

Name and brief description of initiative:**The NIGMS Funded Glue Grant Program**

Purpose: The purpose of this initiative is to make resources available for currently funded scientists to form research teams to tackle complex problems that are of central importance to biomedical science and to the mission of the National Institute of General Medical Sciences (NIGMS) but that are beyond the means of any one research group. A high level of resources is provided allowing investigators to form a consortium, and address the research problem in a comprehensive and highly integrated fashion. Five Glue grants (initiated between 2000-2003) are presently active.

Joint Informatics Efforts:

The NIGMS Glue Grant Program has held two workshops for its Bioinformatics Cores. Dr. Pamela Marino, NIGMS organized an initial workshop April, 28-29, 2002, chaired by Dr. Shankar Subramanian. The NIGMS Glue Grant principle investigators and their key bioinformatics personnel attended. The purpose of this workshop was to discuss cross-cutting issues of mutual interest in bioinformatics, provide NIGMS with information on how these issues are being resolved, and describe future needs and opportunities in bioinformatics. While the four NIGMS-supported glue grants each address distinct scientific problems, to a greater or lesser extent they all work within common model systems and generate multiple, overlapping data sets. This provides impetus for NIGMS to plan how to curate, integrate, and mine the data generated from these projects. Workshop participants identified a large number of issues and potential barriers to progress summarized as short- and long-term needs. There was general consensus as to the need for continued dialogue and sharing of ideas and information among the consortia and that this interaction should be facilitated and coordinated in some manner.

Subsequently, the directors of the bioinformatics cores of the Consortium for Functional Glycomics and the Inflammation and Host Response Glues organized a follow-up meeting (2004), of the Glue bioinformatics groups. This meeting was chaired by Rahul Raman and Charles Cooper, and held at the San Diego Supercomputing Center. Dr Eric Jakobsson, Director of the Center for Bioinformatics and Computational Biology, NIGMS attended. The purpose of the meeting was to facilitate interactions between the Glue Grants bioinformatics teams, identify areas of similar challenges and work towards common solutions. Molecule pages which act as portals to information and the handling of high through-put data (micro-arrays, proteomics) were noted as common themes of the Glues. Annotation of these data sets and development of data analysis tools to process them were discussed in depth. Further meetings of the Glue grant bioinformatics cores is left to the discretion of the core directors.

Summary for Individual NIGMS Sponsored Large Scale Collaborative Project "Glue" Grants

- i. Alliance for Cellular Signaling, AfCS (U54-GM062114)
- ii. Inflammation and the Host Response to Injury (U54 GM062119)
- iii. Protein Carbohydrate Interactions in Cell Communication, CFG (U54-GM062116)
- iv. Cell Migration Consortium (U54-GM064346)
- v. Lipid Maps (U54-GM069338)

i. Alliance for Cellular Signaling, AfCS

www.afcs.org

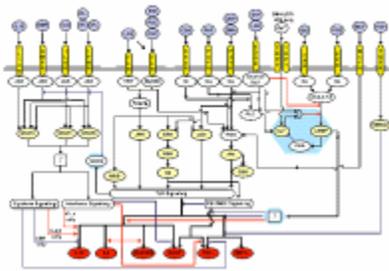
Program Director, Rochelle M. Long, Ph.D., 301-594-3827, longr@nigms.nih.gov

The overall goal of *the Alliance for Cellular Signaling* is to understand as completely as possible the relationships between sets of inputs and outputs in signaling cells that vary both temporally and spatially. The same goal, stated from a slightly different perspective, is to understand fully how cells interpret signals in a context-dependent manner. This will involve identification of all the proteins that comprise the various signaling systems, the assessment of time-dependent information flow through the systems in both normal and pathological states, and finally the reduction of the mass of detailed data into a set of interacting theoretical models that describe cellular signaling.

All AfCS legacy data and datasets are on the AfCS website (www.signaling-gateway.org). All new data developed under Alliance funding will be placed in the public domain via the data download web site (www.afcs.org). Datasets comprised of primary unprocessed data will be provided together with documentation and appropriate metadata to ensure that other investigators can use the AfCS-generated datasets. Consistent with NIH guidelines for "timely release and sharing" of data, full datasets will be available no later than the acceptance for publication of the main findings from the dataset. These datasets will include all data presented in the publication as well as all datasets supporting the conclusions presented in the publication. To complement the publication, the AfCS will periodically post additional datasets that are generated subsequent to the acceptance of the publication and that are deemed relevant to the conclusions of that study. These additional datasets will be available to the public following experimental verification. Once placed within the public domain, all data may be used by any party for research and/or commercial purposes.

