rutgers.edu/Workshops/Detecting/

***Workshop on Machine Learning Approaches for Understanding Gene Regulation***
http://dimacs.rutgers.edu/Workshops/MachineLearning/

**Reports**

**Contributions**

**Contributions within Discipline**

This special focus is of course by nature multi-disciplinary. A major contribution is the impact on the research programs and careers of the participants. Here is a selection of comments from the participants describing this.

"I just came back from a trip abroad that started with the DIMACS workshop on Biomolecular Networks, May 11-13 (followed by a couple of other conferences). It is still premature to tell what kind of collaborations or research results it has generated, but for now I can safely say that it was one of the best scientific meetings I have been to in the last 5 years. The setting (long talks, ample time for interaction, etc.) and selection of topics and participants was outstanding and I can already say that there are many issues that came up that I will definitely pursue in my research with my students here at the Technion and possibly with additional people. Thanks again for this great experience." Ron Y. Pinter, Dept. of Computer Science, Technion, Israel)

"…the workshops at DIMACS are a very convenient way to come up to speed with what's going on in a field. I really appreciate the effort you all put in to make it possible." Matt Wiener, Merck

**Contributions to Other Disciplines**

Since the "discipline" is inherently multidisciplinary, there is no separate entry in this section.

**Contributions Beyond Science and Engineering**

Since this special focus is still very much in its early stages, we have not yet seen contributions beyond science and engineering. We expect that many of the workshops will have an impact on the medical and public health fields.

**Contributions to Human Resources Development**

Many graduate students, undergraduates, and several postdocs are participating in the program. Local graduate students and many non-local students are involved as visitors and workshop/working group attendees. More senior people are also heavily influenced by the project, being exposed to new directions of research. The impact on the careers of the students and faculty is illustrated by a few examples.

"I benefited a lot from the recent workshop on "Biological networks - topology and evolution" organized under the auspices of this special focus. It was good exposure to this developing field and an opportunity to discuss with other participants." Akshay Vashist, Rutgers University, a graduate student whose research was supported by this special focus.

"When I took the Short course, I was already a little bit familiar with Blast and NCBI page. Some parts of the lecture were more like a review for me, however, it help me to understand the way the NCBI and blast pages are organized and how the work; it answered some questions about some search tools, and gave me some "tricks" of how to do search (i.e. how to look for small sequences). The training with hands on the computer was more personal and therefore I could more specific questions about the tools, links, etc. I use the Blast search for my thesis work and do a lot of publications/reference search as well. I also use some other tools, such as looking for genomic sequences and search/study for conserved domains in proteins that I'm studying. In overall, the short course was very helpful for me and I'm applying some learned skills in my research." Maria I. Cruz, Rutgers University

“So far, my only connection with DIMACS or the Special Focus has been to attend the DIMACS training workshop on NCBI web resources earlier this spring. It was a very useful review for me, and an introduction to some sites/tools I was unfamiliar with. I also attended with an eye to training opportunities for undergraduates majoring in Genetics, and for graduate students in the joint Rutgers/UMDNJ Graduate Program in Molecular Biosciences. I think many of our students at both grad and undergrad levels would be attracted to something like this, if it can continue to be made available to them once a year (possibly many more than have been aware of it so far). Alternately or additionally, I'd be interested in more interaction with DIMACS folks for improving Genetics' undergrad course offerings.” Mark A. Brenneman, Department of Genetics, Rutgers University

In addition, the following graduate students have undertaken small research projects under support of the project.

Tiberius Bonates, RU RUTCOR, winter 04/05:
Maximum Patterns and Binarization in Logical Analysis of Data (LAD)

Miguel Mosteiro, RU CS, winter 04/05:
Sensor Networks

Bin Tian, RU Mathematics, winter 04/05:
Computer-aided Drug Discovery

Akshaya Kumar Vashist, RU CS, winter 04/05:
Rapid Automatic Extraction Groups of Orthologous Genes