Report from the Joint ASCR/BERAC Subcommittee on Modeling and Simulation for GTL:

Toward the Development of Predictive Theory and Modeling in Biology

Rick Stevens, Argonne-UChicago (co-chair)
John Wooley, UCSD (co-chair)
Michael Banda, Lawrence Berkeley National Lab
David Galas, Battelle and ISB
Keith Hodgson, Stanford
David Kingsbury, Moore Foundation
Chris Somerville, Carnegie Institution
Barbara Wold, Caltech
Thomas Zacharia, Oak Ridge-UT

Executive Summary

A joint ASCAC/BERAC panel, responding to a charge from the Office of Science, has been analyzing how best to address issues relating to the development of computational models for the DOE Genomics: GTL Program. The general issues concern the goals of the joint program office effort, the barriers to success, and strategies for overcoming these barriers. Following preliminary discussions, the panel identified a team of experts from the community to participate with the panel in a two-day workshop (in October 2007) to discuss the status of current research and potential goals for the future. The panel observed that there has been substantial progress over the past five years in developing techniques for building computational models (*e.g.*, for metabolic and regulatory networks) of microbes, for analysis of high-throughput datasets, and for the integration and visualization of biological datasets. Given the capability of experimental techniques, the sustained progress in computational capabilities, the status of Genomics: GTL projects and the mission of DOE, the panel concluded that further progress is highly likely and will be important for the goals of Genomics: GTL Program. Progress can most readily be accelerated through a focused, joint effort within the Office of Science. The panel makes six recommendations for advancing the program and addressing the questions raised in the charge.

- *The ten-year OMB PART goal for ASCR for the joint modeling and simulation activity of ASCR and BER should be modified for both ASCR and BER to read as follows: (ASCR/BER)* ***By 2018, validate capability to predict phenotype from an organism's genome and to predict genotype from an organism's physiology***

  This PART goal should be accompanied by a specific set of progress metrics.

- *DOE should develop an explicit research program aimed at achieving significant progress on the overarching goal of predictive modeling and simulation in DOE relevant biological systems. This program should be a joint effort between ASCR and BER and should include a diversity of modeling approaches.*

- *DOE should establish an annual conference that focuses on highlighting the progress in predictive modeling in biological systems. This should be an open meeting and separate from any programmatic PI meeting.*

- *The DOE GTL modeling and simulation research program should be supported by an explicit series of investments in modeling technology, databases, algorithms, and software infrastructure needed to address the computational challenges.*

  DOE should establish a mechanism to fund the long-term curation and integration of genomics and related datasets (annotations, metabolic reconstructions, expression data, whole genome screens, phenotype data, etc.) to support, in particular, the needs of modeling and simulation in areas of energy and the environment that are not well supported by NSF and NIH, as well as enabling biological research in general.

- *DOE should work with the community to identify novel scientific opportunities for connecting modeling and simulation at the organism level to modeling and simulation at other spatial and temporal scales.*

Introduction

The joint ASCR-BER AC panel was chartered by Dr. Ray Orbach in February 2007 to address three questions posed in the charge letter (reproduced at the end of this report).

The overall charge was to address the issue of computational models for GTL and how progress could be accelerated through targeted investments in applied mathematics, computer science, and computational biology. Specifically, the panel was asked to address the following questions:

1. *Is the current ASCR long-term goal too ambitious given the status and buy-in from the community?*

2. *What intermediate goals might be more relevant to the two programs?*

3. *What are the key computational obstacles to developing computer models necessary to characterize and engineer microbes for DOE missions such as biofuels and bioremediation?*

The joint subcommittee met for two days in October at the Moore Foundation to hear presentations from researchers on the state of the art of modeling and simulation of microbial organisms and to hear projections of what might be possible over the next ten years. This meeting resulted in the generation of a set of findings and recommendations aimed at positioning the combined efforts of the DOE offices of Advanced Scientific Computing Research and Biological and Environmental Research to address the community's needs better and achieve widespread engagement by the community.

Many of the issues addressed by the panel were foreseen by the community at the beginning of the GTL program. The following excerpt is from the vision workshop for computational and systems biology sponsored by DOE in September 2001.

> Biology is widely noted as the next scientific frontier and as the next "killer application" for high-end computational science. It also will eventually drive both computer science research and the design and investment in high-performance computers and networks. However, funding agencies are still working to refine effective strategies to develop research programs in computational and systems biology. In part, this is because computational biology is still a relatively small subfield of biology and therefore doesn't yet have a large constituency—somewhat like the early days of the genome sequencing programs. As computational biology begins to have more scientific impact on the field and the tools become more widely used, this difficulty will be reduced.
>
> The second challenge is the heterogeneity of computational biology applications. Other scientific communities, such as climate modeling or combustion, typically have a single major computational application that has an unambiguous need for very high performance computing, so that it usually is easy to estimate the improvements that will

be achieved by specific investments in software or hardware. In this case, as was clear from the diversity of talks at the workshop, there is a huge variety of computational biology applications, including databases, sequence annotation, protein structure prediction, biochemical simulations, metabolic network modeling, and many others. Each involves different types of computer science and different barriers to progress, typically not just the need for faster computers and more efficient numerical algorithms.

A number of strategies to develop programs in computational and systems biology were discussed at this workshop. One is to link more clearly the results of quantitative biosciences to national needs. For example, DOE is developing new computational and systems biology programs to support its missions in the roles of microorganisms in climate change and energy production, bioremediation of energy and nuclear materials waste, the health risks of low dose radiation exposure, and the basic bioscience needed for effectively defending against biological attack. Another key strategy is to form partnerships between agencies and offices funding biology and other relevant disciplines. For example, a new partnership has been developed between the DOE Offices of Biological and Environmental Research and the Office of Advanced Scientific Computing Research in developing computational and experimental biosciences programs, including joint grant solicitations and multidisciplinary review teams.

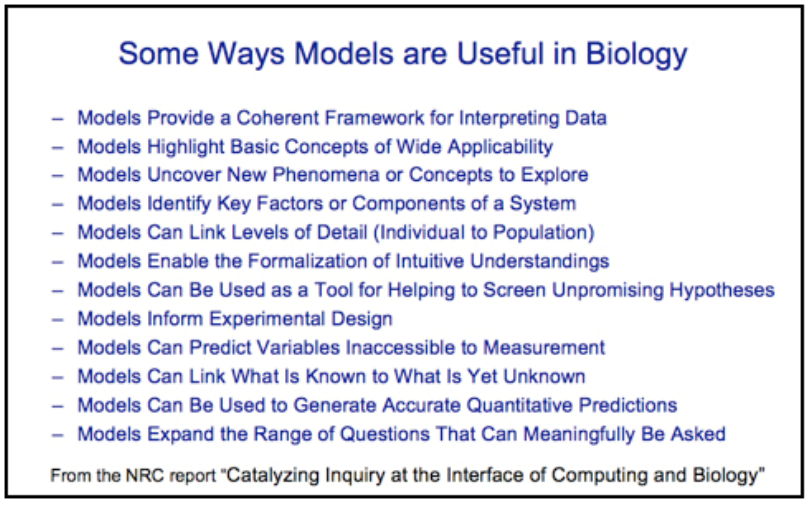
It has been nearly seven years since that 2001 workshop report was written. During this time, the GTL program has made major investments in systems biology projects across the Laboratories and universities. Moreover, numerous major accomplishments have advanced our knowledge of how biological systems function and have provided a framework for effectively coupling experimental, theoretical, and computational programs.

This report addresses the key issues involved in ensuring continued progress in developing a systems level and integrative analysis, the new biology, in support of DOE missions.

Background

In the past decade, bioinformatics has been used to support advanced microbial biotechnology in many ways: computational analysis of wet-lab data, genome sequencing analysis, identification of protein coding genes, genome comparison to identify gene function, the development of genomic and proteomics databases, and the inference ofphenotypes (the higher-level implementation of functions) from genotypes (the gene-level specification of functions). In order to understand higher-level functions, four major types of studies have been undertaken: automated reconstruction and comparison of metabolic pathways; the study of protein-protein and protein-DNA interactions and expression data to understand regulatory and signaling pathways; modeling of the two-dimensional (2D) and three-dimensional (3D) structures of proteins, RNA, and complexes; and modeling of the docking of 3D models of proteins with drugs. Understanding the 3D structure of proteins has had a major impact in our understanding of protein-protein interactions. Studies of protein-protein and protein-DNA interactions have provided a good understanding of binding sites in signaling pathways. Moreover, understanding the interactions between proteins and chemical compounds has already facilitated the development of drugs by design.