The AfCS completed an analysis of context dependent signaling in RAW cells (see the recent paper). The Macrophage-like cells were exposed to 22 different ligands singly or in pairs. And the kinetics of specific protein phosphorylation, second messenger accumulation, global and specific changes in transcription and cytokine

release were all measured and the responses to single and multiple ligand addition were compared. These experiments were designed to reveal the architecture of the signaling circuitry that regulates cellular responses to physiologically relevant ligands. The goal of the AfCS is to understanding the mechanisms that underlie context dependent signaling, and thus underlie the cells complex responses to physiological and pharmacological agents. The resources of the AFCS going forward are limited and thus concentrated on defining and developing mechanistic models for the interactions involved in the operation of the subnet of G protein mediated signaling. They are focused on a few of the "outputs" or end products of G-protein mediated signaling - intracellular cAMP, Calcium, and modulation of Cytokine release.



A global analysis of cross-talk in a mammalian cellular signaling network

Natarajan M, Sternweis PC, Lin KM, Hsueh RC, The Alliance for Cellular Signaling Laboratories, and Ranganathan R
Nature Cell Biology 2006

Cellular information processing requires the coordinated activity of a large network of intracellular signalling pathways.

For informatics collaborations or to explore synergy with other initiatives contact:
Data Modeling and Network Analysis Laboratory
<http://www.signaling-gateway.org/aboutus/DataModelingLab.html>
University of California, Berkeley
Adam P. Arkin, Director

Prepared by R. Long 6/28/06

ii. Inflammation and the Host Response to Injury Large Scale Collaborative Project

<http://www.gluegrant.org>

Program Director, Scott Somers, Ph.D., 301 594-3827 somerss@nigms.nih.gov

Inflammation and the Host Response to Injury is a collaborative program that aims to uncover the biological reasons why patients can have dramatically different outcomes after suffering a traumatic injury. Known in the vernacular as a “glue” grant, this is a large-scale interdisciplinary program attempting to understand and ultimately control the life-threatening problem of inflammation following major trauma or burn. The program brings together major medical and research institutions, with researchers in the fields of surgery, genomics, proteomics, biostatistics, bioinformatics, and computational biology to focus on the molecular biology of inflammation. The scientific goal is to integrate proteomics and genomics biology of inflammation with care of the injured patient, which can only be realized through the integration and simultaneous accomplishment of multiple tasks including:

- * enrollment of sufficient numbers of trauma or burn patients with stringent entry criteria and who receive standardized care across multiple separate clinical sites
- * conducting large-scale cellular and protein assessments necessary to delineate the human immuno-inflammatory phenotype
- * determination of the gene expression profiles using state-of-the-art platforms
- * design and implementation of complex web-enabled databases of clinical, physiological, outcomes, proteomic, and genomic expression and genotype data
- * analysis of these complex data using multiple independent methodologies.

The first funding period (finishing September, 2006) has seen completion and locking of data sets containing demographic and clinical information matched with genomic and proteomic data on whole peripheral blood for both severe blunt trauma or burned patients. The second and last funding cycle is set to start October, 2006 and will continue collection of demographic and clinical data on the same types of injuries, but will perform genomic and proteomic analyses on isolated and purified peripheral blood monocytes, lymphocytes, and neutrophils.

The Information Dissemination and Data Coordination Core provides and supports capabilities for information and data collection and dissemination within the Inflammation and Host Response to Injury (IHRI) Glue Grant community of researchers as well as to external interested parties (Consortium members). Because of the interdisciplinary nature of the program - which includes such diverse fields as genomics, clinical patient care, proteomics, management reporting, and focused educational information distribution - many distinct kinds of information must be captured and stored in databases. Collectively, these systems and the data they contain are referred to as the Trauma-Related Database, or TRDB. The TRDB is a

