Office of Science
Notice 01-21

Advanced Modeling and Simulation of Biological Systems

Department of Energy
Office of Science

Office of Science Financial Assistance Program Notice 01-21; Advanced Modeling and Simulation of Biological Systems

AGENCY: U. S. Department of Energy (DOE)

ACTION: Notice inviting grant applications

SUMMARY: The Offices of Advanced Scientific Computing Research (ASCR) and Biological and Environmental Research (OBER) of the Office of Science (SC), U.S. Department of Energy, hereby announce interest in receiving applications for grants in support of computational modeling and simulation of biological systems. The goal of this program is to enable the use of terascale computers to explore fundamental biological processes and predict the behavior of a broad range of protein interactions and molecular pathways in prokaryotic microbes of importance to DOE. This goal will be achieved through the creation of scientific simulation codes that are high performance, scalable to hundreds of nodes and thousands of processors, and able to evolve over time and be ported to future generations of high performance computers. The research efforts being sought under this Program Notice will take advantage of extensive information inferred from the complete DNA sequence, such as the genetics and the biochemical processes available for a well-characterized prokaryotic microbe; for example, *Escherichia coli* (*E. coli*). This notice encourages applications from the disciplines of applied mathematics and computer science in partnership with microbiology, molecular biology, biochemistry and structural and computational biology to combine information available on a well characterized prokaryotic microbe with advanced mathematics and computer science to enable this new understanding. This announcement is being issued in parallel with Program Notice 01-20, the Microbial Cell Project. Together, they represent a planned first step in an ambitious effort to understand the functions of the proteins in a prokaryotic microbial cell, to understand their interactions as they form pathways that carry out DOE-relevant activities, and to eventually build predictive models for microbial activities that address DOE mission needs.

DATES: Preapplications referencing Program Notice 01-21 should be received by February 21, 2001. Earlier submissions will be gladly accepted. A response to timely preapplications will be communicated to the applicant by March 9, 2001.

Formal applications in response to this notice should be received by 4:30 p.m., E.D.T., April 24, 2001, to be accepted for merit review and funding in FY 2001.

ADDRESSES: Preapplications referencing Program Notice 01-21 should be sent to Dr. Walter M. Polansky, Office of Advanced Scientific Computing Research, SC-32, Office of Science, U.S. Department of Energy, 19901 Germantown Road, Germantown, MD 20874-1290; e-mail is acceptable for submitting preapplications using the following address: walt.polansky@science.doe.gov.

Formal applications referencing Program Notice 01-21, should be forwarded 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 01-21. This address must be used when submitting applications by U.S. Postal Service Express Mail or any commercial mail delivery service, or when hand-carried by the applicant.

FOR FURTHER INFORMATION CONTACT:

Dr. Walter M. Polansky, Office of Advanced Scientific Computing Research, SC-32, Office of Science, U.S. Department of Energy, 19901 Germantown Road, Germantown, MD 20874-1290; telephone: (301) 903-5995, e-mail: walt.polansky@science.doe.gov.

Dr. John Houghton, Office of Biological and Environmental Research, Office of Science, U.S. Department of Energy, 19901 Germantown Road, Germantown, MD 20874-1290; telephone: (301) 903- 8288, e-mail: john.houghton@science.doe.gov.

The full text of Program Notice 01-21 is available via the World Wide Web using the following web site address:
http://www.science.doe.gov/production/grants/grants.html.

SUPPLEMENTARY INFORMATION:

Extraordinary advances in computing technology in the past decade have set the stage for a new era in scientific computing. Within the next five to ten years, computers running at 1 to 10 trillion floating point operations per second (Tops) will become available. Using such computers, it will be possible to dramatically extend

explorations of fundamental processes as well as advance the ability to predict the behavior of a broad range of complex biological systems.

The primary mission of the Office of Advanced Scientific Computing Research is to discover, develop, and deploy the computational and networking tools that enable researchers in the scientific disciplines to analyze, model, simulate and predict complex phenomena important to the Department of Energy. In carrying out this mission, ASCR:

- Maintains world leadership in areas of scientific computing research relevant to the missions of the Department of Energy;
- Integrates the results of advanced scientific computing research into the natural sciences and engineering;
- Provides world class supercomputer and networking facilities for scientists working on problems that are important to the missions of the Department.