Three approaches have generally been used to undertake the research described above: computational search and alignment techniques to compare a new genome against the set of known genes to support the annotation and determination of the structure and function of genes in a newly sequence genome; applied mathematical modeling techniques such as data mining, statistical analysis, neural networks, genetic algorithms, and graph matching techniques to identify common patterns, features, and high-level functions; and increasingly, the integration of sequence analysis, database, and search techniques with mathematical modeling and simulation. See Figure 1.

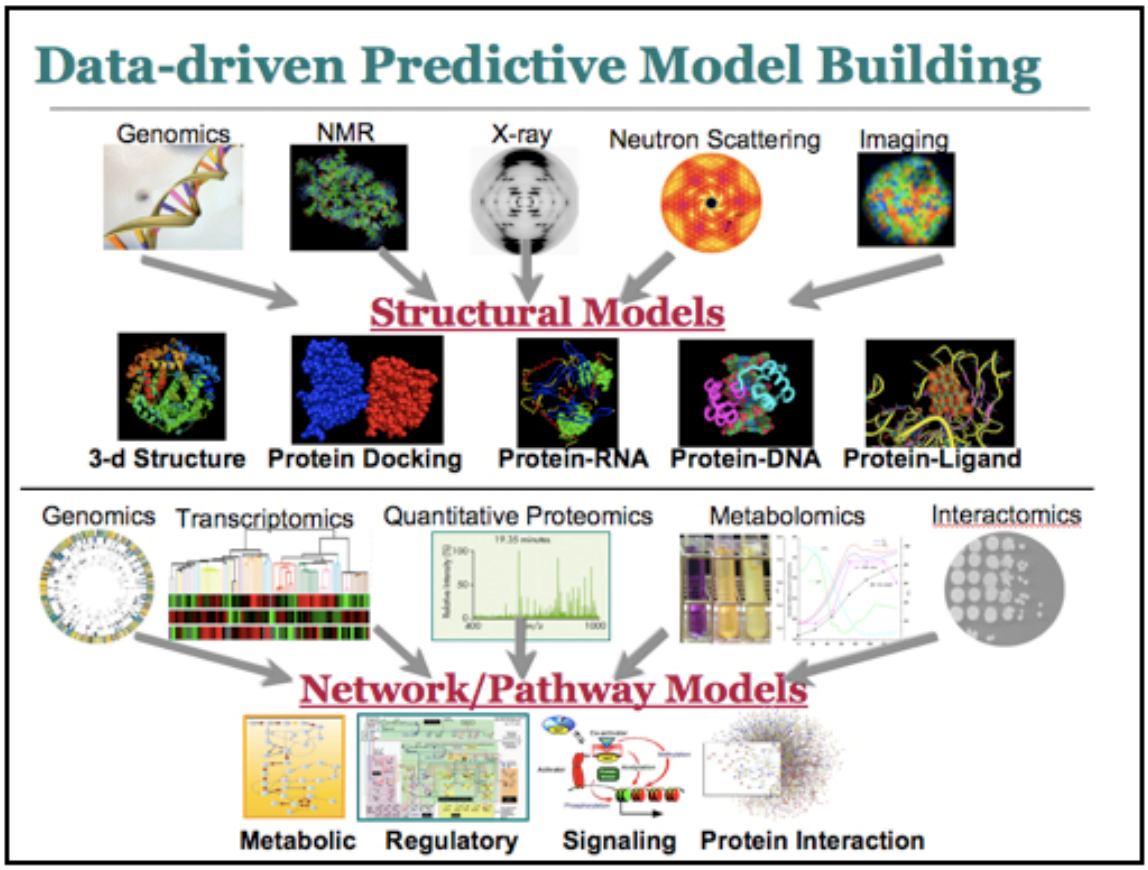


This report focuses on the third approach, which is now mature due to the simultaneous advances in biology and computing: the idea of integrating bioinformatics approaches with mathematical modeling and simulation to enable and accelerate advances in systems-level understanding of complex biological systems.

Results are presented as a set of findings and recommendations. The findings aim to capture the significant accomplishments, opportunities, and remaining challenges in the areas of modeling and simulation in microbiology and cellular biology that are particularly essential for the Genomics: GTL Program. The presentations and other material convincingly show that tremendous progress has been made in developing predictive models of cellular processes during the past decade. It is now possible in some cases to create models of sequenced microbes that can predict growth rates on various substrates, predict genes that are essential for growth, and predict transcriptional responses of the modeled organism to specific types of environmental changes. These models represent early steps toward realizing a broader vision of being able to predict an organism's phenotype (set of expressed traits and biochemical behavior) from an understanding of its genome and the environment in which it is found.

These early pioneering efforts are still under development, and their predictive ability varies depending on the specific model and organism. However, significant progress clearly is being made in developing a broad range of predictive models, and more progress is likely in the immediate future. Nevertheless, at least two factors are currently slowing the research efforts of the community.

The first factor is that small groups with limited funding are pursuing the vast majority of the current work, with uncertain levels and duration of funding often requiring the integration of multiple funding sources from multiple agencies over time to maintain progress. This prevents the development of long term teams with critical masses of disciplinary and interdisciplinary researchers combining theory, modeling and experimentation needed to advance the goal (Fig. 2).



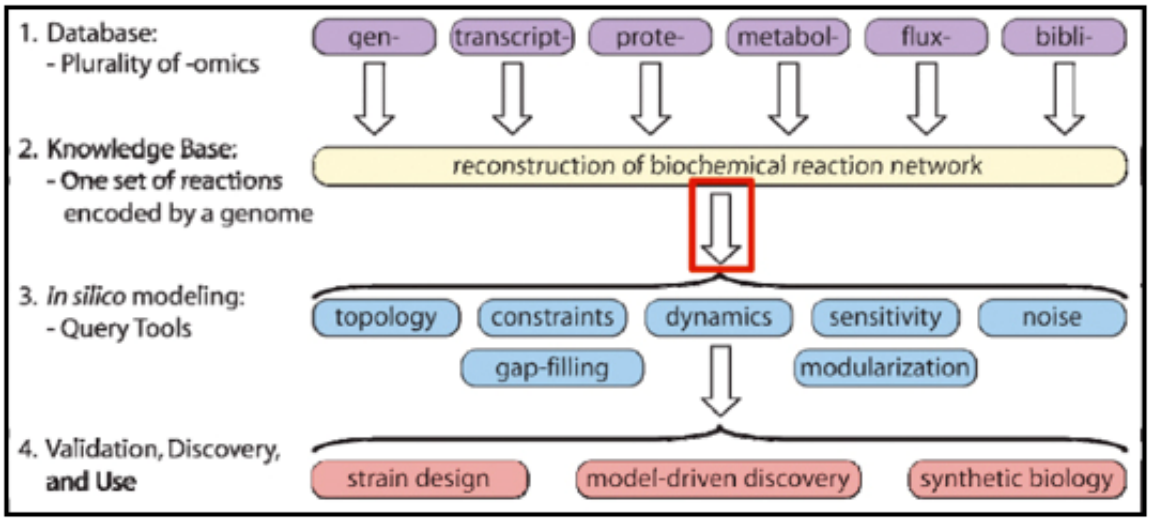
The second limiting factor is that DOE is not yet a significant provider of funding for the development of modeling and simulation research for biological systems, even though the advancement of modeling and simulation would directly advance the DOE mission areas of computational science and applied mathematics and would improve our understanding and control of biological systems relating to energy and the environment. Thus, the community's work in this area is in many cases indirectly or even only tangentially related to DOE mission goals.

The modeling efforts discussed by the committee span different levels of biological organization and detail, ranging from symbolic models of metabolic and regulatory networks, to flux balance analysis, to stochastic models of specific cellular subsystems. All of these modeling approaches are relevant to DOE problem areas. While the field of biomolecular modeling (structural modeling) is relatively mature and accounts for significant fractions of supercomputing time allocations at both DOE and NSF centers, it is limited in its ability alone to contribute to the challenge of developing of systems-level representations of organisms. The focus of this report is on the development of more integrated models that - instead of modeling a single protein or protein complexes - focus on models most appropriate for supporting systems biology research (e.g., metabolic pathways, transcription regulatory networks, signaling, and development).