data warehouse for data and information coming from separate but related applications and databases. Ultimately, it is the creation and support of the TRDB that is the IDDC's primary product. In most cases, data in the TRDB is accessible via the World Wide Web, using an ordinary Web browser rather than specialized software, so that the data and information generated by this research is easily available to authorized users and registered consortium members at non-IHRI research centers. Since collected data can vary in size – that is, it may be very large if selected indiscriminately – the TRDB provides the capability of obtaining data as organized information. In other cases, such as analyzing large numbers of microarray data files, specialized software runs directly on the investigator's local computer. The IDDC has developed and adapted several systems that support data capture activities. The Glue Grant Laboratory Information Management System (GLIMS) captures and tracks information on laboratory samples at sample collection and processing sites. It is the keystone system for sample and patient identification, and for tracking sample movement and processing between IHRI sites. A secure file transfer protocol (SFTP) site is maintained to support the transfer of large data files such as microarray images and for results from subsequent analytic processing while maintaining complete confidentiality. The Glue Grant Proteomics Data System (GPDS) is used to capture proteomics results. The Trial/DB clinical database system is used to capture patient clinical information; this database system was adapted from public-domain software developed by researchers at Yale University under grants. The Information Dissemination and Data Coordination Core continues to develop and adapt software and systems for capturing, presenting and tracking data. As the program's research activities result in generating large amounts of data, the IDDC continues to develop, integrate, and improve tools and services to improve researcher efficiency and effectiveness. The IDDC continues to develop systems to track and verify data files and incorporate data and analytical results into the TRDB data warehouse. With the support of software systems and the Data Curator, the IDDC does preparatory data processing to provide input to the [Computational Analysis and Modeling \(CAM\) Core](#). CAM Core results are subsequently integrated with the TRDB. The IDDC also provides program management reports regarding clinical data collection, sample tracking and processing, analytical data collection, analytical data processing and collection and processing of analytical results. The IDDC Core also designs, develops, edits, and maintains the Web site you are now viewing, as well as controlled-accessed Web pages that are only available to authorized participating investigators or registered members of the data-access Consortium. The Web site is used to publish relevant information regarding the program's research, program information and activities. The security of these Web pages and the underlying data is strictly maintained – so that all information is available only to the appropriate parties – and for the physical protection of all data, including backup copies of the data at remote sites.

The Computational Analysis and Modeling, or CAM, Core has been assigned the responsibility for processing the data generated by the other Core groups working on the Glue Grant. The purposes of these analyses are to extract knowledge predictive of patient outcomes and gain true biological understanding of the underlying

mechanisms leading to these different outcomes, as well as test the different protocols used by other cores for accuracy and reproducibility. The CAM relies heavily on known statistical methods in the broad areas of variance analysis, such as Analysis of Variance (ANOVA); clustering methods, such as Principal Component Analysis (PCA), and Hierarchical Clustering; signal-to-noise metrics such as Intra-class Correlation Coefficient (ICC) and Coefficient of Variation (CoV); and tests for significance such as t-tests, permutation tests and different cross-validations, etc. This core is also responsible for developing novel statistical approaches, either by using extant statistical techniques in novel ways, or developing new methods to analyze the data. Some of these have been the use of various filters to reduce the data sets to groups of more significant or interesting genes, Fisher Discriminate Analysis to rank the discriminatory genes, and PCA to find time-profile patterns of expression. Due to the size of the ever-expanding data set and the complexity in analysis, the CAM Core was split during the competitive renewal into a Clinical Biostatistics Core and a Data Integration Core, the latter being a flexible infrastructure to allow building small teams pairing biologists with pertinent analytical expertise.

The data collected by the program is freely available to Consortium members, but given that much of it involves human subjects, investigators must obtain local IRB approval for the specific project under consideration prior to receiving access. All programs, analytical tools, and experimental or clinical standard operating procedures are either posted on the web site or available upon request.

For informatics collaborations or to explore synergy with other initiatives contact:

Ronald G. Tompkins, M.D., D.Sc., MGH

Prepared by S Somers 06/28/2006

iii. Protein Carbohydrate Interactions in Cell Communication - The Consortium for Functional Glycomics (CFG)

<http://www.functionalglycomics.org/static/consortium/main.shtml>

Program Director, Pamela A. Marino, Ph.D 301 594-3827 marinop@nigms.nih.gov

The Protein Carbohydrate Interactions in Cell Communication - Consortium for Functional Glycomics, seeks to understand the role of carbohydrate-protein interactions at the cell surface, in cell-cell communication. To elucidation the functions of glycan-binding proteins which mediate biology at the mammalian cell surface, the Consortium integrates the efforts of its scientific cores with its approximately 300 participating investigators to address a defined set of specific aims.

A description of the cores can be found at:

<http://www.functionalglycomics.org/static/consortium/organization/sciCores/coreCtoH.shtml>

The specific aims of the CFG can be found at:

<http://www.functionalglycomics.org/static/consortium/consortium.shtml>

The Bioinformatics Component of the CFG is responsible for acquiring, storing and disseminating all Consortium related information (both raw and processed data) and creating bioinformatics tools for data mining and prediction. For this purpose it has developed a Consortium web site and constructed complex relational databases that integrate the diverse data sets generated by the Scientific Cores and Participating Investigators of the Consortium, as well as data sets from other public databases. Bioinformatics components include: A Central Database; Web-based user interfaces to the Central Database for data entry and data access via simple queries; Three specialized databases; and Bioinformatic tools. The CFG's databases and tools provide a major resource for both the Consortium and the larger scientific community. The three specialized databases integrate genomics, proteomics and glycomics-related information (both from the Consortium's Central database and other public databases) pertaining to Carbohydrate Binding Protein, Glycoenzymes, and Carbohydrate structure.

