Office of Science
Notice DE-FG01-04ER04-32

Genomics:GTL

Department of Energy

Office of Science Financial Assistance Program Notice DE-FG01-04ER04-32; Genomics:GTL

AGENCY: U.S. Department of Energy

ACTION: Notice inviting grant applications.

SUMMARY: The Office of Biological and Environmental Research (OBER) and the Office of Advanced Scientific Computing Research (ASCR) of the Office of Science (SC), U.S. Department of Energy (DOE), hereby announce their interest in receiving applications for research from large, well integrated, multidisciplinary research teams (see Supplementary Information below) that support the Genomics:GTL research program (http://www.doegenomestolife.org/). A central theme of the entire Genomics:GTL program is to develop the necessary experimental and computational capabilities to enable a predictive understanding of the behavior of microbes and microbial communities of interest to DOE. To this end, proposals that integrate strong experimental biology and computational science research components are strongly encouraged.

DATES: Statements of intent to apply, including information on collaborators and areas of proposed research and technology development should be submitted by Monday, October 25, 2004.

Formal research applications are due by 4:30 PM Eastern Time Tuesday, January 18, 2005.

ADDRESS: Statements of intent to apply should be sent to Ms. Kim Laing by email at: kim.laing@science.doe.gov with copies to Dr. David Thomassen at: david.thomassen@science.doe.gov and Dr. Gary Johnson gary.johnson@science.doe.gov.

Formal applications referencing Program Notice DE-FG01-04ER04-32 must be sent electronically by an authorized institutional business official through DOE's Industry Interactive Procurement System (IIPS) at: http://e-center.doe.gov. IIPS provides for the posting of solicitations and receipt of applications in a paperless environment via the Internet. In order to submit applications through IIPS your business official will need to register at the IIPS website. **IIPS offers the option of using multiple files, please limit submissions to one volume and one file if possible, with a maximum of no more than four PDF files.** The Office of Science will include attachments as part of this notice that provide the appropriate forms in PDF fillable format that are to be submitted through IIPS. Color images should be submitted in IIPS as a separate file in PDF format and identified as such. These images should be kept to a minimum

due to the limitations of reproducing them. They should be numbered and referred to in the body of the technical scientific grant application as Color image 1, Color image 2, etc. Questions regarding the operation of IIPS may be e-mailed to the IIPS help desk at: HelpDesk@pr.doe.gov or you may call the help desk at (800) 683-0751. Further information on the use of IIPS by the Office of Science is available at: http://www.science.doe.gov/grants.html.

If you are unable to submit an application through IIPS, please contact the Grants and Contracts Division, Office of Science at: (301) 903-5212 or (301) 903-3064, in order to gain assistance for submission through IIPS or to receive special approval and instructions on how to submit printed applications.

FOR FURTHER INFORMATION CONTACT: Dr. David Thomassen, telephone: (301) 903-9817, E-mail: david.thomassen@science.doe.gov, Office of Biological and Environmental Research, SC-72, U.S. Department of Energy, SC-72/Germantown Building, 1000 Independence Avenue SW, Washington, DC 20585-1290 and Dr. Gary Johnson, telephone: (301) 903-5800, E-mail: gary.johnson@science.doe.gov, Office of Advanced Scientific Computing Research, SC-31/Germantown Building, 1000 Independence Avenue SW, Washington, DC 20585-1290.

A complementary request for proposals from DOE national laboratory led teams has been issued http://www.sc.doe.gov/grants/LAB04_32.html.

SUPPLEMENTARY INFORMATION

This solicitation will support the establishment of large, well integrated, multidisciplinary (e.g., biology, computer science, mathematics, computational science, engineering, informatics, biophysics, biochemistry) research teams. Applicants are invited to include, where appropriate, partners from multiple institutions, including DOE National Laboratories, universities, private research institutions, and companies. Successful applications will include a detailed management plan describing the responsibility of and relationship between all participating institutions and investigators, a strategy for maximizing communication and exchange of information between investigators, a data and information management plan, and project milestones.

