Office of Science
Notice 02-13

Genomes to Life

Department of Energy

Office of Science Financial Assistance Program Notice 02-13; Genomes to Life

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 Genomes to Life research program (http://www.doegenomestolife.org/). A central theme of the entire Genomes to Life 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. In such proposals, the leadership role may rest either with experimentation or with computation.

DATES: Statements of intent to apply, including information on collaborators and areas of proposed research and technology development should be submitted by March 1, 2002. Research applications are due by 4:30 PM E.D.T. Tuesday May 7, 2002.

ADDRESS: Statements of intent to apply should be sent to Ms. Joanne Corcoran by email at: joanne.corcoran@science.doe.gov with copies to Dr. David Thomassen at: david.thomassen@science.doe.gov and Dr. Walter Polansky at walt.polansky@science.doe.gov. Formal applications, referencing Program Notice 02-13, should be sent to: U.S. Department of Energy, Office of Science, Grants and Contracts Division, SC-64, 19901 Germantown Road, Germantown, MD 20874-1290, ATTN: Program Notice 02-13. This address must be used when submitting applications by U.S. Postal Service Express, commercial mail delivery service, or when hand carried by the applicant. (For safety reasons, the Washington, DC area continues to experience delays in the processing of all U.S. Mail. Please check the Office of Science, Grants and Contacts Web Site at: http://www.sc.doe.gov/production/grants/grants.html for the latest updates regarding the processing of U.S. Mail.)

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, 19901 Germantown Road, Germantown, MD 20874-1290 and Dr. Walter Polansky, telephone: (301) 903-5800, E-mail:

walt.polansky@science.doe.gov, Office of Advanced Scientific Computing Research, SC-31, U.S. Department of Energy, 19901 Germantown Road, Germantown, MD 20874-1290.

A complementary request for proposals from DOE national laboratory led teams has been issued http://www.sc.doe.gov/production/grants/LAB02_13.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 will 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 non-laboratory research partners beginning in FY 2002, each application should include a complete set of forms for each non-laboratory 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.sc.doe.gov/production/grants/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 Focus

The Genomes to Life research program will cut across components of each of the goals described in the Genomes to Life program plan, available on the web at: http://www.doegenomestolife.org/. Applicants should refer to the program plan for additional information on the overall organization of the Genomes to Life program. Individual applications should address one or more of the individual research elements described below.

Other useful web sites include:

MCP Home Page – http://microbialcellproject.org

Microbial Genome Program Home Page - http://www.sc.doe.gov/production/ober/microbial.html

DOE Joint Genome Institute Microbial Web Page - http://www.jgi.doe.gov/JGI_microbial/html/

GenBank Home Page - http://www.ncbi.nlm.nih.gov/

Human Genome Home Page - http://www.ornl.gov/hgmis

Microbes of Interest to DOE. The initial focus of Genomes to Life should be on microbes (including fungi) directly relevant to DOE mission needs in energy (cleaner energy, biomass conversion, carbon sequestration), bioweapons defense (biothreat agents or their close relatives), or the environment (cleanup of metals and radionuclides at DOE sites). Research in Goals 1 and 2 should take advantage of and focus on microbes whose complete DNA sequence is already known. Research in Goal 3 should focus on microbes or microbial communities of interest to, directly relevant to, or that would contribute substantially to an ability to address DOE mission needs. Selected, well-justified research using yeast may also be appropriate as a means of quickly generating data that addresses the needs of this solicitation and of the Genomes to Life Program. However, the use of yeast as a long-term research focus is not encouraged.

Data and Other Results. Any data and results that are generated through the investigations into Goals 1 through 4 that are appropriate to share with the broader community should be provided in timely, open, and machine-readable format where possible. 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). 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. Any code development should be open source. Teams should be amenable to the adoption of open data standards and interoperability requirements, as they evolve and are specified by the Genomes to Life program.