The primary mission of the Office of Biological and Environmental Research is to advance environmental and biomedical knowledge connected to energy production, development, and use. In carrying out this mission, OBER:

- Contributes to the environmental remediation and restoration of contaminated environments at DOE sites through basic research in bioremediation, microbial genomics, and ecological science;
- Provides new knowledge that will widen DOE's options for clean and affordable energy through research in microbial genomics and bioinformatics;
- Advances our understanding of and finds solutions for the effects of energy production and use on the environment through research in global climate modeling and simulation, the role of clouds in climate change, carbon cycle and carbon sequestration, atmospheric chemistry, and ecological science;
- Helps protect the health of DOE workers and the public by advancing our understanding of the health effects of energy production and use through basic research in key areas of the life sciences including functional genomics and structural biology as well as low dose radiation research;
- Seeks to develop new applications of radiotracers in diagnosis and treatment and supports biomedical engineering research focused on fundamental studies in medical imaging, biological and chemical sensors, laser medicine, new biocompatible materials, informatics, and artificial organs.

The scope and complexity of the proposed projects will likely require close collaboration among researchers from the biological sciences, computational sciences, computer science, and applied mathematics disciplines. Accordingly, this solicitation calls for the creation of scientific simulation teams, or collaborations, as the

organizational basis for a successful application. Partnerships among universities, national laboratories, and industry are encouraged but not required. A scientific simulation team is a multi-disciplinary, and perhaps multi-institutional, group of people who will:

- create scientific simulation codes that take full advantage of terascale computers,
- work closely with other research teams and centers to ensure that the best available mathematical algorithms and computer science methods are employed, and
- manage the work of the team in a way that will foster good communication and decision making.

Biological systems and their regulatory and metabolic pathways are complex. The details of many biological processes are not well understood, and the resulting computations will require new algorithms, computational biology tools, and extraordinary computing resources. The successful development of the new tools will require the sustained efforts of multi-disciplinary teams, and applications of these tools will require Tops-scale and beyond supercomputers, as well as the considerable expertise required to use them. Although forms of these computational tools already exist, considerable research in mathematics and computer science remains to be done in order to develop reliable, robust, efficient, and widely applicable versions of these tools.

Data analysis, computational modeling and simulation will play critical roles in the future of biological research. Large sets of genomic data will be generated by the on-going DNA sequencing efforts at large genome centers around the world. These data will be analyzed and combined with different types of biological data, including information on structure, expression, and function to develop a more comprehensive understanding of biological systems. Homology-based protein structure correlations identified by pattern searches will be used to predict the structures of the proteins coded by the new genome sequences and will be invaluable for ascertaining protein function and for identifying more distant homologies than are possible by simple sequence comparisons. For selected biochemical processes, computational modeling will be used for a range of applications, from elucidating the mechanisms of enzymatic reactions to identifying the energetic principles underlying macromolecular interactions. Computer models of entire cells and microbial ecosystems will also use the understanding gained about biomolecular processes to predict likely behaviors of organisms under different conditions.

A goal for the research solicited here is to develop a predictive understanding of biological systems using a well characterized prokaryotic microbial cell, for example,

E. coli, as a model system. Given the immense complexity of even the simplest microbes, fully predictive models that provide quantitatively accurate estimates of each chemical component of a cell will remain a challenge for subsequent generations of researchers. Hence, in the foreseeable future, the modeling of cellular processes will instead be performed at a level beyond that of the individual chemical reactions, perhaps at the level of functional building blocks that can be pieced together or linked into higher order models. At this level, cellular pathways are described either qualitatively as being present or absent, or quantitatively, in terms of the average concentrations and rates of activity derived from experimental data. Despite their lack of chemical detail, such models will provide a powerful tool for integrating and analyzing the very large new biological data sets and, under some conditions, predicting cellular behavior under changing conditions. Just as importantly, these high level models will provide a means of inducing and testing the general principles of cellular function.