Research partners at individual universities, private research institutions and companies, and DOE National Laboratories may be funded directly by DOE but will be reviewed as part of the overall research application submitted by the lead research institution. To facilitate funding of individual research partners each application should include a complete set of forms for each research institution as described in the instructions contained in the Grant Application Guide, the Guide and Forms are available on the web at: http://www.science.doe.gov/grants.html. This includes:

- Signed Face Page (DOE F 4650.2 (10-91))
- Budgets for each year, (using DOE F 4620.1)
- Budget Explanation
- Biographical Sketches (limit 2 pages per senior investigator)
- Description of Facilities and Resources
- Current and Pending Support for each senior investigator

- Other institutional forms as described

Research partners at DOE National Laboratories do not need to complete a signed face page or the "other institutional forms" included in the Grant Application Guide.

Research Focus

Research funded here will cut across components of each of the goals described in the Genomics:GTL program plan, available on the web at: http://www.doegenomestolife.org and is intended to complement and advance ongoing research in the Genomics:GTL program (see http://www.doegenomestolife.org/research/index.shtml#research).

Microbes of Interest to DOE. The focus of Genomics:GTL is on nonpathogenic microbes (including fungi) directly relevant to DOE mission needs in energy (cleaner energy, biomass conversion, carbon sequestration, and the global carbon cycle both terrestrial and ocean) or the environment (immobilization of metals and radionuclides at DOE sites). When possible, research should take advantage of and focus on microbes whose complete DNA sequence is already known or microbial communities of interest to, directly relevant to, or that would contribute substantially to an ability to address DOE mission needs. Applicants should identify proposed high throughput DNA sequencing needs, if any, in their application. Applicants should also provide a clear, scientifically justified description for their choice of microbe(s) in the context of DOE mission needs as outlined above.

Data and Other Results. Data and results that are generated through these investigations that are appropriate to share with the broader community should be provided in timely, open, and machine-readable format where possible. At a minimum, data should be freely distributed to academics, including the right to redistribute. Accepting a licensing agreement that requires proper attribution is allowable. Microbial DNA sequence data will be publicly released according to the "Data Release Requirements: Microbial Genome Sequencing Projects" (http://www.sc.doe.gov/production/ober/EPR/data.html). Research plans should be included that describe the procedures and policies the teams will institute to make the data and results available and interoperable with other significant sources of relevant data. Teams should be amenable to the adoption of open data standards and interoperability requirements, as they evolve and are specified by the Genomics:GTL program.

Software Development and Distribution. Software developed by research teams that is appropriate for distribution beyond the developing team shall be made available to the biological and computational community. It is the intent of the GTL Program that this software be accessible, useful, affordable, and interoperable with other software and data to the maximum extent possible. At a minimum, executables need to be distributed freely to academics. A licensing requirement for proper attribution prior to distribution is allowable. Applications should include plans for assuring availability, stating whether: the software will be available as binary or source code, a fee will be charged for the use of the software, some users (e.g., commercial) will be charged while others not, in what way derivative products will be treated, etc. Statements such as that by the International Society for Computational Biology on Bioinformatics Software Availability, http://www.iscb.org/pr.shtml, may be used for reference.

High Throughput Analysis of Microbial Multi Protein Complexes.

The vast majority of the processes and functions of living systems involve the interactions of multiple proteins. In some cases proteins work in an independent yet coordinated manner. However, in many cases, likely the majority, proteins work as part of multi protein complexes comprised of small to large numbers of proteins and perhaps other molecules. Some complexes are tightly bound and others only loosely associated with components coupled for widely varying times. Understanding the role of individual proteins in a biological system and the overall processes and functions carried out by that system requires that we have the research tools to identify and characterize all of the multi protein complexes used by a given biological system to carry out its various processes and functions. We also need computational tools, based on experimental data that enable us to predict the functions and behaviors of complex biological systems beginning with genome sequence data.

Research is sought to demonstrate the feasibility of establishing high throughput approaches for the isolation and characterization of microbial multi protein complexes. Applicants should describe a strategy designed to increase the throughput of complex isolation and characterization in their research project so that:

- >5,000 attempts to isolate and analyze complexes per year are made after three to five years, and/or
- at least 85% of the stable multi protein complexes in a single microbe can be identified and characterized within a single year.