Enabling the construction and curation of these diverse biological models are numerous efforts to capture, integrate, and annotate the wealth of genomics data for hundreds of

microbial organisms. These databases contain data from both model organisms and organisms that do not have extant user communities.

Each model building effort is also coupled directly or indirectly to one or more experimental efforts that result in the generation of diverse datasets for model development and validation (Fig. 3). Currently, however, no coordinated mechanism exists to capture these datasets and make them available to the community in a sustainable fashion. While NIH supports a variety of biological databases through the National Center for Biotechnology Information (NCBI) for biomedical research, DOE has no corresponding activity for capturing data on organisms and experiments relevant to DOE applications that would not naturally be archived and curated in the NIH databases.



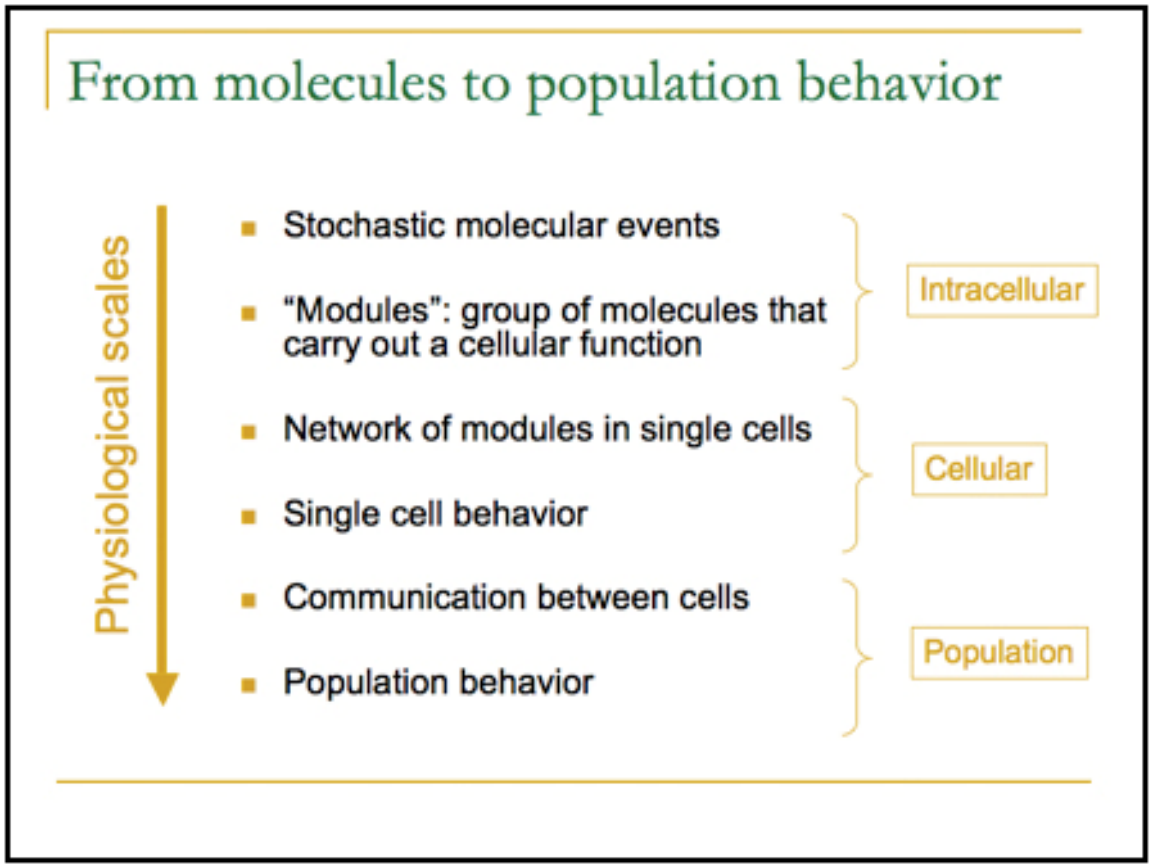
This conclusion was pointed out in the September 2001 Vision for GTL report:

> The clear consensus was that these earlier efforts were limited by a lack of experimental data and the means to verify the models quantitatively. There also was agreement on the key requirements necessary to create a successful new biology. The methods and results of quantitative and predictive biology must:
>
> 1. Be guided by the important biological questions of the day;
>
> 2. Tightly integrate computational analysis and experimental characterization of biological systems;
>
> 3. Draw on multiple types of experimental information and computational analyses;
>
> 4. Be made accessible to those not extensively trained in computational simulation; and
>
> 5. Ultimately use computation and modeling to drive hypothesis formulation, experiment design, and data collection.

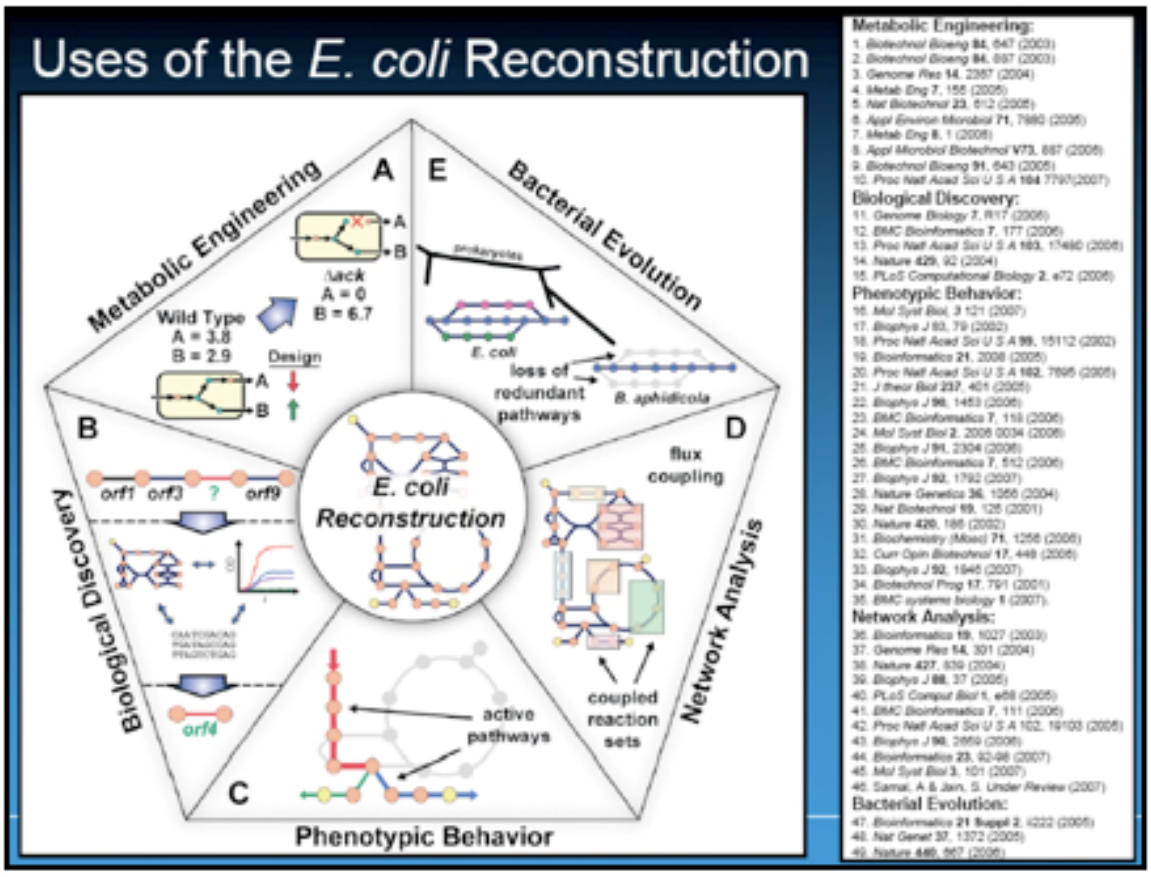
> Key also will be the need for scientists trained to be part of such a multidisciplinary research program—ideally this new generation of scientists will be equally "intellectually comfortable" in both biology and computation.

One way to view the impact of modeling on microbiology is through the lens of improving our understanding of biological systems-level functions.

