Computational Cell Biology – Challenges and Opportunities for an Emerging Field

A Report Based on a Roundtable Discussion at the
First International Symposium on Computational Cell Biology
Cranwell Resort
Lenox, Massachusetts, March, 2001

Prepared by: Dr. Leslie M. Loew, Dr. Ann Cowan, and Dr. Ion Moraru
University of Connecticut Health Center

Cell Biology is the most immediate beneficiary of the flood of data that is emerging from the genomics and proteomics enterprises. However, this very wealth of data threatens to overwhelm the bench biologist, because traditional scientific methodologies would be impeded by the complexity and sheer volume of the information available to build hypotheses. Furthermore, if a complex hypothesis can be constructed, it may be extraordinarily difficult to grasp its implications or to develop experimentally testable predictions from it. Indeed, even simple cell biological systems can require quantitative analyses that are beyond the repertoire of traditional tools available to the experimentalist. Therefore, cell biologists have begun to tentatively explore computational approaches toward organizing data into quantitative models. This has also led to a sudden interest in interactions with mathematicians and physicists.

However, the different cultures of the biological science community and the physical and mathematical science communities represent a challenging barrier to effective collaboration. Biologists usually stretch the limits of reductionism in their science. They revel in the enormity, variety, mutability, and diversity of the biological universe. Mathematicians and physical scientists value unifying principles and have developed analytical approaches toward extracting common patterns from seemingly unrelated phenomena. One community prefers to consider only spherical cows while the other considers such an exercise to be without value. One community views the results of experiments as the only worthwhile end, while the other often considers an abstract mathematical construct to be the ultimate representation of reality. Both approaches are totally appropriate within their respective contexts. However, when theorists have made forays into biology, their efforts have been generally met with indifference from biologists.

Nonetheless, necessity is now making bedfellows of these two camps, as judged by the excitement generated during the first International Symposium on Computational Cell Biology (CCB) held at the Cranwell Resort in Lenox, Massachusetts this past March. The gathering included nearly equal representation from theorists and experimentalists and was sponsored by the National Center for Research Resources (NCRR) of the National Institutes of Health (NIH) through a conference grant. Support for student travel grants was also obtained from the National Science Foundation (NSF). All 125 registration slots were filled.

This document summarizes ideas from a roundtable discussion held near the end of the meeting. This forum, originally suggested by NCRR director Dr. Judith Vaitukaitis, focused on directions and needs for the emerging field of computational cell biology. Dr. Vaitukaitis had previously agreed to chair the discussion; however, she was unable to attend the session. Therefore, Dr. Abraham Levy of NCRR's Division of Biomedical Technology led a lively discussion that identified a series of issues and opportunities.

Issues

The discussion identified issues that were both cultural and technical. The often-interrelated challenges to the field are listed below. Many of the technical issues are individually common to other interdisciplinary technology-driven biomedical research fields that benefit from computational approaches (e.g., large information-rich data objects in neuroimaging, complexity in genetic databases, computationally intense calculations in molecular dynamics, etc.). However, collectively, they are unique to computational cell biology.

Although the number of scientists interested in the field of computational cell biology is growing and the need for the field is clear, mechanisms for establishing a research community are largely absent. The success of the CCB Symposium attests to the demand for appropriate forums for scientists to share ideas and establish interactions. An immediate response was to establish a LISTSERV® in computational cell biology that is now being hosted by the Pittsburgh Super Computer Center and to start planning for the second CCB Symposium in 2003; both of these initiatives received unanimous support by the attendees of the discussion forum.

Also related to the challenge of building and supporting a community are the particular requirements associated with sharing complex models and communicating the results of modeling research. Firstly, complex spatial models can produce enormously large datasets. In our own Virtual Cell modeling, each datum is a double floating point number representing the value of one of dozens of variables at each of several hundred time points; furthermore, each variable at each time point can be mapped onto a 3-D grid corresponding to the geometry of the cellular system. Thus, we have produced datasets as large as 10Gb from one simulation. Secondly, the publication of models in available traditional flat formats, even in the form of electronic publications, does not permit the full exercise, validation, and extension of models and simulations by the research community. Interactive peer-reviewed publication formats are technically feasible and will realize the full potential of the Internet as an ideal medium for scientific communication.

A constant theme of the discussion was that the modelers and the bench researchers need to work with a common purpose and in close cooperation. This will eliminate the cultural divisions of the past. It will also assure that the appropriate quantitative approaches will be brought to bear on the most important biological issues. Therefore, to move the field forward, mathematicians, physicists, computer scientists, and engineers will have to be employed within biomedical research centers and departments. However, the institutional environments for biomedical research can generally be alien to these fields, both culturally and professionally. A physical separation can isolate these scientists from their peers; yet, unless they are immersed in the research milieu where the experiments are being carried out, they will be unable to develop a sufficient appreciation of the biology to be effective collaborators. On the other hand, within a purely biomedical research center, their scholarly contributions may be insufficiently valued, being viewed as dependent on the creativity of their biology colleagues. Also, the salary scales and opportunities for advancement may not match those available in their own established disciplines.