Applicants should consider the following items when preparing their applications though it is not expected that all items will be included in an application, that all items will be given equal weight, or that this is an exclusive list of topics for consideration:

- Methods to quantitate a cell's proteins as a function of state or time.
- Alternative methods for isolating multi protein complexes since no single method is likely to be effective for all complexes.
- Methods for stabilizing the components of unstable complexes.
- Methods for identifying the individual components that comprise a complex.
- Methods for characterizing the stoichiometry of the components that comprise a complex.
- Strategies for determining whether all the individual components required for function are present in a complex following its isolation.
- Methods for isolating and characterizing difficult complexes such as those that are short-lived or that are found embedded in membranes.
- Methods for determining the overall quaternary structure of complexes (Note: it is not the intent of this solicitation to support research on high resolution atomic-level structure determination of individual proteins or multi protein complexes though established techniques may be used in a supporting role in proposed research).
- Experimental and computational methods for determining the function of complexes.
- Methods for determining the temporal and spatial relationships of complexes within microbial cells over time and under different conditions.

- Development of predictive models or "rule sets" that can be used to predict the number and nature of multi protein complexes that a microbe is likely to form based on its genome sequence.
- Strategies for automating the isolation and analysis of complexes.
- Strategies for managing data flow, data analysis, data sharing, and sample and process management.

Milestones of annual research progress and success as well as overall project milestones should be included as part of the research plan.

Genome-Scale Analysis of Biochemical Pathways in Microbes and Microbial Communities

Microbes have capabilities and carry out processes of direct interest and importance to DOE that include energy production, the remediation of contaminants found at former weapons sites, and their central role in the global carbon cycle. Microbes or microbial communities exist that:

- normally carry out these various functions with differing efficiencies;
- can be stimulated, with the addition of a simple nutrient for example, to carry out one of these functions much more efficiently than they normally do or under environmental conditions when these functions might normally be suppressed;
- can be modified or engineered genetically to carry out a completely new function or to carry out a function much more efficiently than they normally would or under conditions when it might normally be suppressed; or
- possess several complimentary pathways for performing identical or closely related functions and select among them for the most useful and efficient depending on environmental conditions.

In each of these cases we need to understand not only the molecular/biochemical pathways directly responsible for these functions but also regulatory mechanisms used to affect the activity levels of specific pathways of interest.

Experimental and computational research is sought that will provide the knowledge and tools needed to predict and reconstruct a microbe's regulatory networks for the control of multi protein complexes, for metabolic pathways, or for the entire organism beginning with knowledge of its DNA sequence. Similarly, research is sought that will provide the knowledge and tools needed to develop computational tools to predict the metabolic, physiologic, and behavioral characteristics of microbial communities from community DNA sequence data. Research is not being sought in this Notice to study the regulation or function of single or small numbers of genes or single metabolic pathways. Applicants should describe a strategy that will, within three to five years:

- enable the construction of a computational model for a microbe or microbial community's central metabolism within a few days or weeks after the genomic sequence has been determined;
- lead to the development of predictive and testable models of regulatory networks for key processes such as environmental sensing, contaminant processing, carbon dioxide

fixation, cellulose degradation, or hydrogen or ethanol production for individual microbes or microbial communities.

Applicants should consider the following items when preparing their applications, though it is not expected that all items will be included in an application, that all items will be given equal weight, or that this is an exclusive list of topics for inclusion:

- Methods for developing "simple" connection diagrams to understand dynamic regulatory interactions between different microbial proteins, multi protein complexes, and metabolic pathways.
- Methods for including a stoichiometric component in these "connection diagrams" including, for example, a link to quantitative proteomics or metabolomics.
- Use of traditional molecular tools, such as the generation of deletion mutants and high throughput genetics, in a high throughput format to develop accurate models of key biologic processes.
- Strategies for leveraging genome sequence information from closely related microbes to identify and characterize pathway and regulatory networks.
- Strategies that build and integrate models that represent varying levels of biological or biochemical detail.
- Methods grounded in genomic and/or proteomic analyses to develop predictive capabilities for phenotypic profiling of microbes and microbial communities.
- Strategies for using computational approaches to predict the effects of adding or removing specific components to or from pathways.
- Methods to image metabolites, proteins, protein complexes, community member interactions, and gene expression in microbial communities.
- Methods for coupling data at many scales, e.g., the functional characterization of a microbial community, field studies and characterization of the environment where the community is found.
- Strategies for verifying and validating models and model predictions.
- Strategies for managing data flow, data analysis, data sharing, and sample and process management.

Milestones of annual research progress and success as well as overall project milestones should be included as part of the research plan.