These include:

The Central Database

<http://www.functionalglycomics.org/glycomics/publicdata/home.jsp>

A powerful relational database that stores all the program information generated by the Scientific Cores and Participating Investigators. Web-based user interfaces facilitate data entry and data access via simple queries of the Central database

GBP database

<http://www.functionalglycomics.org/glycomics/molecule/jsp/gbpMolecule-home.jsp>

Glycan Binding Protein Database: Molecule pages for 150 glycan-binding proteins provide seamless access to diverse sets of data generated by the Cores in addition to information obtained from public sources. After selection of a GBP of interest from the menu, a tabbed menu interface provides facile navigation to relevant data in the CFG database and public databases. As new CFG data become available they are automatically linked to this interface.

Carb Database

<http://www.functionalglycomics.org/glycomics/molecule/jsp/carbohydrate/carbMoleculeHome.jsp>

Glycan Database: Over 7500 entries for N-linked, O-linked and synthetic glycans contain structural and chemical information, as well as PubMed links to the source articles. Glycans are displayed in the CFG symbol nomenclature, the IUPAC nomenclature and several other nomenclature formats. Multiple search options are available including a substructure search to find all glycans with a sequence of interest. Any structure pulled up can be used as the basis for another substructure search of the entire database. This database is also linked to the glycan array results for identification of hits.

Glycoenzyme database

<http://www.functionalglycomics.org/static/gt/gtdb.shtml>

This database contains a set of glycosylation pathways pages for different glycoconjugate classes. Arrows between structures link to the appropriate CAZY database page for the GT. A revised version of the database with highly annotated 'composite' structures and links to all known GT genes is near release. The enhanced version will provide users with a tabbed interface for each GT gene, analogous to the GBP database.

Bioinformatic tools for statistical analysis, data mining and to construct predictive models based on the diverse data sets ranging from molecular level to the physiology associated with GBP-carbohydrate interactions are being developed. These databases and tools provide a major resource for both the Consortium and the larger scientific community.

Participation in the cross-glycobiology grant bioinformatics meetings informed and allowed the CFG to adopt established standards for data common to multiple glycomics such as gene microarray data. The MAIME standard for meta data description, and established file formats for storing raw data collected from CFG gene microarrays, are utilized by the CFG databases. Unique data sets specific to the CFG include: glycan array binding data with cross-references to glycan structures database; annotated MALDI-MS data; and a variety of mouse phenotyping data such as histology images, and summary worksheets for hematology and immunology. The extensive utility of the unique CFG datasets, disseminated in their

current formats could potentially provide excellent data standards for such datasets. Progress has also been made in discussions between the major glycomics initiatives (CFG, EuroCarbDB, CCRC, KEGG and HGPI) to develop XML data standards for exchange of glycan structural information between relevant glycan databases developed by each of these initiatives.

The CFG Nomenclature committee evaluated symbol nomenclatures in wide use for carbohydrates and consulted with a variety of interested parties. They eventually chose to adopt for the CFG (with modification) the nomenclature originally put forth by Stuart Kornfeld, as subsequently adapted and modified (second edition) by the Editors of "Essentials of Glycobiology." In response to concerns from the Oxford group and the IUPAC committee, the CFG changed the "Essentials of Glycobiology" nomenclature colors used for symbols, such that documents can be copied or printed in black and white. The "Essentials" nomenclature as modified by the CFG is now under formal consideration by the IUPAC nomenclature committee. See: <http://glycomics.scripps.edu/CFGnomenclature.pdf>. The CFG's efforts in this area were highlighted in a C&E Cover Story which noted "A key achievement of the Consortium for Functional Glycomics (CFG) was its development last year of a new set of symbols to represent sugars. 'Until now, everyone has tended to use different symbols,' CFG Director James C. Paulson says. CFG 's nomenclature consists of a series of colored geometric symbols, one for each type of monosaccharide, and characters that indicate the stereochemistry and connection points of glycosidic linkages between sugars in oligosaccharides. The symbols can still be interpreted correctly in black-and-white copies." see: <http://pubs.acs.org/cen/coverstory/83/8332carbohydrates.html>

Standard ontology's have also been utilized for development of GBP and glycosyltransferase molecule pages which integrate genome, proteome and glycome level information pertaining to each molecule. Discussions with Dr. Shankar Subramanian and others as a part of the cross-glyc grant bioinformatics meetings as well as discussions with the CFG Bioinformatics Advisory Board enabled the development of standard ontology's for the CFG molecule pages. Standard ontology's are also used to describe knock-out mouse lines generated by Mouse Transgenic and Mouse Phenotyping Cores (these ontology's are similar to those utilized by the Jackson lab and other major mouse transgenic research labs.