Three levels of modeling are included in this solicitation: (1) molecular simulations of protein function and macromolecular interactions, (2) semi-quantitative simulations of metabolic networks in whole cells, and (3) quantitative kinetic models of biochemical pathways. The latter simulations are much more demanding in terms of the empirical data and computer power required and therefore, will initially be limited to relatively small, well characterized pathways. Since both of these levels of modeling depend on having the (nearly) complete parts lists provided by the fully annotated genome sequences, combined with gene function, expression information and phenotypic data about an organism, the focus of this solicitation will be on *E. coli* or another well-characterized and studied prokaryotic microbe.

1) Molecular simulations of protein function and macromolecular interactions. The ultimate biological models would be molecular-level simulations of each biochemical process. There are many challenges to molecular-level simulations of biological processes, including the large size of biomolecules and the wide range of time scales of many biological processes, as well as the subtle energetics and complex milieu of biochemical reactions. Moreover, many biochemical reactions occur far from equilibrium and are regulated by both transport of the reactants and subsequent processing of the products. Finally, there remains a wide gulf between the detailed chemical data needed for initiating and validating biomolecular simulations and the data available on many biological processes and environments. Despite these challenges, there are a vast number of biochemical processes for which chemical simulations will have a major impact on our understanding. These problems include the elucidation of the energetic factors underlying protein-protein or protein-DNA interactions and the dissection of the catalytic function of certain enzymes. The promise of such modeling studies is rapidly growing as a result of the development of

linear-scaling computational chemical methods and molecular modeling software for massively parallel computers. Additionally, molecular modeling will be used to determine the principles that underlie protein-protein interactions, and ultimately to predict likely protein binding sites.

2) Semi-quantitative simulations of metabolic networks. This modeling approach follows the engineering tradition of making maximal use of limited information by combining highly simplified models with successive constraints to identify an "envelope" of expected behaviors of the system under different conditions. A fundamental tenet of such modeling is that the very complex molecular details of biology combine to form robust and relatively simple rules for behavioral responses. Such models are iteratively refined as more functional data and constraints become available from experiments that are themselves guided by the model's predictions.

Since such modeling depends only on the nature of the reactants and products (i.e., the stoichiometry) of the metabolic transformations, rather than the rates of these reactions (kinetics), most of the necessary data for building the model can be derived directly from annotated genomes, in some cases using artificial intelligence based pathway synthesis algorithms. These data are typically encoded in a "stoichiometry matrix" relating specific reaction products to metabolic reactions. Numerical analysis of this matrix can identify the entire repertoire of theoretically possible metabolic capabilities of a given genotype, for example, what nutrients are essential and what metabolic pathways are non-redundant. Such information, although qualitative, has enormous potential value. It will allow the inference of phenotypic properties directly from the functionally annotated genotype, help in the optimization of product yield in bio-reactors, and provide a predictive basis for engineering organisms with novel capabilities. Additionally, such analysis can be used to improve and validate tentative functional annotations. Even in the absence of stoichiometric data, mathematical analysis of metabolic networks can shed light on overall biological function. A number of successful models have already been developed for *E. coli* using both stoichiometric data, based on a network analysis, and constraint-based approaches.

Unlike the kinetic pathway described below, computing speed is not typically a limiting factor in molecular pathway analysis. Instead, the primary bottleneck to progress is the availability of functionally annotated genomes and the human talent trained in both the biological sciences and the art of developing and applying such mathematical models. The choice of a well-characterized prokaryotic organism as a model biological system for this solicitation minimizes the challenges associated with the first bottleneck.