Functioning models (of all types) require self-consistent representations of biological functions (metabolic, regulatory, or signaling networks) to produce correct results. By building integrated models and testing them against experimental data, it is possible and likely that inconsistencies in gene or gene product function assignments will be found and corrected. This situation has happened in cases where attempts to build flux balance analysis models of particular organisms have uncovered errors in biochemistry databases regarding reactions and gene product reaction mapping. Further, it has resulted in the generation of numerous conjectures regarding missing genes (i.e., functions known to be present in the organism through experiment and required for self-consistent models, but not yet mapped to a known gene product). In this way modeling itself acts like a large consistency check on our collective understanding of gene function assignments. See Figure 4.



Much current work in bioengineering (modification of metabolic and regulatory pathways) is guided mainly by the intuition of the researcher, perhaps with simple causal models of the system. Experience in other disciplines (*e.g.*, electronics design) shows that progress can be greatly accelerated when simple computational tools and models (*e.g.*, early VLSI [should this be defined?] CAD tools) become available that approximate the system enough to be useful as replacements for trial and error. Large-scale computing is also beginning to making it feasible to model ecosystems by aggregating models of individuals. With access to petascale computing capabilities this technique may begin to be applied to natural environments such as soils and to artificial environments such as bioreactors, in order to understand the interactions between different types of organisms and their ability to cooperatively metabolize compounds important for carbon cycling. See Figure 5.



With the number of completed genome sequences reaching 1,000 in the next few years, a new class of biological problem can be addressed: reconstructing the function of entire genomes and building models that enable the prediction of phenotypes from the genotype. With advances in modeling it may become feasible to quickly produce a whole genome-scale model for a new sequenced organism and begin to understand the organism's lifestyle prior to culturing the organism.

Current systems and models can address one growth condition at a time and sweep through a narrow range of control parameters or through single- or double-gene

knockouts to compute essentiality or coessentiality. Metabolic modeling tools use constrained optimization solvers. On the largest systems it is possible to find coessential gene predictions and do limited parameter searches. Modeling of small consortia is just becoming possible. Modeling the evolution of cellular networks is not yet feasible – but is nearly there. Indeed, the creation and adoption of reliable and relatively simple models of transcription regulation and metabolic flux analysis are on the verge of enabling a dramatic increase in the productivity of researchers and engineers tasked with improving strains for industrial use or exploring the limits of pathway engineering.

For example, it is now routine to sequence a bacterial genome and within a few weeks have a basic understanding of the organism's metabolism. It is also becoming feasible to reconstruct a bacterial transcription regulatory network from gene expression and transcription factor binding experimental data in a few months with existing levels of computing. And it is becoming possible to reconstruct a genome from environmental sequence data even if we cannot yet culture the organism. The vision is to integrate reconstructions in a series of models at various levels of abstraction, ranging from flux models (constrained optimization models) to full network time-dependent PDE [define] models with parameter estimation that can be used for a variety of predictive simulations. With access to next-generation computing systems it will be possible not only to build more complex models but also to optimize them for a variety of engineering purposes, enabling the era of computer-aided design of organisms for both science and industry.

Findings and Recommendations

The findings and recommendations are organized in three groups related to the charge questions.

Charge Question 1

> ***Is the current ASCR long-term goal too ambitious, given the status and buy-in from the community?***

The committee did not find the ASCR long-term goal to be too ambitious. However, two factors suggest that it should be modified.

The first is the observation that many of the projects funded by the GTL program have not ranked the development of integrative modeling and simulation as a key strategy for advancing the scientific goals of their research. Therefore, it is unlikely that many of those projects would spontaneously achieve the ASCR goal. This is not to say that no community is willing and able to do so; rather, those groups that are committed to advancing the goal of predictive modeling and simulation for biological systems of interest to DOE have not yet been a specific target of GTL funding.

The second is that the wording of the ASCR long-term goal is ambiguous and not easily subject to measurement. Thus, it may not be the ideal statement of a goal that would be subject to monitoring or assessment.

Finding #1. Modeling and simulation clearly are beginning to play a critical role in integrating the understanding of biological mechanisms at multiple levels, including specific cellular subsystems such as metabolism, motility, signaling, regulation, differentiation, and development—all of which are critical areas of understanding relevant to advancing DOE mission areas. Moreover, the community clearly is ready to take big steps in the direction of more complete models incorporating more detailed biological mechanisms and to apply these models to more areas of biological science. It should also be noted that integrative modeling of biological systems complements the relatively well-developed field of atomistic modeling (*e.g.*, molecular dynamics), which can contribute to DOE mission areas in biology but which is not sufficient to meet the long-term bioengineering goals alone.

Finding #2. While considerable progress in advancing integrative modeling has occurred during the past decade (as witnessed in the high quality of presentations heard by the subcommittee), this progress has been driven largely by a relatively small number of research groups that have been successful at piecing together research support from a number of disparate sources (e.g., NIH, NSF, DOE, DARPA). Currently there is no long-term research program of appropriate scale aimed explicitly at developing biological modeling and simulation capabilities relevant to DOE missions.

Finding #3. The ASCR-supported components of the GTL program are not currently supporting projects in applied mathematics or computer science primarily targeted at developing integrated modeling and simulation capabilities for microbes or plants.

Recommendation 1. *The ten-year OMB PART goal for ASCR and BER for the joint modeling and simulation activity of ASCR and BER should be modified to read as follows:*

(ASCR/BER) By 2018, ***validate capability to predict phenotype from an organism's genome and to predict genotype from an organism's physiology***

This PART goal should be accompanied by a specific set of metrics of progress. Such metrics could include for a given organism the number of correct metabolic phenotype measurements predicted, the fraction of an organism's genes and gene products included in a model, the number of transcription regulatory elements in a model, the number of correct gene expression experiments predicted, and the fraction of correct predictions of essential genes, number of organisms for which predictive models can be generated.

Recommendation 2. *DOE should develop an explicit research program aimed at achieving significant progress on the overarching goal of predictive modeling and simulation in DOE relevant biological systems. This program should be a joint effort between ASCR and BER and should include a diversity of modeling approaches.*

The program should leverage existing experimental activities as well as support the development of new experimental activities that are directly tied to the needs of developing predictive models. This new research program should be aimed at advancing the state of the art of cell modeling directly, should include significant participation from biologists and mathematicians, computer scientists, and engineers; and should be indirectly coupled to the more applied goals of bioenergy, carbon cycle research, or bioremediation.

This program will need to be supported at a large enough scale that a multiple-target approach can be pursued that will enable progress on many intermediate goals simultaneously by different research groups.

Recommendation 3. *DOE should establish an annual conference that focuses on highlighting the progress in predictive modeling in biological systems. This should be an open meeting and separate from any programmatic PI meeting.*

One goal of the meeting would be to establish a series of scientific indicators of progress in predictive modeling, similar to successful indicators associated with the competitive assessment of structure prediction (CASP). These types of measures will enable the community to benchmark progress on methods and will be critical to assessing the impact of the research program on fundamentally advancing the state of the art. Example

metrics could include predicting essentiality in microbial genomes, predicting gene expression patterns in novel environments, predicting flux values through key reactions in microorganisms, or predicting yields in metabolic engineering scenarios.

Charge Question 2

> ***What are potential intermediate goals that might be more relevant to the two programs?***

Intermediate goals that could be considered more relevant for the two programs fall into two general areas.

The first area is building needed tools, curated databases, and computational and collaborative infrastructure that directly accelerate the communities' ability to develop models and simulations. Examples of these are tools for curation of genomes and reconstruction of metabolic networks, integrated databases enabling the community to share data needed to build and test models and validation datasets, and mathematical libraries and core model components that would enable many groups to leverage the work of others.

The second area is focusing on a targeted set of biological modeling and simulations problems that build on each other and that over time would expand the modeling capabilities in the appropriate directions. Examples of these are models of cellular metabolism, motility, global transcription regulation and differentiation, and life-cycle development. Each of these models could play a role in advancing toward the overarching goal of a complete cell model that can be used to predict phenotypic traits or behaviors of a cell from genomic and other "omic" data sources.

Finding #4. Integrative modeling and simulation efforts are highly dependent on the curation of genomics data and associated integrated pathway and protein databases that support metabolic reconstruction, interpretation of microarrays, and other experimental data. These databases are the foundation for the development of models and provide the critical biological context for a given organism or problem. Through resources like NIH's NCBI and the dozens of community-led database projects, there is reasonable coverage of model organisms (*e.g.*, *Escherichia coli* and *Saccharomyces cerevisiae*) and pathogens; however, there is not the same level of support for curating the data associated with organisms related to energy and the environment.