The CFG uses Graphical User Interface (GUI) based forms to accept meta data and data from the Scientific Cores. Adoption of GUI's resulted from discussions with the bioinformatics cores of other glue grants such as AfCS which use similar interfaces and workflows. The CFG uses ids (in addition to the primary database keys) for the different database objects and three-tier software architecture and Oracle relational databases for implementing large scale bioinformatics platforms.

The CFG has generated molecule pages for each carbohydrate binding protein that include unique CFG data, as well as links to relevant data in public databases. The

dissemination page for the CFG's gene microarray data contains links to several online data analysis tools (downloadable in appropriate data format) for inspection of CFG generated data. See: <http://www.functionalglycomics.org/glycomics/publicdata/microarray.jsp>. Specific bioinformatics tools are being developed in collaboration with CFG bridging projects. Tools to annotate and interactively view MALDI-MS glycan profile data with cross-references to the glycan structures database are under development in collaboration with Dr. David Goldberg (PARC). Further, tools that integrate glycan structural information with appropriate glycosylation pathways, and predict structures based on gene expression of glycosyltransferases are being developed in collaboration with Dr. Minoru Kanehisa (KEGG initiative). These tools will be translated into SOAP and API interfaces for use by the public in the near future.

The CFG's bioinformatics group interacts with leading investigators and other consortia both within the US and internationally. The CFG has ongoing bioinformatics collaborations with:

EurocarbDB- <http://www.eurocarbdb.org>

GlycosciencesDE - <http://www.glycosciences.de>

CCRC - <http://www.crc.uga.edu/>

KEGG - <http://www.genome.jp/>

The CFG has organized numerous and ongoing meeting sessions and workshops, providing a platform for open discussion of analytic standards and database needs including:

2003, Dr. Anne Dell, a member of the CFG steering committee, organized a workshop on *Glycomic Databases and Associated Algorithms* for the Annual Meeting of the Society for Glycobiology. Powerpoint presentations of the workshop talks can be viewed at <http://www.bio.ic.ac.uk/research/glycobiology/>. Speakers included: Anne Dell (Imperial College, London, UK), Rahul Raman (Massachusetts Institute of Technology, Massachusetts, US), Naoyuki Taniguchi (Osaka University Medical School, Osaka, Japan), Willi von der Lieth (German Cancer Research Centre, Heidelberg, Germany), Yann Guerardel (Academia Sinica, Taipei, Taiwan), Will York (Complex Carbohydrate Research Centre, Georgia, US), Edward Dennis (University of San Diego, California, US), David Goldberg (Palo Alto Research Centre, California, US), Hailong Zhang (University of New Hampshire, New Hampshire, US), and Tony Merry (Oxford Glycobiology Institute, Oxford, UK)

2004, The CFG held a joint meeting with the Japanese Glycomics Consortia (Honolulu, Hawaii). Program: <http://glycomics.scripps.edu/pi2004agenda.pdf> Dr.'s Jim Paulson (CFG) and Naoyuki Taniguchi (Japanese Center for Glycomics and Glycotechnology) co-chaired. This meeting was an excellent forum for discussion of strategic objectives and progress made by both these Consortia. Common areas for collaboration were identified.

2005, The CFG is working with the Human Disease Glycomics/Proteome Initiative. Twenty one laboratories world-wide have analyzed a specific glycans using MS to develop standards for screening as well as glycoinformatics approaches. A workshop for this initiative was held and the meeting abstract can be found at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=16400694&query_hl=1&itool=pubmed_docsum

