

**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, BerkeleyLab

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 met in October 2007 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 effort, the barriers to success and strategies for overcoming these barriers. Experts from the community were involved in a two-day workshop and discussed the status of current research and potential goals for the future. The panel concluded that there has been substantial progress over the past five years in the development of techniques for building computational models of microbes (e.g. metabolic networks and regulatory networks), and analysis of high-throughput datasets and for integration and visualization of biological datasets and that further progress is highly likely given the state of experimental techniques and progress in computational capabilities. The panel makes six recommendations for advancing the program and addressing the charge questions.

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

(ASCR) By 2018, demonstrate significant advances in the capability to predict an organism's phenotype from its genome sequence, through advances in genome sequence annotation, whole genome modeling and simulation, and integrated model-driven experimentation.

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 support the long-term curation and integration of genomics and related datasets (annotations, metabolic reconstructions, expression data, whole genome screens, phenotype data, 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.*
- *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.*

Introduction

The ASCR-BERAC joint 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 we were 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 better address the community's needs.

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, not only of how biological systems function, but also of how experimental, theoretical, and computational programs can be coupled effectively.

This report addresses key issues involved in ensuring continued progress in developing the new biology.

Background

In the past decade, bioinformatics has been used to support advanced microbial biotechnology in many ways: computationally analysis of wet-lab data, genome sequencing analysis, identification of protein coding genes, genome comparison to identify the gene function, development of genomic and proteomics databases, and inference of phenotypes (higher-level functions) from genotypes (gene-level 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 pathways; modeling of the 2D and 3D structure 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: 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; 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 the integration of sequence analysis, database, and search techniques with mathematical modeling.

Some Ways Models are Useful in Biology

- Models Provide a Coherent Framework for Interpreting Data
- Models Highlight Basic Concepts of Wide Applicability
- Models Uncover New Phenomena or Concepts to Explore
- Models Identify Key Factors or Components of a System
- Models Can Link Levels of Detail (Individual to Population)
- Models Enable the Formalization of Intuitive Understandings
- Models Can Be Used as a Tool for Helping to Screen Unpromising Hypotheses
- Models Inform Experimental Design
- Models Can Predict Variables Inaccessible to Measurement
- Models Can Link What Is Known to What Is Yet Unknown
- Models Can Be Used to Generate Accurate Quantitative Predictions
- Models Expand the Range of Questions That Can Meaningfully Be Asked

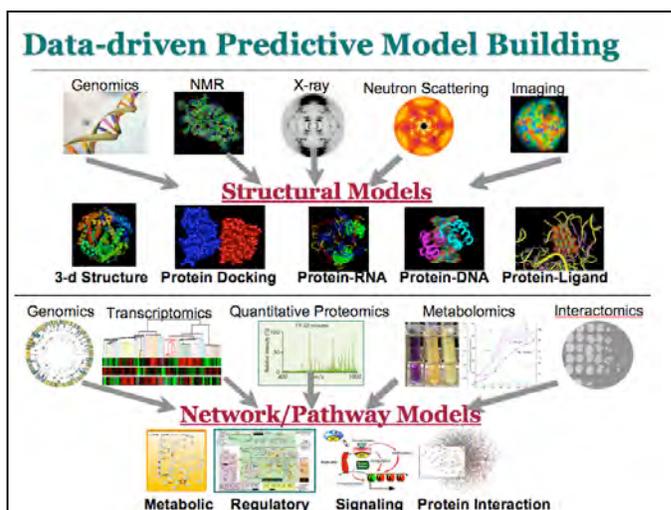
From the NRC report "Catalyzing Inquiry at the Interface of Computing and Biology"

It is the last of these ideas that we are concerned with in this report: the idea of integrating bioinformatics approaches with mathematical modeling and simulation to support advances in systems-level understanding of complex biological systems.

We present our results 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. The presentations and other material clearly 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 future. Nevertheless, at least two factors are holding the community back.