Finding #5. Modeling and simulation in microbial systems have advanced in many areas simultaneously. For some systems, there are useful predictive models for core metabolism, global transcription regulation, signaling and motility control, and life-cycle development and differentiation. However, there are not yet many integrated models that include two or more of these capabilities. Also, the successful examples in each case are typically limited to a few model systems and have not been generally extended to the hundreds of organisms relevant to DOE whose genomes are now available.

Recommendation 4. *The GTL modeling and simulation research program should be supported by an explicit series of investments in modeling technology, databases, algorithms, and software infrastructure needed to address the computational challenges.*

The appropriate early targets for a comprehensive attack on predictive biological modeling are specific functions of microbial organisms (e.g., cellular metabolism, motility, global transcription regulation and differentiation, and life-cycle development). The focus should include advancing the predictive skill on well-studied models (*e.g., E. coli, B. subtilis)* but begin to extend to those organisms that stretch the capability beyond the existing well-studied model systems (*e.g., Clostridium, Shewanella, Synechocystis*) and small consortia (communities) of microorganisms relevant to DOE missions, such as those associated with bioremediation, carbon sequestration, and nitrogen fixation and fermentation and degradation.

The subcommittee also recommends that lower eukaryotes (e.g., diatoms, coccolithopores, single-cell fungi) and plants be included as targets in longer-term modeling and simulation goals. Such inclusions will advance the goals of Genomics:GTL by strengthening efforts to integrate the modeling and advancing systems-level and synthetic knowledge for microbes and plants.

Charge Question 3.

> ***What are the key computational obstacles to developing computer models of the major biological understandings necessary to characterize and engineer microbes for DOE missions such as biofuels and bioremediation?***

Finding #6. A number of obstacles remain to reaching the visionary goal of a predictive model useful for engineering of an organism derived largely from its genome and related data. Five of the most relevant ones follow.

First, there is a lack of integrated genomics databases and the associated computational methods for supporting curation, extension, and visualization of comparative data explicitly focused on supporting the development of modeling and simulations for DOE-relevant organisms.

Second, for current systems-level computational analyses (e.g., flux balance analysis) work is needed to further integrate additional cellular physio-chemical constraints into modeling frameworks in order to generate computational predictions with greater accuracy towards defining the actual physiological state of the cell.

Third, there is a lack of robust mathematical frameworks and software implementing those frameworks for integrating models of metabolism with those of gene regulation that are two of most highly developed areas of modeling and simulation at the whole cell level, but whose mathematical representations are quite different.

Fourth, the multiscale mathematics and associated software libraries and tools for integrating processes in cellular models of disparate scales (e.g., molecular scale to that of the whole cell and microbial community) that would enable the modeling community to begin development of integrated whole-cell-scale models with atomistic simulations of specific mechanisms are lacking.

Fifth, a computational and analytical theory for framing all of computational biology is lacking. Such a theory should incorporate evolution as the basis for understanding and interpreting the results from comparative analysis. For example, the algorithms needed to make rapid progress on questions such as understanding the major forces governing the evolution of metabolism and regulatory networks have not yet been developed. Understanding these forces will be critical to creating the stable engineered strains needed for large-scale bioproduction of materials.

Recommendation 5. *DOE should establish a mechanism to support the long-term curation and integration of genomics and related datasets (annotations, metabolic reconstructions, expression data, whole genome screens, etc.) to support biological research in general and specifically the needs of modeling and simulation in particular in areas of energy and the environment that are not well supported by NSF and NIH.*

This mechanism should target the creation of a state-of-the-art community resource for data of all forms that are relevant to organisms of interest to DOE. This should be a joint activity of ASCR and BER, with ASCR responsible for the database and computational infrastructure to enable community annotation and data sharing. It should also leverage the work of established groups.

Recommendation 6. *DOE should work with the community to identify novel scientific opportunities for connecting modeling and simulation at the organism level to modeling and simulation at other space and temporal scales.*

Examples that could be investigated include integration of microbial models into ocean and terrestrial ecology models which in turn are coupled to global climate models, and models of bioremediation environments that can couple organism metabolic capabilities to external biogeochemistry. This multiscale coupling in beginning to be explored, but much more can be done, and it is likely to yield significant scientific insight.

Potential Impact according to Budget Levels made available

Of course, a significant commitment of funds would have a large impact and could well deliver on the objectives well ahead of schedule. However, to manage the potential difficulties with funding levels, the subcommittee has developed an estimate of what might be achieved for a given total funding level, jointly and equally supported by ASCR and BER. See Table 1 and its Addendum.

Table 1. Consideration of Funding Levels and Outcomes for an Expanded and Focused Office of Science Partnership on GTL	
Level ($M)	**Recommended Action (Funding and HQ effort is assumed as shared equally between BER and ASCR offices, save in baseline case)**
5	Initiate next step for GTL Bioenergy Centers: fund a GTL Knowledge Base (KB) by adding a Coordination Center (CC) to create GTL KB and a user-friendly interface (Portal); CC will work w the three GTL Centers, focusing on their common datasets and goals. This is a minimum effort to sustain GTL advances. BER to choose a maximum of three pilot projects to meet most important needs. Pilot activities chosen through needs-analyses by the three BioEnergy Centers. The KB CC will also provide $0.6M to each Center. The coordination funds are to work toward common standards and a single interface, while immediately providing BioIT support for Internet-based access and also, scientific connections to experimentalists. The KB will interact with other major knowledge resources.
10	Expand CC; advance Core Knowledge Environment, include all GTL projects; flexible approach adds the capability to add any future GTL Centers exist. The shared DOE effort would establish a more powerful GTL KB and the addition of additional pilots. GTL could begin funding some modeling activities and the KB would be better linked to other major knowledge resources. Include ASCR community collaborations on software tool development to improve usefulness of the KB. Increase the amount of annual BioIT support from the CC to GTL projects to ensure presence of strong, internal bioIT efforts. Initiate funding for some visualization ASCR-supported tools for data analysis. Include focus on simulation and modeling within GTL PI Annual Meeting
30	Expand rapidly the level of support for computational approaches (algorithm and software packages) across ASCR community. Provide major funding for well-interconnected collaborations among experimentalists and quantitative scientists. Establish joint Solicitations w other federal agencies to enable more comprehensive contributions of computing and apply wide range of biology Advances toward GTL goals through a more extensive KB representing the fully needed range of reference data beyond GTL data. Expand modeling and simulation to separate annual meeting with a validation/CASP-like process and include the broad Biological Community.
50	Full implement of the opportunities in joint subcommittee report for robust individual and group efforts in algorithm and software. Validation of software and of computational predictions would be intrinsic to collaborations among GTL biologists and computing. Progress would now be driven and focused by a deep engagement among the GTL experimentalists and the modeling groups. Extensive collaborations, comprehensive validations, and connections to the broader efforts in systems and synthetic biology enable GTL KB to contribute to wide range of DOE missions and GTL science, as a whole, to serve society

Table 1 Addenda. Description of the Consequences for DOE GTL from the Implementation of the Set of Levels of Funding	
Level ($M)	**Impact**
5	The baseline would include only the initiation of the proposed GTL KB, sustaining minimal expectation of GTL Bioenergy Centers. There would be NO implementation of the findings of the study by the Joint Subcommittee re potential contributions of a Science Partnership that would bring computing fully into GTL effort and stimulate progress.
10	The specifics of the pilot projects to nucleate and build the GTL Knowledge Base itself, the coordination activities, the integration of GTL data, and the response to needs of the three Bioenergy Centers are given in the GTL KB report for BER. More rapid progress by GTL activities, Centers, could now begin, in which key community software tools would be provided by ASCR funding.
15	At this level, a partnership to accelerate and expand the impact of GTL in meeting DOE mission needs can begin. Besides extending GTL modeling and simulation research, the KB will be able to work more effectively with the entire biology community and with the ASCR computational science community can contribute needed software and a powerful, readily accessed, internet based, GTL Knowledge Base can be created by the distributed DOE community to enhance the outreach of GTL and well as internal progress.
30	This funding level, 30M/yr jointly, would provide an increased implementation of the goals outlined in this report, and position the DOE to build key partnerships with other agencies in order to broaden support for software for systems and synthetic biology, which GTL researchers would incorporate and integrate. Contributions in GTL-empowered biology funded by other agencies would be included in GTL Knowledge for enhanced value for DOE investigators and underpin DOE applied mission needs. Validation of modeling and simulation by way of open, competitive review at the Annual Meeting will enhance recognition by the Experimental Biology community, and thus, increasing the GTL impact on science and society and opening new collaborations.
50	The optimum level, 50M/yr, would allow full delivery for DOE of the opportunities for GTL from modeling and simulation; the full GTL effort would now include intensive collaborative modeling and simulation efforts; GTL would include the integration of the Experimental work and the Computational Modeling and Simulation efforts, that is, there would be feedback from computing to the GTL Center's experimental decisions and foci, and GTL experimental findings would inform and direct modeling to achieve full Impact. All DOE /energy-related organisms would be included in computational effort. At this level, the Office of Science of the DOE would play a Leadership role for Nation and significantly catalyze the growth of GTL and its impact on the societal goals of DOE mission efforts on bioremediation, carbon sequestration and the delivery of the bioenergy vision