2005, The CFG held *its Annual Participating Investigator Meeting* at NIH. The bioinformatics session and subsequent round table discussion that were part of this meeting focused on the status and future plans for the CFG databases, and explored ways in which the CFG's efforts could be integrated into worldwide efforts to provide database resources to the glycobiology community. Many of the major databases efforts in glycobiology were represented, particularly in the carbohydrate structure and glycosyltransferase database space. Session talks provided the basis and led into a round table discussion, (chaired by Rahul Ramin and attend by NLM) of standards which need to be developed for the community. Speakers included: Rahul Raman (Massachusetts Institute of Technology) Bioinformatics and the construction of relational databases for functional glycomics; Anne Dell (Imperial College London) High throughput mouse and human glycomics; David Goldberg (Palo Alto Research Center) Automated analysis of glycan spectra; Minoru Kanehisa (Kyoto University) Integration of genomic, chemical, and pathway information for glycoinformatics; Hisashi Narimatsu (National Institute of Advanced Industrial Science and Technology) Enzymatic synthesis and high throughput structural analysis of glycans using glycogene library; Claus-W. von der Lieth (German Cancer Research Center) Informatics in glycomics: status and perspectives; James C. Paulson (The Scripps Research Institute) Functional glycomics: decoding information in the genome; Pauline Rudd (University of Oxford) Detailed analysis of the O-glycosylated regions of secretory IgA1 and gelatinase B using HPLC/MS based strategies combined with the GBI oligosaccharide structural database provides insights into immunity and metastasis.

2005, The CFG's Ram Sasisekharan and Rahul Raman organized a Satellite Meeting of the Society for Glycobiology Annual Meeting entitled: *Bioinformatics in Glycomics – an Integrated Approach to Glycans*. Invited talks addressed glycan analysis, data integration and large scale research initiatives in glycomics. Rahul Raman (CFG) and Ram Sasisekharan (CFG), in collaboration with Willi von der Lieth (EuroCarbDB) and Will York (CCRC) also organized an informal mini-informatics session following this Satellite Meeting. Other participants included: Naoyuki Taniguchi (HGPI), Minoru Kanehisa (KEGG) and Hisashi Narimatsu (National Institute of Advanced Industrial Science and Technology, Japan). This session was critical for development of a standardized XML data exchange format to interchange information on glycan structures between these leading glycomics initiatives. The proposal for this format is currently being reviewed.

2006, The CFG's Rahul Raman and James Paulson are participating along with Dr's Stein (NIST); von der Lieth (Heidelberg); York (CCRC); Zhang (UNH) and Lapadula

(UNH) in the bioinformatics session of the *Charles Warren Workshop on Glycoconjugate Analysis*. This workshop will facilitate an open discussion of current technical limitations in analytical methods and informatics tools, with an eye toward setting goals.

2006, The CFG's James Paulson and Ram Sasisekharan are serving on the organizing committee and the CFG is co-sponsoring a workshop, *Frontiers in Glycomics: Bioinformatics and Biomarkers in Disease*, being held at the NIH. This workshop convenes worldwide experts in the glycomics field to: Assess the status and future needs for bioinformatics and database resources covering the structure and biology of glycoconjugates and assess current glycan structure analysis technology for elucidating the mammalian glycome and for discovery of biomarkers in disease. see: <http://glycomics.scripps.edu/NIHBioMarker/BioMarkerProgram.html>

Major accomplishment of the CFG's bioinformatics core to date include: Development of databases and interfaces to capture and disseminate the diverse data generated by CFG and enabling integration among these diverse datasets with information from public databases via molecule pages. The database interfaces capture the depth of the data by going from data summaries to individual components such as expression of specific genes or identifying specific glycan structures which bind with high affinity to specific proteins and; Development of interfaces to facilitate online tracking of resource requests made to the CFG by the scientific community (including the participating investigators) and track research progress of the participating investigators based on the CFG data generated and publications. The CFG's databases are positioned so as to be utilized in different ways for development of data analysis and mining tools. The development of tools is already in progress and will continue to be a major focus of the Bioinformatics Core in the near future. Preparation of XML standards for representing and exchanging glycan structure information between glycomics database is in progress. Power point presentations on the CFG's ongoing efforts can be found at:

http://www.functionalglycomics.org/static/consortium/CFG_Bioinformatics_Progress_2006.ppt

These slides highlight the cross-referencing and interconnectivities between CFG databases. In addition, each of the data home pages has individual navigation demo slides in Power Point.

The May 2006 issue of *Glycobiology* has a special focus on Bioinformatics in Glycomics

<http://glycob.oxfordjournals.org/content/vol16/issue5/index.dtl>

Publications arising from this initiative include:

Goldberg D, Bern M, Li B and Lebrilla CB. [Automatic determination of O-glycan structure from fragmentation spectra](#). *J Proteome Res.* 2006;5:1429-1434.

Raman R, Venkataraman M, Ramakrishnan S, Lang W, Raguram S and Sasisekharan R. [Advancing glycomics: Implementation strategies at the Consortium for Functional Glycomics](#). *Glycobiology*. 2006;16:82R-90R.

von der Lieth C-W, Lutteke T and Frank M. [The role of informatics in glycobiology research with special emphasis on automatic interpretation of MS spectra](#). *Biochim Biophys Acta*. 2006;1760:568-577.