3) Quantitative kinetic models of biochemical pathways. Although the metabolic network modeling described above can provide useful qualitative information on

possible behavioral characteristics of organisms, a fully predictive understanding of biological processes will require quantitative information about the dynamics of each sub-process. In other words, network analysis can suggest what metabolic transformations may be possible, but full kinetic details are required to determine which pathways are most important under the given conditions. Such models will require detailed empirical data, including *in vivo* reaction rates and substrate concentrations for each step in the biological system to be simulated. Additionally, these simulations are highly computationally demanding; for example, the simulation of a regulatory circuit involving only several dozen parameters required the use of a parallel supercomputer. These experimental and computational requirements will prohibit such quantitative simulations of whole cells in the foreseeable future. Nevertheless, for selected critical cell subsystems, such simulations offer the promise of quantitative predictions of cellular response and will constitute a rigorous validation of the completeness of our understanding the processes under investigation.

Kinetic models have been applied to a handful of specific cellular pathways that demonstrate both the benefits and technical challenges of such simulations. One of the most complex examples to date has been a full kinetic analysis of the lytic versus lysogenic pathways in phage lambda infected *E. coli* cells. The heart of the decision circuitry for this pathway contains only four promoter sites modulated by five gene transcripts, yet the kinetic model required nearly forty empirical rate constants and a number of other parameters. Additionally, to be computationally tractable, this model involved a number of simplifying assumptions, including approximating the cell as a well-stirred homogeneous mixture. Despite these assumptions and the large number of empirical parameters this model yielded reasonably accurate results for the lytic/lysogenic fractions at different levels of viral infection.

An important outcome of this previous work is to highlight the significant differences between the modeling methodologies necessary for biochemical pathways and those used for macroscopic chemical processes (e.g., in optimizing industrial chemical processes.) In the latter the chemical concentrations can be assumed to be continuous and therefore the kinetics can be simulated using ordinary differential equations. In contrast, the very small numbers of individual signaling molecules in biological regulatory pathways require the use of discrete stochastic simulations. Indeed, a number of seemingly non-deterministic features in gene expression have been ascribed to the inherently stochastic fluctuations in the concentrations of very small numbers of regulatory signals.

Overall, both the kinetic models and the metabolic network analysis will provide a means of combining and evaluating the consistency of large sets of biological data. Each requires detailed functional annotation of whole genomes and well as phenotypic data under a wide variety of conditions.

In a parallel solicitation, the Microbial Cell Project (see Program Notice 01-20) supports key DOE missions by building on the successful DOE Microbial Genome Program that has furnished microbial DNA sequence information on microbes relevant to environmental remediation, global carbon sequestration (e.g., CO_2 fixation), complex polymer degradation (e.g., cellulose and lignins), and energy production (fuels, chemicals, and chemical feedstocks). These microbial genome sequences provide a finite set of "working parts" for a cell and the challenge now is to understand how these parts are assembled into functional pathways and networks to accomplish activities of interest to the DOE. The traditional reductionist experimental approach has defined specific steps or stages within many physiological processes; however, the availability of whole genomes affords the opportunity to integrate these individual pathways into a larger physiological or whole organism framework. The Microbial Cell Project seeks to integrate available information about individual processes and regulatory complexes to understand the intracellular environment, in which these pathways and networks exist and function. The DOE Microbial Cell Project is part of a coordinated Federal effort called the Microbe Project involving elements from several other Federal agencies. The long-term goal is that research funded in this program and in the Microbial Cell Project will converge so that simulations and models can be developed in organisms and for biochemical pathways important for the DOE mission.

This notice takes advantage of decades of research on *E. coli* (or a similarly well characterized prokaryotic microbe) providing much of the biological information needed to begin developing more comprehensive models of biological systems. It is anticipated that the applied mathematicians and computer scientists will need to partner with biologists in the initial phases of algorithm development, as well as in the design of biological tests to validate models that are developed, including predictions made using these models. Links to some of the vast amount of information available on *E. coli* can be found at http://genprotec.mbl.edu/start and http://web.bham.ac.uk/bcm4ght6/res.html.