Conclusions

The time is right, given a convergence of the DOE-funded advances in biology and in mathematics and computing, for a major effort by the DOE Office of Science to unite the disparate accomplishments of the past decade into a larger-scale activity aimed at addressing the grand challenges of modeling and simulation for the Genomics: GTL program. This new activity can build on advanced capabilities from programs such as ASCR's SciDAC and the base computing research activities, and at the same time, couple them with the path-breaking research supported by the Genomics: GTL program. This new thrust would provide DOE with a long-term research capability that will help realize the community's Genomics: GTL vision and provide a basis for cooperative agreements with agencies such as NIH and NSF in the fast-emerging area of systems and synthetic biology.

Participants and Agenda

Joint ASCAC – BERAC Panel Meeting, October 4-5, 2007
Site: the Gordon and Betty Moore Foundation (GBMF) Office, San Francisco

Panel Members:
Michael Banda — LBNL
David Galas — ISB/Battelle
Keith Hodgson — Stanford
David Kingsbury — Moore Foundation
Chris Somerville — Stanford
Rick Stevens — ANL/UChicago
Barbara Wold — Caltech
John Wooley — UCSD
Thomas Zacharia — ORNL

Invited Presenters:
Nitin Baliga — ISB/U Washington
Rich Bonneau — NYU/Courant
Paramvir Dehal — LBNL/UCB
Justin Donato — U Wisconsin
Thierry Emonet — Yale U
Adam Feist — UCSD
Mick Follows — MIT
Peter Karp — SRI
Harley McAdams — Stanford
Sue Rhee — Carnegie InstiCtution
Nagiza Samatova — ORNL

Agenda:

Thursday October 4, 2007

9:-10:00	Executive Session – Coffee etc. will be available
10-10:45	Sue Rhee
10:45-11:30	Nitin Baliga
11:30-12:15	Harley McAdams
12:15-1:30	Lunch and discussion of the morning sessions
1:30-2:15	Rich Bonneau
2:15-3	Thierry Emonet
3-3:15	BREAK
3:15-4	Adam Feist
4-4:45	Justin Donato
4:45-6	Discussion of the p.m. sessions and wrap up of the first day
6-	Dinner (location TBD)

Friday, October 5, 2007

8-8:30	Executive Session – Coffee etc. will be available
8:30-9:15	Paramvir Dehal
9:15-10	Mick Follows
10-10:15	BREAK
10:15-11	Nagiza Samatova
11-11:45	Peter Karp
11:45-1	Lunch and discussion of the morning sessions
1-4:00	Executive Session, writing, etc.

Bibliography

Austin, D. W., M. S. Allen, et al. (2006). "Gene network shaping of inherent noise spectra." Nature **439**(7076): 608-11.

Baliga, N. S. (2001). "Promoter analysis by saturation mutagenesis." Biol Proced Online **3**: 64-69.

Baliga, N. S., M. Pan, et al. (2002). "Coordinate regulation of energy transduction modules in Halobacterium sp. analyzed by a global systems approach." Proc Natl Acad Sci U S A **99**(23): 14913-8.

Baliga, N. S., R. Bonneau, et al. (2004). "Genome sequence of Haloarcula marismortui: a halophilic archaeon from the Dead Sea." Genome Res **14**(11): 2221-34.

Baliga, N. S., S. J. Bjork, et al. (2004). "Systems level insights into the stress response to UV radiation in the halophilic archaeon Halobacterium NRC-1." Genome Res **14**(6): 1025-35.

Baliga, N. S., S. P. Kennedy, et al. (2001). "Genomic and genetic dissection of an archaeal regulon." Proc Natl Acad Sci U S A **98**(5): 2521-5.

Bard, J., S. Y. Rhee, et al. (2005). "An ontology for cell types." Genome Biol **6**(2): R21.

Bare, J. C., P. T. Shannon, et al. (2007). "The Firegoose: two-way integration of diverse data from different bioinformatics web resources with desktop applications." BMC Bioinformatics **8**: 456.

Becker, S. A., A. M. Feist, et al. (2007). "Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox." Nat Protoc **2**(3): 727-38.

Berardini, T. Z., S. Mundodi, et al. (2004). "Functional annotation of the Arabidopsis genome using controlled vocabularies." Plant Physiol **135**(2): 745-55.

Bernaerts, K., E. Dens, et al. (2004). "Concepts and tools for predictive modeling of microbial dynamics." J Food Prot **67**(9): 2041-52.

Bonneau, R., D. J. Reiss, et al. (2006). "The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo." Genome Biol **7**(5): R36.

Bonneau, R., M. T. Facciotti, et al. (2007). "A predictive model for transcriptional control of physiology in a free living cell." Cell **131**(7): 1354-65.

Bonneau, R., N. S. Baliga, et al. (2004). "Comprehensive de novo structure prediction in a systems-biology context for the archaea Halobacterium sp. NRC-1." Genome Biol **5**(8): R52.

Boyd, P. W., T. Jickells, et al. (2007). "Mesoscale iron enrichment experiments 1993-2005: synthesis and future directions." Science **315**(5812): 612-7.

Cakir, T., C. Efe, et al. (2007). "Flux balance analysis of a genome-scale yeast model constrained by exometabolomic data allows metabolic system identification of genetically different strains." Biotechnol Prog **23**(2): 320-6.

Caspi, R., H. Foerster, et al. (2006). "MetaCyc: a multiorganism database of metabolic pathways and enzymes." Nucleic Acids Res **34**(Database issue): D511-6.

Caspi, R., H. Foerster, et al. (2008). "The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases." Nucleic Acids Res **36**(Database issue): D623-31.

Chen, S. L., W. Lee, et al. (2004). "Codon usage between genomes is constrained by genome-wide mutational processes." Proc Natl Acad Sci U S A **101**(10): 3480-5.

Clarke, B., M. Lambrecht, et al. (2003). "Arabidopsis genomic information for interpreting wheat EST sequences." Funct Integr Genomics **3**(1-2): 33-8.

Collier, J., H. H. McAdams, et al. (2007). "A DNA methylation ratchet governs progression through a bacterial cell cycle." Proc Natl Acad Sci U S A **104**(43): 17111-6.

Cordes, E. E., M. A. Arthur, et al. (2005). "Modeling the mutualistic interactions between tubeworms and microbial consortia." PLoS Biol **3**(3): e77.

Covert, M. W., C. H. Schilling, et al. (2001). "Metabolic modeling of microbial strains in silico." Trends Biochem Sci **26**(3): 179-86.

Crosson, S., H. McAdams, et al. (2004). "A genetic oscillator and the regulation of cell cycle progression in Caulobacter crescentus." Cell Cycle **3**(10): 1252-4.