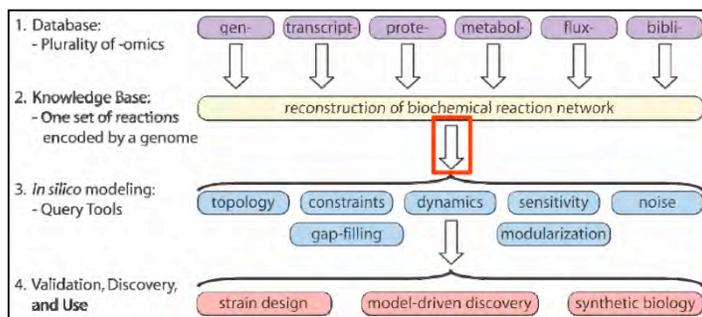
The first factor is that small groups with limited funding are pursuing the vast majority of the current work, with uncertain levels and duration of funding often requiring the integration of multiple funding sources from multiple agencies over time to maintain progress. This prevents the development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

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 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 (i.e., 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. 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 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:

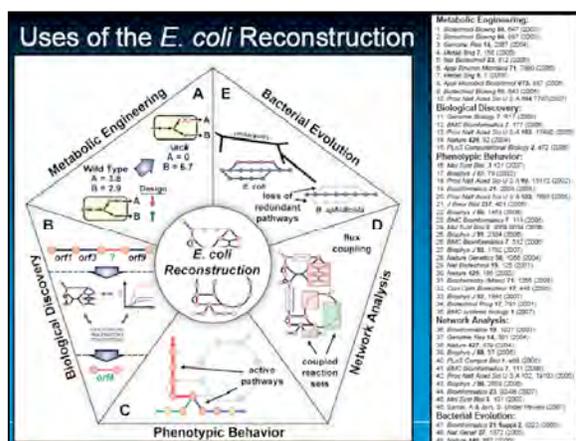
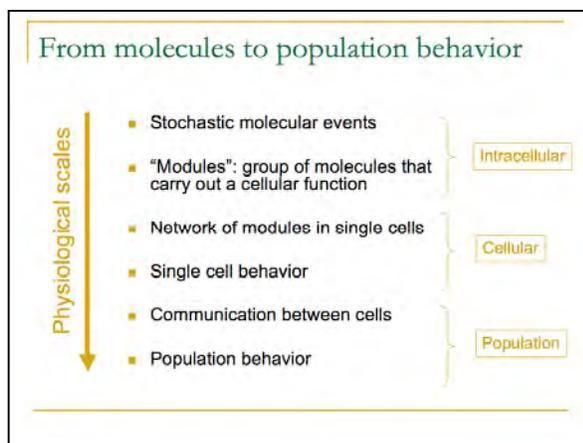
The clear consensus was that these earlier efforts were limited by a lack of experimental data and the means to verify the models quantitatively. There also was agreement on the key requirements necessary to create a successful new biology. The methods and results of quantitative and predictive biology 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.



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. We know from experience in other disciplines (e.g., electronics design) that progress can be greatly accelerated when simple computational tools and models (e.g., early VLSI 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.

With the number of completed genome sequences reaching 1,000 in the next few years, we are on the verge of a new class of biological problem: 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 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. Our 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 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

We organize the findings and recommendations 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 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 the target of GTL funding focused largely on developing the modeling capabilities.

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 our 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. We also note 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 for the joint modeling and simulation activity of ASCR and BER should be modified to read as follows:*

(ASCR) By 2018, demonstrate significant advances in the capability to predict an organism's phenotype from its genome sequence, through advances in genome sequence annotation, whole genome modeling and simulation, and integrated model-driven experimentation.

This PART goal should be accompanied by a specific set of metrics of progress. Such metrics could include for a given organism the fraction of an organism's genes and gene products included in a model, number of correct metabolic phenotype measurements predicted, number of transcription regulatory elements in a model, number of correct gene expression experiments predicted, and 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 equal 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, 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 support accelerating 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 we have useful predictive models for core metabolism, global transcription regulation, signaling and motility control, and life-cycle development and differentiation. However, we do not yet have 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 push on 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.

We also recommend that the lower eukaryotes (e.g., diatoms, coccolithophores, single-cell fungi) and plants be included as targets in longer-term modeling and simulation goals.

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. Here we describe four of the most relevant ones.

First, we lack 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, we lack 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.

Third, we lack 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.

Fourth, we lack a computational and analytical theory for framing all of computational biology. Such a theory should incorporate evolution as the basis for understanding and interpreting the results from comparative analysis. For example, we have not yet developed the algorithms needed to make rapid progress on questions such as understanding the major forces governing the evolution of metabolism and regulatory networks. 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, phenotype data, 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 is beginning to be explored, but much more can be done, and it is likely to yield significant scientific insight.

Conclusions

The time is right for DOE to make a major push to bring together what has been accomplished over 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 couple them with the path-breaking research supported by the Genomics:GTL program. This new thrust can 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 biology.

Participants and Agenda

October 4-5, 2007, Joint Panel Meeting at the Moore Foundation 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 Institution
Nagiza Samatova — ORNL