Goal 1 -- Identify and Characterize the Molecular Machines of Life – the Multiprotein Complexes that Execute Cellular Functions and Govern Cell Form

Current structural genomics or proteomics efforts generally focus on individual proteins, either one at a time or at a genomic scale, or as pairs of interacting proteins. An initial focus of the Genomes to Life program will be to develop and implement research strategies and technologies that will enable the systematic identification, characterization, and, eventually, understanding of all the multi protein molecular machines in an organism. A research plan should be described that will lead, within five years, to the development of the capability to measure and characterize thousands of molecular machines per year. The initial focus of this research should be on microbial processes with application to DOE needs (see section on Microbes of Interest to DOE). The research plan should describe how the proposed research and technology and computational tool development will, within the next four to six years, enable at least 80% of the molecular machines in a single microbe to be identified and characterized within a single year.

An overarching goal of the Genomes to Life program is to develop computational tools, based on experimental data, that enable us to predict the functions and behaviors of complex biological systems beginning with genome sequence data. In the context of Goal 1, computational tools are

needed to predict the inventory of molecular machines, and the functions of those machines, likely to be found in a microbe whose DNA sequence is known. This could include development of computational modeling tools, including high performance implementations of techniques analogous to Rosetta-type algorithms and threading programs to characterize the molecular machinery on the scale of complete microbial organisms. Significant effort should be devoted to the development of high-precision computational models able to identify the principal components and functions of characterized molecular machines. These computational approaches will also provide an important future interface with the projected increases in the rate of protein structure determination to understand the molecular details of protein interactions in molecular machines.

Milestones of progress and success should be included as part of the research plan. Pilot studies that test and compare several different research and technology strategies are encouraged along with a decision plan to choose and expand the most promising strategies.

Understanding the role that these molecular machines play within an organism will require information on both the interactions of molecular machines and on the physical and temporal location and behavior of molecular machines within cells. Research plans should be described that will lead to high-throughput strategies, technologies, and computational tools for achieving these goals. Investigators conducting research on these goals should describe how they will work in close collaboration with or maintain a detailed awareness of the progress of investigators who are developing high-throughput strategies for identifying molecular machines. Pilot studies that test and compare several different research and technology strategies are encouraged along with a decision plan to choose and expand the most promising strategies.

Experimental research is not being requested to determine the three-dimensional, high-resolution structure of individual proteins or multi protein molecular machines. As the number of high resolution protein structures in the Protein Data Bank increases dramatically over the next five years, that information will serve as an important starting point for characterizing the molecular details of protein-protein interactions within and between individual molecular machines.

Goal 2 -- Characterize Gene Regulatory Networks

Understanding the structure and function of an organism's molecular machines is a limited, though substantial, first step towards a predictive understanding of the organism's complex functions. This will only come by understanding the complex gene regulatory networks that govern the coordinated formation and behavior of molecular machines and their individual protein subunits. A goal of Genomes to Life is to develop large-scale research strategies, technologies, and computational tools needed to identify all the components of gene regulatory networks with an initial focus on cis-acting regulatory sequences. Although the principal focus should be on microbial processes with application to DOE needs (see section on Microbes of Interest to DOE), these studies will likely benefit from comparative genomics approaches that may cross species.

Again, an overarching goal of the Genomes to Life program is to develop computational tools, based on experimental data, that enable us to predict the functions and behaviors of complex

biological systems beginning with genome sequence data. In the context of Goal 2, computational tools are needed to predict regulatory networks for the molecular machines and their component proteins identified in Goal 1. A major goal is to be able to predict and reconstruct regulatory networks for molecular machines, metabolic pathways, or entire organisms beginning with knowledge of the organism's DNA sequence. Determination and verification of regulatory interactions will be enabled by the development of the integrated computational approaches assembling many types of experimental information together with relevant computational algorithms.

These studies should be closely integrated with genome-scale proteomics efforts or efforts to identify all of an organisms's molecular machines and their dynamic behavior within cells. Pilot studies that test and compare several different research and technology strategies are encouraged along with a decision plan to choose and expand the most promising strategies.