The mathematical and computer science challenges in this effort span a broad range of the current research topics in both fields. A few examples of possible areas include: advanced techniques for data fusion; algorithms for solution of low dimensional dynamical systems in the presence of uncertainty; applications of computational geometry and topology to pattern recognition and analysis; advanced concepts in discrete state machines; and control theory. It must, however, be emphasized that the preceding list is only a list of possible examples and does not reflect any prioritization of areas. Collaboration and Coordination

Applicants are encouraged to collaborate with researchers in other institutions, such as: universities, industry, non-profit organizations, Federal laboratories and Federally

Funded Research and Development Centers (FFRDCs), including the DOE National Laboratories, where appropriate, and to include cost sharing wherever feasible. Further information on preparation of collaborative proposals is available in the Application Guide for the Office of Science Financial Assistance Program that is available via the World Wide Web at:
http://www.science.doe.gov/production/grants/Colab.html.

Preapplications

Potential applicants are strongly encouraged to submit a brief preapplication that consists of two to three pages of narrative describing the research objectives, the technical approach(es), and the proposed team members and their expertise. The intent in requesting a preapplication is to save the time and effort of applicants in preparing and submitting a formal project application that may be inappropriate for the program. Preapplications will be reviewed relative to the scope and research needs outlined in the summary paragraph and in the SUPPLEMENTARY INFORMATION. The preapplication should identify, on the cover sheet, the title of the project, the institution, principal investigator name, telephone, fax, and e-mail address. No budget information or biographical data need be included, nor is an institutional endorsement necessary. A response to each timely preapplication will be communicated to the Principal Investigator by March 9, 2001.

Program Funding

It is anticipated that up to $2 million will be available for all awards in Fiscal Year 2001. Multiple year funding is expected, also contingent on availability of funds and progress of the research; pending the availability of future funding, it is anticipated that this initiative will reflect a long term commitment to understanding the workings of a microbial cell. Awards are expected to range from $250, 000 to $600,000 per year with terms of one to three years. The DOE is under no obligation to pay for any costs associated with the preparation or submission of an application. DOE reserves the right to fund, in whole or in part, any, all, or none of the applications submitted in response to this Notice. Applications received by the Office of Science under its normal competitive application mechanisms may also be deemed appropriate for consideration under this announcement and may be funded under this program.

Merit Review

Applications will be subjected to scientific merit review (peer review) and will be evaluated against the following evaluation criteria which are listed in descending order of importance 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 to the above evaluation criteria, applications will also be evaluated on the following:
5. The robustness of the organizational framework if a consortium is proposed;
The evaluation under item 2, Appropriateness of the Proposed Method or Approach, will also consider the following elements:
a) clarity of the plan in detailing areas of work to be addressed by biologists, computational scientists, applied mathematicians, computer scientists and computer programmers;
b) quality of the plan for effective collaboration among participants;
c) viability of the plan for verifying and validating the models developed, including verification using experiment results; and
d) quality and clarity of the proposed work schedule and project deliverables.

The evaluation will include program policy factors such as the relevance of the proposed research to the terms of the announcement and the agency's programmatic needs. Note, external peer reviewers are selected with regard to both their scientific expertise and the absence of conflict-of-interest issues. Non-federal reviewers will often be used, and submission of an application constitutes agreement that this is acceptable to the investigator(s) and the submitting institution.

Submission Information

The Project Description must be 25 pages or less, exclusive of attachments. It 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. The application must include letters of intent from collaborators (briefly describing the intended contribution of each to the research), and short curriculum vitaes, consistent with NIH guidelines, for the applicant and any co-PIs.

To provide a consistent format for the submission, review and solicitation of grant applications submitted under this notice, the preparation and submission of grant applications must follow the guidelines given in the Application Guide for the Office of Science Financial Assistance Program, 10 CFR Part 605. Access to SC's Financial Assistance Application Guide is possible via the World Wide Web at:
http://www.sc.doe.gov/production/grants/grants.html.

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 Office of Science, as part of its grant regulations (10 CFR 605.11(b)) requires that a grantee funded by SC and performing research involving recombinant DNA molecules and/or organisms and viruses containing recombinant DNA molecules shall comply with the NIH "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.

Other useful web sites include:

MCP Home Page - http://microbialcellproject.org

Microbial Genome Program Home Page - http://www.er.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

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 January 26, 2001, Volume 66, Number 18, Pages 7890-7894.