Dassarma, S., S. P. Kennedy, et al. (2001). "Genomic perspective on the photobiology of Halobacterium species NRC-1, a phototrophic, phototactic, and UV-tolerant haloarchaeon." Photosynth Res **70**(1): 3-17.
Dehal, P. S. and J. L. Boore (2006). "A phylogenomic gene cluster resource: the Phylogenetically Inferred Groups (PhIGs) database." BMC Bioinformatics **7**: 201.
Deich, J., E. M. Judd, et al. (2004). "Visualization of the movement of single histidine kinase molecules in live Caulobacter cells." Proc Natl Acad Sci U S A **101**(45): 15921-6.
Edwards, J. S. and B. O. Palsson (2000). "Metabolic flux balance analysis and the in silico analysis of Escherichia coli K-12 gene deletions." BMC Bioinformatics **1**: 1.
Emonet, T., C. M. Macal, et al. (2005). "AgentCell: a digital single-cell assay for bacterial chemotaxis." Bioinformatics **21**(11): 2714-21.
Facciotti, M. T., D. J. Reiss, et al. (2007). "General transcription factor specified global gene regulation in archaea." Proc Natl Acad Sci U S A **104**(11): 4630-5.
Facciotti, M. T., V. S. Cheung, et al. (2004). "Specificity of anion binding in the substrate pocket of bacteriorhodopsin." Biochemistry **43**(17): 4934-43.
Feist, A. M., C. S. Henry, et al. (2007). "A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information." Mol Syst Biol **3**: 121.
Feist, A. M., J. C. Scholten, et al. (2006). "Modeling methanogenesis with a genome-scale metabolic reconstruction of Methanosarcina barkeri." Mol Syst Biol **2**: 2006 0004.
Follows, M. J., S. Dutkiewicz, et al. (2007). "Emergent biogeography of microbial communities in a model ocean." Science **315**(5820): 1843-6.
Garcia-Hernandez, M., T. Z. Berardini, et al. (2002). "TAIR: a resource for integrated Arabidopsis data." Funct Integr Genomics **2**(6): 239-53.
Goo, Y. A., J. Roach, et al. (2004). "Low-pass sequencing for microbial comparative genomics." BMC Genomics **5**(1): 3.
Goryanin, II, G. V. Lebedeva, et al. (2006). "Cellular kinetic modeling of the microbial metabolism." Methods Biochem Anal **49**: 437-88.
Green, M. L. and P. D. Karp (2004). "A Bayesian method for identifying missing enzymes in predicted metabolic pathway databases." BMC Bioinformatics **5**: 76.
Green, M. L. and P. D. Karp (2007). "Using genome-context data to identify specific types of functional associations in pathway/genome databases." Bioinformatics **23**(13): i205-11.
Heffelfinger, G. S., A. Martino, et al. (2002). "Carbon sequestration in Synechococcus Sp.: from molecular machines to hierarchical modeling." Omics **6**(4): 305-30.
Henson, M. A. (2003). "Dynamic modeling of microbial cell populations." Curr Opin Biotechnol **14**(5): 460-7.
Heuett, W. J. and H. Qian (2006). "Combining flux and energy balance analysis to model large-scale biochemical networks." J Bioinform Comput Biol **4**(6): 1227-43.
Hottes, A. K., L. Shapiro, et al. (2005). "DnaA coordinates replication initiation and cell cycle transcription in Caulobacter crescentus." Mol Microbiol **58**(5): 1340-53.
Hottes, A. K., M. Meewan, et al. (2004). "Transcriptional profiling of Caulobacter crescentus during growth on complex and minimal media." J Bacteriol **186**(5): 1448-61.
Hu, P., E. L. Brodie, et al. (2005). "Whole-genome transcriptional analysis of heavy metal stresses in Caulobacter crescentus." J Bacteriol **187**(24): 8437-49.
Huala, E., A. W. Dickerman, et al. (2001). "The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant." Nucleic Acids Res **29**(1): 102-5.
Ishii, N., M. Robert, et al. (2004). "Toward large-scale modeling of the microbial cell for computer simulation." J Biotechnol **113**(1-3): 281-94.
Iyer, R., N. S. Baliga, et al. (2005). "Catabolite control protein A (CcpA) contributes to virulence and regulation of sugar metabolism in Streptococcus pneumoniae." J Bacteriol **187**(24): 8340-9.
Jeffries, T. W., I. V. Grigoriev, et al. (2007). "Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast Pichia stipitis." Nat Biotechnol **25**(3): 319-26.

Judd, E. M., L. R. Comolli, et al. (2005). "Distinct constrictive processes, separated in time and space, divide caulobacter inner and outer membranes." J Bacteriol **187**(20): 6874-82.
Karp, P. D. (2004). "Call for an enzyme genomics initiative." Genome Biol **5**(8): 401.
Karp, P. D., C. A. Ouzounis, et al. (2005). "Expansion of the BioCyc collection of pathway/genome databases to 160 genomes." Nucleic Acids Res **33**(19): 6083-9.
Karp, P. D., I. M. Keseler, et al. (2007). "Multidimensional annotation of the Escherichia coli K-12 genome." Nucleic Acids Res **35**(22): 7577-90.
Kato Marcus, A., C. I. Torres, et al. (2007). "Conduction-based modeling of the biofilm anode of a microbial fuel cell." Biotechnol Bioeng **98**(6): 1171-82.
Kauffman, K. J., P. Prakash, et al. (2003). "Advances in flux balance analysis." Curr Opin Biotechnol **14**(5): 491-6.
Kaur, A., M. Pan, et al. (2006). "A systems view of haloarchaeal strategies to withstand stress from transition metals." Genome Res **16**(7): 841-54.
Keseler, I. M., J. Collado-Vides, et al. (2005). "EcoCyc: a comprehensive database resource for Escherichia coli." Nucleic Acids Res **33**(Database issue): D334-7.
Klotz, B., D. L. Pyle, et al. (2007). "New mathematical modeling approach for predicting microbial inactivation by high hydrostatic pressure." Appl Environ Microbiol **73**(8): 2468-78.
Korobkova, E. A., T. Emonet, et al. (2006). "Hidden stochastic nature of a single bacterial motor." Phys Rev Lett **96**(5): 058105.
Krummenacker, M., S. Paley, et al. (2005). "Querying and computing with BioCyc databases." Bioinformatics **21**(16): 3454-5.
Laub, M. T., L. Shapiro, et al. (2007). "Systems biology of Caulobacter." Annu Rev Genet **41**: 429-41.
Le, T. T., S. Harlepp, et al. (2005). "Real-time RNA profiling within a single bacterium." Proc Natl Acad Sci U S A **102**(26): 9160-4.
Le, T. T., T. Emonet, et al. (2006). "Dynamical determinants of drug-inducible gene expression in a single bacterium." Biophys J **90**(9): 3315-21.
Lee, J. M., E. P. Gianchandani, et al. (2006). "Flux balance analysis in the era of metabolomics." Brief Bioinform **7**(2): 140-50.
Lee, T. J., Y. Pouliot, et al. (2006). "BioWarehouse: a bioinformatics database warehouse toolkit." BMC Bioinformatics **7**: 170.
Li, H., A. Coghlan, et al. (2006). "TreeFam: a curated database of phylogenetic trees of animal gene families." Nucleic Acids Res **34**(Database issue): D572-80.
McAdams, H. H. (2006). "Bacterial stalks are nutrient-scavenging antennas." Proc Natl Acad Sci U S A **103**(31): 11435-6.
McCollum, J. M., G. D. Peterson, et al. (2006). "The sorting direct method for stochastic simulation of biochemical systems with varying reaction execution behavior." Comput Biol Chem **30**(1): 39-49.
McGrath, P. T., H. Lee, et al. (2007). "High-throughput identification of transcription start sites, conserved promoter motifs and predicted regulons." Nat Biotechnol **25**(5): 584-92.
McGrath, P. T., P. Viollier, et al. (2004). "Setting the pace: mechanisms tying Caulobacter cell-cycle progression to macroscopic cellular events." Curr Opin Microbiol **7**(2): 192-7.
Mohamed, M. M. and K. Hatfield (2005). "Modeling microbial-mediated reduction in batch reactors." Chemosphere **59**(8): 1207-17.
Mueller, L. A., P. Zhang, et al. (2003). "AraCyc: a biochemical pathway database for Arabidopsis." Plant Physiol **132**(2): 453-60.
Mukhopadhyay, A., A. M. Redding, et al. (2007). "Cell-wide responses to low-oxygen exposure in Desulfovibrio vulgaris Hildenborough." J Bacteriol **189**(16): 5996-6010.
National Research Council (2005), "Mathematics and 21st Century Biology," a NAS NRC Board on Mathematical Sciences' study committee chaired by M. Olson, plus fourteen others and J. Wooley (supported by ASCR).