Computational methods and capabilities to advance understanding of complex biological systems and predict their behavior

A central goal of the Genomics:GTL program is to develop computational models that enable an accurate prediction of the behavior of microbes and microbial communities of interest to DOE. An inherent challenge in modeling biological complexity is to make use of the diversity and volume of data that will be available for inclusion in these models. This diversity and volume may soon confront our abilities to efficiently manage, query, and analyze the data. Data types include, but are not limited to, DNA sequence and annotation, gene expression, protein expression and modification, protein interactions, gene regulation, information related to cell and

community metabolism, diverse imaging modalities, microbe-microbe interactions, microbe-environment interactions, and multiple environmental conditions all over diverse time scales.

High throughput experimental strategies that are essential to the success of Genomics:GTL will also generate very large volumes of both raw and revised data, including DNA sequence, mass spectrometry, imaging, and many others, that will be difficult to manage given current capabilities.

Research is sought that will develop a model, or series of computational models, that captures the biological complexities of the salient functions of a microbial organism of DOE interest. As a part of the model development, a broad range of biological data is expected to be used. It is expected that this effort will be very closely coordinated with other research efforts that provide much of the experimental data to populate the model. Through the application of these diverse data to the development of models and through other means as necessary, the research team will collaborate with data providers to develop data management strategies for the biological data including interfaces, common ontologies, data standards, and other tools to help make the data accessible and useful to the broader biological community. There will be a particular emphasis on nascent sources of data that are likely to grow rapidly in volume and importance.

Development of these new models will require that applicants solve, in parallel, a number of algorithmic, mathematical, and computational challenges that currently impede progress in order to develop models for complex biological systems. As part of their research projects, applicants should consider finding solutions to some of the following challenges, though it is not expected that all items will be included in an application, that all items will be given equal weight, or that this is an exclusive list of topics for inclusion:

- New data storage solutions that can accommodate large and diverse data volumes and data types and that can provide rapid data analysis, query, retrieval, and transmission.
- Methods for data representations, standards, and controlled vocabularies for the diverse, bulk data types anticipated in Genomics:GTL, including, but not limited to, gene expression, imaging, mass spectrometry, metabolic profiling, DNA sequence and annotation.
- Development of new types of databases that can accommodate large data volumes, great schema complexity, and rapid query retrieval.
- New data analysis methods and algorithms that can accommodate large and diverse data volumes and data types as they are incorporated into new biological models.
- Strategies for integrating diverse data types, e.g., gene expression, imaging, mass spectrometry, metabolism, into common models and simulations.
- Strategies of incorporating new data, of types not previously integrated, into evolving models of biological processes and functions without starting model development anew.
- Strategies for dealing with incomplete, sparse, and potentially inaccurate data in biological models.
- Methods for visualizing complex and multi-dimensional data.

Applications for this component of Notice DE-FG01-04ER04-32 will **not** include a large, independent, laboratory-based research component. It is expected that applicants will interface

their proposed modeling research with ongoing laboratory-based research projects, especially those funded by the Genomics:GTL program (http://www.doegenomestolife.org/research/index.shtml#research).

It is expected that some of the computational tools developed here will be executed on existing computer resources with little need for additional computational power. Other tools may require particularly compute-intensive resources. Special consideration will be given to the development of computational tools that can be ported across high- performance computing environments, including computing capabilities that are not yet available but are expected soon.

Appropriate attention should also be paid to attributes such as modularity, interoperability, and scalability.

Milestones of annual research progress and success as well as overall project milestones should be included as part of the research plan.

Program Funding

Up to $10 million of FY 2005 funds will be available, contingent upon availability of appropriated funds. Multi-year funding of grant awards is expected, and is also contingent upon the availability of appropriated funds, progress of the research, and continuing program need. It is anticipated that individual research grants for the first two components of this Notice (multi protein complexes and genetic regulatory network analysis) will be funded at a level of approximately $1-6 million per year (total costs) for 3 to 5 years and that research grants for the third component of this Notice (predictive model development) will be funded at a level of approximately $1-2 million per year (total costs) for 3 to 5 years. Applicants should also describe a scientifically justified scale-up plan to maximize technology development and research productivity.