Goal 3 -- Characterize the Functional Repertoire of Complex Microbial Communities in their Natural Environments at the Molecular Level

Understanding the structure and functional capabilities and diversity of complex microbial communities is key to using the diverse functions and capabilities of microbes to address DOE mission needs. However, the majority of microbes of importance and interest to DOE have not been isolated, purified, and cultured. An initial goal of Genomes to Life is to use high throughput DNA sequencing and computational approaches to determine the genetic and functional diversity of individual uncultured microorganisms and of microbial communities. It is anticipated that the majority of high throughput DNA sequencing required for this Goal will be conducted at the DOE Joint Genome Institute. An estimate of the amount of DNA sequencing that will be required should be included as part of the application. Funds for high throughput DNA sequencing should not be included as part of the budget request for individual applications as funds will be provided directly to the Joint Genome Institute for Genomes to Life sequencing needs.

The organisms and microbes chosen for sequencing should be chosen to help make an initial determination of:

- The extent and patterns of phylogenetic and genetic diversity in microbial communities from different environments.
- Whether microbial communities conserve metabolic function in spite of extensive individual phylogenetic diversity and whether a microbial community's metabolic functions correlate with the physical properties of its environmental niche.
- Improvements in the ability to infer the metabolic, physiologic, and behavioral characteristics of a microbe or microbial community from its DNA sequence (including improvements in the ability to infer gene function from DNA sequence).

Just as development of computational tools to predict the inventory, functions, and regulation of molecular machines from genome sequence data is a key part of Goals 1 and 2, development of computational tools to predict the metabolic, physiologic, and behavioral characteristics of microbial communities from community DNA sequence data is a key part of Goal 3. It is

expected that some of the computational tools developed will be executed on existing computer resources with little need for additional computational power. However, 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.

A scientific and experimentally based strategy for selecting the microbes and microbial communities proposed for analysis should be provided. Estimates of the number and diversity of uncultured microbes and microbial communities chosen for sequencing during the first three years of the project should be made. A strategy for estimating the degree of sequence coverage for DNA isolated from microbial communities should be provided.

Goal 4 -- Develop the Computational Methods and Capabilities to Advance Understanding of Complex Biological Systems and Predict their Behavior

Computational capabilities, including data management, modeling of complex biological systems, and prediction of biological responses, underpin all of Genomes to Life. In particular, the needs include:

- Computational research on analysis and modeling of the structure and function of molecular machines, as integrated with the research to be conducted under Goal 1 above, with an emphasis on the interactions among the proteins and other molecules that make up these machines. This could also include investigations into prediction of functions of the molecular machines through the use of consensus groupings or proxies, such as analogs to "Rosetta" or threading-type methods used for predicting the structure of single proteins.
- Computational research on models and simulations of metabolic pathways, regulatory networks, and whole-cell functions, as integrated with the research to be conducted under Goal 2 above. This may include computational tools to integrate data from a wide variety of high-throughput experimental data, such as mass spectrometry, protein arrays, cross linking, and Nuclear Magnetic Resonance data with other biological data, such as genome annotation and experimental genetic data, such as results from knockout experiments.
- Computational research in support of sequencing environmental samples to be conducted under Goal 3 above. Computational tools will be needed to analyze the output of the simultaneous sequencing of multiple organisms. This will include a need to infer properties of the environmental sample, such as the presence or absence of both certain classes of organisms and certain functional capabilities, such as particular metabolic pathways.
- Computational research in support of biological databases and database tool development. Any applications for subprojects to augment or develop databases will be judged primarily on the degree that they contribute to the successful completion of the team's research conducted as part of Goals 1, 2, and 3 above. The subprojects will also be judged on the predicted utility of the database and tools to the broader community and to the degree that the tools contribute to the broader goal of database interoperability.

- It is expected that some of the computational tools developed in Goal 4 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 be paid to attributes such as modularity, interoperability, and scalability.

Program Funding

Up to $15 million is available in FY 2002, contingent upon availability of appropriated funds. It is anticipated that individual research grants will be funded at a level of $1-4 million per year. Applications 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 description of Narrative 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 subproject or 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/production/grants/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.

The application must contain an abstract or project summary, letters of intent from collaborators, and short curriculum vitas consistent with NIH guidelines.

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 (10-91))
- 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.

John Rodney Clark
Associate Director of Science
for Resource Management

Published in the Federal Register February 4, 2002, Volume 67, Number 23, Pages 5095-5099.