National Research Council (2005), “Catalyzing Inquiry at the Interface of Computing and Biology,” a NAS NRC Board on Computer Sciences study committee chaired by J. Wooley, plus ten others and J. Schwaber (supported by BER among others).
National Research Council (2008), “The Role of Theory in Advancing 21st Century Biology,” a NAS NRC Board on Life Sciences study committee chaired by D. Galas, plus fifteen others and G. Wagner.
Ng, W. V., S. P. Kennedy, et al. (2000). "Genome sequence of Halobacterium species NRC-1." Proc Natl Acad Sci U S A **97**(22): 12176-81.
Ostrouchov, G. and N. F. Samatova (2005). "On FastMap and the convex hull of multivariate data: toward fast and robust dimension reduction." IEEE Trans Pattern Anal Mach Intell **27**(8): 1340-3.
Paley, S. M. and P. D. Karp (2006). "The Pathway Tools cellular overview diagram and Omics Viewer." Nucleic Acids Res **34**(13): 3771-8.
Pan, C., G. Kora, et al. (2006). "ProRata: A quantitative proteomics program for accurate protein abundance ratio estimation with confidence interval evaluation." Anal Chem **78**(20): 7121-31.
Patnaik, P. R. (2001). "Microbial metabolism as an evolutionary response: the cybernetic approach to modeling." Crit Rev Biotechnol **21**(3): 155-75.
Pouliot, Y. and P. D. Karp (2007). "A survey of orphan enzyme activities." BMC Bioinformatics **8**: 244.
Price, M. N., P. S. Dehal, et al. (2007). "Orthologous transcription factors in bacteria have different functions and regulate different genes." PLoS Comput Biol **3**(9): 1739-50.
Price, M. N., P. S. Dehal, et al. (2008). "Horizontal gene transfer and the evolution of transcriptional regulation in Escherichia coli." Genome Biol **9**(1): R4.
Prommer, H., M. E. Grassi, et al. (2007). "Modeling of microbial dynamics and geochemical changes in a metal bioprecipitation experiment." Environ Sci Technol **41**(24): 8433-8.
Ramirez, I. and J. P. Steyer (2008). "Modeling microbial diversity in anaerobic digestion." Water Sci Technol **57**(2): 265-70.
Reiser, L., L. A. Mueller, et al. (2002). "Surviving in a sea of data: a survey of plant genome data resources and issues in building data management systems." Plant Mol Biol **48**(1-2): 59-74.
Reiss, D. J., M. T. Facciotti, et al. (2008). "Model-based deconvolution of genome-wide DNA binding." Bioinformatics **24**(3): 396-403.
Reiss, D. J., N. S. Baliga, et al. (2006). "Integrated biclustering of heterogeneous genome-wide datasets for the inference of global regulatory networks." BMC Bioinformatics **7**: 280.
Rhee, S. Y. (2000). "Bioinformatic resources, challenges, and opportunities using Arabidopsis as a model organism in a post-genomic era." Plant Physiol **124**(4): 1460-4.
Rhee, S. Y. and B. Crosby (2005). "Biological databases for plant research." Plant Physiol **138**(1): 1-3.
Rhee, S. Y., W. Beavis, et al. (2003). "The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community." Nucleic Acids Res **31**(1): 224-8.
Schilling, C. H., J. S. Edwards, et al. (2000). "Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems." Biotechnol Bioeng **71**(4): 286-306.
Schlueter, S. D., M. D. Wilkerson, et al. (2005). "Community-based gene structure annotation." Trends Plant Sci **10**(1): 9-14.
Schmid, A. K., D. J. Reiss, et al. (2007). "The anatomy of microbial cell state transitions in response to oxygen." Genome Res **17**(10): 1399-413.
Shannon, P. T., D. J. Reiss, et al. (2006). "The Gaggle: an open-source software system for integrating bioinformatics software and data sources." BMC Bioinformatics **7**: 176.
Shannon, P., A. Markiel, et al. (2003). "Cytoscape: a software environment for integrated models of biomolecular interaction networks." Genome Res **13**(11): 2498-504.
Shastri, A. A. and J. A. Morgan (2005). "Flux balance analysis of photoautotrophic metabolism." Biotechnol Prog **21**(6): 1617-26.
Sloan, W. T., S. Woodcock, et al. (2007). "Modeling taxa-abundance distributions in microbial communities using environmental sequence data." Microb Ecol **53**(3): 443-55.

Stolyar, S., S. Van Dien, et al. (2007). "Metabolic modeling of a mutualistic microbial community." Mol Syst Biol **3**: 92.
Thimm, O., O. Blasing, et al. (2004). "MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes." Plant J **37**(6): 914-39.
Tyler, B. M., S. Tripathy, et al. (2006). "Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis." Science **313**(5791): 1261-6.
Vallino, J. J. (2003). "Modeling microbial consortiums as distributed metabolic networks." Biol Bull **204**(2): 174-9.
Viollier, P. H., M. Thanbichler, et al. (2004). "Rapid and sequential movement of individual chromosomal loci to specific subcellular locations during bacterial DNA replication." Proc Natl Acad Sci U S A **101**(25): 9257-62.
Wahr, J. A., M. Gerber, et al. (2001). "Allosteric modification of oxygen delivery by hemoglobin." Anesth Analg **92**(3): 615-20.
Weston, A. D., N. S. Baliga, et al. (2003). "Systems approaches applied to the study of Saccharomyces cerevisiae and Halobacterium sp." Cold Spring Harb Symp Quant Biol **68**: 345-57.
Whitehead, K., A. Kish, et al. (2006). "An integrated systems approach for understanding cellular responses to gamma radiation." Mol Syst Biol **2**: 47.
Yu, G. X., B. H. Park, et al. (2005). "An evolution-based analysis scheme to identify CO2/O2 specificity-determining factors for ribulose 1,5-bisphosphate carboxylase/oxygenase." Protein Eng Des Sel **18**(12): 589-96.
Yu, G. X., B. H. Park, et al. (2005). "In silico discovery of enzyme-substrate specificity-determining residue clusters." J Mol Biol **352**(5): 1105-17.
Yu, G. X., G. Ostrouchov, et al. (2003). "An SVM-based algorithm for identification of photosynthesis-specific genome features." Proc IEEE Comput Soc Bioinform Conf **2**: 235-43.
Zhang, B., N. C. VerBerkmoes, et al. (2006). "Detecting differential and correlated protein expression in label-free shotgun proteomics." J Proteome Res **5**(11): 2909-18.

Charge Letter

Under Secretary for Science
Washington, DC 20585

FEB 22 2007

Dr. Jill P. Dahlburg, Chair
Naval Research Laboratory, Code 1001
4555 Overlook Avenue
Washington, DC 20375

Dear Dr. Dahlburg:

I am requesting that the Advanced Scientific Computing Advisory Committee (ASCAC) undertake two charges this year.

1) The August 15, 2003 charge to ASCAC instituted a Committee of Visitors (COV) to assess the program management of major elements of Advanced Scientific Computing Research (ASCR) program every two to three years. The first two COV reviews - of the research program and the facilities efforts – resulted in a number of improvements to the processes. Following on these reviews I now ask ASCAC to conduct a COV review of the SciDAC efforts within ASCR. A report to ASCAC should be planned for the Fall 2007 ASCAC meeting.

2) Next, I would like ASCAC to convene a Joint Panel with the Biological and Environmental Research Advisory Committee (BERAC) to examine the issue of computational models for GTL and how progress could be accelerated through targeted investments in applied mathematics, computer science and computational biology. The Joint Panel should consider whether the current ASCR long-term goal described below is too ambitious given the status and level of buy-in from the community. The Joint Panel should also discuss possible intermediate goals that might be more relevant to the two programs. Finally, the Joint Panel should identify the key computational obstacles to developing computer models of the major biological understandings necessary to characterize and engineer microbes for DOE missions such as biofuels and bioremediation. The context for this charge is the fact that both ASCR and Biological Environmental Research (BER) have long-term goals for developing GTL models:

    a. (ASCR) By 2015, demonstrate progress toward developing, through the Genomes to Life partnership with the Biological and Environmental Research program, the computational science capability to model a complete microbe and a simple microbial community.
    b. (BER) By 2015, provide sufficient scientific understanding of plants and microbes to develop robust new strategies to produce biofuels, clean up waste, or sequester carbon. This includes research that supports the development of computational models to direct the use and design of improved organisms carrying out these processes.

In 2006, both ASCAC and BERAC reviewed progress toward the respective goals. BERAC rated progress as "excellent" and ASCAC rated progress as "good" with "concerns regarding adequate communication and reasonable timescales for achieving research objectives."

To inform the FY 2009 Budget Request, I would like a full report on the findings and recommendations at the August 2007 ASCAC meeting. I appreciate ASCAC's willingness to undertake this important activity.

Sincerely,

Raymond L. Orbach