Merit and Relevance Review

Applications will be subjected to scientific merit review (peer review) and will be evaluated against the following evaluation criteria listed in descending order of importance as codified at 10 CFR 605.10(d):

1. Scientific and/or Technical Merit of the Project;
2. Appropriateness of the Proposed Method or Approach;
3. Competency of Applicant's Personnel and Adequacy of Proposed Resources;
4. Reasonableness and Appropriateness of the Proposed Budget.

In addition, applications will be evaluated for the robustness of their organizational framework and coordination plan.

The evaluation will include program policy factors such as the relevance of the proposed research to the terms of the announcement and the Department's programmatic needs. External peer reviewers are selected with regard to both their scientific expertise and the absence of

conflict-of-interest issues. Non-federal reviewers may be used, and submission of an application constitutes agreement that this is acceptable to the investigator(s) and the submitting institution.

Applications

These large, multi investigator applications will be reviewed as individual research projects consisting of several individual subprojects. The research description (see project description below) for individual subprojects should be no more than 20 pages each, exclusive of attachments. The combined research descriptions for all individual subprojects for each application should be no more than 100 pages, exclusive of attachments. In addition, each application should contain a project overview, not to exceed 20 pages, that contains an overall project summary, research integration plan, management plan, data and information management plan, and a communication plan. Each research team should identify a single scientific coordinator or point of contact for its application.

Each subproject description must contain an abstract or project summary on a separate page with the name of the applicant, mailing address, phone, Fax, and E-mail listed. Each project must include letters of intent from outside collaborators briefly describing the intended contribution of each to the research and short curriculum vitaes, consistent with National Institutes of Health (NIH) guidelines, for all principal investigators and any co-PIs.

Information about the development and submission of applications, eligibility, limitations, evaluation, selection process, and other policies and procedures may be found in the Application Guide for the Office of Science Financial Assistance Program and 10 CFR Part 605. Electronic access to the Guide and required forms is made available via the World Wide Web at: http://www.science.doe.gov/grants.html. DOE is under no obligation to pay for any costs associated with the preparation or submission of applications if an award is not made.

Adherence to type size and line spacing requirements is necessary for several reasons. No applicants should have the advantage, or by using small type, of providing more text in their applications. Small type may also make it difficult for reviewers to read the application. Applications must have 1-inch margins at the top, bottom, and on each side. Type sizes must be 10 point or larger. Line spacing is at the discretion of the applicant but there must be no more than 6 lines per vertical inch of text. Pages should be standard 8 1/2" x 11" (or metric A4, i.e., 210 mm x 297 mm).

Applicants are expected to use the following ordered format to prepare Applications in addition to following instructions in the Application Guide for the Office of Science Financial Assistance Program. Applications must be written in English, with all budgets in U.S. dollars.

- **Face page** (DOE F 4650.2(10-03))
- **Project abstract** (no more than one page)
- **Budgets** for each year and a summary budget page for the entire project period (using DOE F 4620.1)
- **Budget explanation**
- **Budgets and budget explanation** for each collaborative subproject, if any

- **Project description** (includes goals, background, research plan, preliminary studies and progress, and research design and methodologies)
    - Goals
    - Background
    - Research plan
    - Preliminary studies and progress (if applicable)
    - Research design and methodologies
- **Literature cited**
- **Collaborative arrangements** (if applicable)
- **Biographical sketches** (limit 2 pages per senior investigator)
- **Description of facilities and resources**
- **Current and pending support** for each senior investigator

The Office of Science, as part of its grant regulations, requires at 10 CFR 605.11(b) that a recipient receiving a grant to perform research involving recombinant DNA molecules and/or organisms and viruses containing recombinant DNA molecules shall comply with the National Institutes of Health "Guidelines for Research Involving Recombinant DNA Molecules", which is available via the world wide web at: http://www.niehs.nih.gov/odhsb/biosafe/nih/rdna-apr98.pdf, (59 FR 34496, July 5, 1994), or such later revision of those guidelines as may be published in the Federal Register.

DOE policy requires that potential applicants adhere to 10 CFR Part 745 "Protection of Human Subjects" (if applicable), or such later revision of those guidelines as may be published in the Federal Register.

The Catalog of Federal Domestic Assistance number for this program is 81.049, and the solicitation control number is ERFAP 10 CFR Part 605.

Martin Rubinstein
Grants and Contracts Division
Office of Science

Posted on the Office of Science Grants and Contracts Web Site
September 24, 2004.