Taniguchi N. [From glycobiology to systems glycobiology: International network with Japanese scientists through consortia](#). *IUBMB Life*. 2006;58:269-272.

Hashimoto K, Goto S, Kawano S, Aoki-Kinoshita KF, Ueda N, Hamajima M, Kawasaki T and Kanehisa M. [KEGG as a glycome informatics resource](#). *Glycobiology*. 2006;16:63R-70R.

Lutteke T, Bohne-Lang A, Loss A, Goetz T, Frank M, von der Lieth C-W. [GLYCOSCIENCES.de: An internet portal to support glycomics and glycobiology research](#). *Glycobiology*. 2006;16:71R-81R.

Raman R, Raguram S, Venkataraman G, Paulson JC and Sasisekharan R. [Glycomics: an integrated systems approach to structure-function relationships of glycans](#). *Nat Methods*. 2005;2:817-824.

Goldberg D, Sutton-Smith M, Paulson J and Dell A. [Automatic annotation of matrix-assisted laser desorption/ionization N-glycan spectra](#). *Proteomics*. 2005;5:865-875.

Kawano S, Hashimoto K, Miyama T, Goto S and Kanehisa M. [Prediction of glycan structures from gene expression data based on glycosyltransferase reactions](#). *Bioinformatics* 2005;21:3976-3982.

von der Lieth CW. An endorsement to create open access databases for analytical data of complex carbohydrates. *Journal of Carbohydrate Chemistry*. 2004;23:277-297.

von der Lieth CW, Bohne-Lang A, Lohmann KK and Frank M. [Bioinformatics for glycomics: status, methods, requirements and perspectives](#). *Briefings in Bioinformatics*. 2004;5:164-178. (Visit [Journal Homepage](#).)

Sasisekharan R and Myette JR. [The sweet science of glycobiology](#). *American Scientist*. 2003;91:432-441.

Yao T. [Bioinformatics for the genomic sciences and towards systems biology. Japanese activities in the post-genome era](#). *Progress in Biophysics & Molecular Biology*. 2002;80:23-42. Reprinted with permission from Elsevier. (You may visit [Progress in Biophysics & Molecular Biology](#) and [ScienceDirectTM](#).)

Biochemical Society of Great Britain. 69th Symposium, University of York; 2001.
[Glycogenomics: the impact of genomics and informatics on glycobiology](#). London:
Portland Press; 2002.

All CFG publications:

<http://www.functionalglycomics.org/static/consortium/news.shtml#researcharticles>

For informatics collaborations or to explore synergy with other initiatives contact:
Principle Investigator, James Paulson 858 784-9633 jpaulson@scripps.edu
Bioinformatics Core Coordinator, Ram Sasisekharan 617 258-9494 rams@mit.edu
Bioinformatics Core Director, Rahul Raman 617 452-3257 rraman@mit.edu

Prepared by Pamela A. Marino 6/28/06

iv. Cell Migration Consortium, CMC

<http://www.cellmigration.org/index.shtml>

Program Director, James Deatherage Ph.D., 301 594-0828 deatherj@nigms.nih.gov
The Cell Migration Consortium is dedicated to accelerating progress in migration-related research by fostering collaborative, interdisciplinary research activities. The goal is to develop novel reagents, technologies, data and information. By facilitating and catalyzing interdisciplinary research on migration, the Consortium aims to develop information and reagents that will be useful for all investigators in migration-related areas. The Consortium serves to promote interactions among investigators in common areas as well as those working in different sub-disciplines, e.g., adhesion, cytoskeleton, signaling, and vesicle trafficking. It also promotes the entry of new investigators from other disciplines, e.g., modeling, imaging, biomaterials, chemistry, and proteomics, into the field of migration research through collaborative research projects, meetings and web-based interactive activities.

Bioinformatics' Objectives of the Cell Migration Consortium are:

- to support the organization, integration, and dissemination of information by providing hardware and software infrastructure to manage sequence, expression, and structural information produced by the Consortium.
- to develop WWW enabled applications to foster collaboration and share data, common and customized sequence analysis tools to identify and characterize new protein sequences, and standard relational database techniques to manage and store protein sequence and expression data.
- to build a molecule-centric knowledgebase for organizing data available on target proteins from all of the research projects.