A related issue is the dearth of comprehensive interdisciplinary training programs that combine cell biology with mathematics, computer science, and/or physics. Undergraduate and graduate programs in cell biology are almost universally devoid of any quantitative training opportunities, let alone requirements for basic courses in kinetics or differential equations. On the other hand, even bioengineering training programs do not often offer opportunities for comprehensive courses of study in cell biology. The only programs that can offer truly quantitative training in cell biology are the occasional biophysics graduate programs.

Finally, an issue that was consistently raised was the need for cell biologists to produce more numbers, vis-à-vis intracellular concentrations, kinetic rates, and binding constants. Such data are often difficult to obtain and have been insufficiently valued by the biomedical research community. It is also difficult to find such information in the literature and to fully assess its validity, accuracy, and pertinence to the biological system under study. However, experimental numbers are the best inputs to models as they necessarily provide the most effective constraints to the parameter space. Although most biological systems are quite robust, the subsystems may not be. To fully comprehend the behavior of a complex system from knowledge of its component parts, a researcher would need a detailed quantitative description of those parts and the interactions between them.

Proposed Initiatives

(1) An infrastructure for the communication and sharing of models is a high priority. This will require a comprehensive plan for the deployment of computational resources, network access, and software that will facilitate model interchange and retrieval.

- Interchange standards will have to be developed so that models can be visualized and exercised within a variety of software and hardware platforms. Existing efforts to develop SBML (http://www.cds.caltech.edu/erato/sbml/docs/index.html) and CELLML (http://www.cellml.org/) may prove to be a reasonable starting point for expansion of such standards.
- Also at the software/computer science level is the design of database systems to permit both the storage and retrieval of models and the ability to search for model components (molecules, reactions, pathways, kinetic data, geometries, etc.). Development of a peer-review system for the population of the database will also be required.
- National computational resources that are dedicated to the special needs of cell biology simulations are critical. Such resources will house the requisite high performance computing hardware as well as software libraries to enable the key numerical methods used in computational cell biology. The key to success for such centers is the availability of personnel who have expertise in the biology as well as with the hardware and software.
- Access to Internet2 and appropriately fast local network architectures will need to be made available to individual research labs that are committed to computational modeling. This would include grant supplements to support both the appropriate hardware/software as well as personnel for system administration.

It would be important to investigate whether this proposed infrastructure could be jump-started by leveraging the existing NIH initiatives in image database development and deployment – notably the BIRN project at NCRR.

(2) To meet the challenge of creating a new culture in which interdisciplinary collaborative research is strengthened, it would be important to support a grant instrument that fosters the creation of *Computational Cell Biology Centers* within existing biomedical research institutions. These centers should be closely aligned with experimental cell biology research programs with clear collaborations pre-identified. The grant program should provide comprehensive support for faculty research, graduate and post-doctoral training, and a hardware infrastructure that includes appropriate IT personnel. Although the bulk of any grant award would be directed to the research within the Center, supplementary funds should be allocated to the collaborating experimental laboratories to specifically support the generation of quantitative data on intracellular binding constants, reaction rates, channel activities, and molecular concentrations.

(3) A grant program should be targeted at journals containing substantial cell biology content to develop a fully interactive web-based publication format for models and other quantitative information. This will allow the cell biology community to view, validate, exercise and mine models and simulation results. As part of this initiative, a peer-review mechanism will need to be formulated. This program might be co-sponsored by the NLM, which should, in any event, be intimately involved in the design of archiving mechanisms for this publication. Additionally, this publication format should be coordinated with the model database described in item (1).

(4) A proposal that was advanced at the Cranwell Symposium and that has also been the subject of further discussion within the CCB LISTSERV® is to identify a "grand challenge" as a focus for the development of modeling tools and strategies. The idea is to establish a consortium similar to those being established under the National Institute of General Medical Sciences (NIGMS) "Glue Grants" initiative. (Because these programs are meant to provide resources to bring people together, they have been nicknamed "glue grants.") The Consortium will have identified a multi-level, multi-scale problem in cell biology that could benefit from a large set of complementary computational approaches. The experimentalists in the consortium are to provide both the quantitative input data for models and the experimental tests of model predictions. The computational scientists in the consortium will develop new computational, mathematical, and physical frameworks for modeling the biology as well models of specific biological systems. The distinction of this initiative from the NIGMS "Glue Grants" is that the emphasis here is on the development of new computational approaches. The data gathering will be carried out within existing research programs of the participating laboratories; however, a commitment to the generation of quantitative data should be demonstrated. While the output of such an initiative will be inspired by the specific cell biological focus of the consortium, scientific computing algorithms and professional-quality software that will have a much broader applicability should be required products.