Specific Bioinformatics efforts include:

The Cell Migration Knowledgebase (CMKB) <http://data.cellmigration.org/cmckb/>
This is a database of key facts about proteins, families, and complexes involved in cell migration. This ongoing project provides a large amount of automated and curated data, collected from numerous online resources that are updated monthly. These data include names, synonyms, sequence information, summaries, CMC research data, reagents, structures, as well as protein family and complex details.

The parent CMC Website contains sections for research data from the following CMC Initiatives: Discovery (Proteomics), Discovery (Genetic Screens), Structure, Signaling/Biosensors, Transgenic/Knockout Mice, Biomaterials, and Modeling.

The Bioinformatics group of the Cell Migration Consortium has developed tools for mass spectrometry data management, analysis, and presentation for use by the proteomics groups of the Protein Discovery Initiative. For data management, an experiment and analysis data is entered into a relational database. MS peptide results are compared to a human protein database and peptide hits are displayed in a comparison format. In addition, new search tools for identifying experiments, post-translational modifications, and protein hits have been developed. Protein identification software is available under the MIT Open Source license from Dr. Pearson wrp@virginia.edu . In addition, the Bioinformatics group has developed tools for microarray data management, analysis, and presentation for use by the Gene Expression group of the Protein Discovery Initiative.

Information of CMC modeling and modeling software can be found at:

<http://www.cellmigration.org/resource/modeling/>

http://www.cellmigration.org/resource/modeling/model_resources.shtml#softwares

For informatics collaborations or to explore synergy with other initiatives contact:
Principal Investigator, Dr. Alan R. (Rick) Horwitz Horwitz@virginia.edu

Prepared by Jim Deatherage 6/28/06

v. Lipid Metabolites and Pathways Strategy, LIPID MAPS

<http://www.lipidmaps.org/index.html>

Program Director, Jean Chin Ph.D., 301 594-2485; email chinj@nigms.nih.gov

The goal of the LIPID MAPS glue grant is to develop an integrated system capable of characterizing global changes in lipid metabolites (lipidomics) in the cell. Lipids, which are central to the regulation of cell function, the structure and function of cell membranes, and the understanding of many diseases, include the diverse classes of fatty acids and eicosanoids, glycerophospholipids, neutral lipids, sterols,

sphingolipids and glycosphingolipids, and isoprenoids. Changes in the levels and/or locations of one lipid metabolite have been shown to affect those of other lipid classes. The broad specific aims of the lipidomics glue grant are to: (a) identify and quantitate all the lipids in the macrophage cell; (b) identify and quantitate temporal lipid responses to cell activators; (c) discover and characterize novel and minor lipids; (d) discover and quantitate relationships among the levels of specific lipids, proteins, and genes; (e) create a LIPID MAPS database accessible to all scientists; and (f) discover lipid networks and generate new hypotheses about metabolism.

Two data bases have been developed:

A Structure Database <http://www.lipidmaps.org/data/structure/index.html> in which representative examples from each category of the LIPID MAPS Lipid Classification scheme can be searched by substructure http://www.lipidmaps.org/data/structure/structure_search.php

A Proteome Database <http://www.lipidmaps.org/data/proteome/index.cgi> which was published http://nar.oxfordjournals.org/cgi/content/full/34/suppl_1/D507. This is a lipid-associated protein sequences with annotations from UniProt, EntrezGene, ENZYME, GO, KEGG and other public resources. Browse or search by species, lipid class association, and/or keywords.

The consortium has developed a lipid classification scheme <http://www.jlr.org/cgi/content/short/46/5/839> comprised of eight categories, containing distinct classes and subclasses of molecules, chemically based and driven by the distinct hydrophobic and hydrophilic elements that compose the lipid.

Tools are available for analysis of mass spectrometry data that allow one to find mass, number of carbons, number of double bonds, abbreviation, MS/MS product ions (neutral loss), formula, and ion based on input criteria, with links to structure and isotopic distribution <http://www.lipidmaps.org/tools/ms/> In addition, tools to draw structures <http://www.lipidmaps.org/tools/structuredrawing/> are available via the consortium website.

Microarray data and analytical tools are downloadable from the website, <http://www.lipidmaps.org/data/microarray/codelink/index.html>, <http://biome.sdsc.edu:8090/vampire/>, <http://www.lipidmaps.org/data/results/raw2647/microarray/agilent/>, <http://www.lipidmaps.org/data/results/raw2647/microarray/goby/>, <http://www.lipidmaps.org/data/results/raw2647/microarray/annotations/>

For informatics collaborations or for exploring synergy with other initiatives, contact: Principle Investigator, Edward Dennis Ph.D at 858 534-3055 or edennis@ucsd.edu
Prepared by Jean Chin, 6/28/2006