CHAPTER 6

EDUCATION, NATIONAL PROGRAMS, AND INFRASTRUCTURE IN SYSTEMS BIOLOGY

Marvin Cassman, Douglas Lauffenburger, and Frank Doyle

INTRODUCTION

The future of systems biology will depend on three critical elements: education of a new generation of scientists who have both biological and mathematical training; the availability of funding that operates outside of disciplinary boundaries; and the availability of a supportive infrastructure that can accommodate the needs of an intrinsically interdisciplinary research area. This chapter will consider the status of these issues both in the U.S. and elsewhere.

EDUCATION

Not surprisingly for such a hot field, systems biology has spurred interest from thousands of researchers, some just starting their careers, others well established but looking for an opportunity to become involved. Because of the need to couple computational analysis techniques with systematic biological experimentation, more and more universities are offering PhD programs that integrate both computational and biological subject matter.

Page 62 lists a set of current educational programs in this field. Several of the early endeavors in the U.S. are closely associated with bioengineering. The Computational & Systems Biology graduate program was established at the Massachusetts Institute of Technology (MIT) in 2004 as a three-way partnership among the biological engineering, biology, and electrical engineering and computer science departments. The University of California, San Diego (UCSD) offers a systems biology track within its Bioengineering graduate program. Other nascent programs are not closely associated with formal engineering programs but instead arise out of life sciences or medical sciences fields. Examples of these include the new programs at Harvard Medical School, the Institute for Systems Biology, Oxford University, and Biocentrum Amsterdam.

Given the pace of the field, it is certainly too early to endorse a particular syllabus as the correct or best option. However, the study of systems biology must lead to a rigorous understanding of both experimental biology and quantitative modeling. Programs might require that all students, regardless of background, perform hands-on research in both computer programming and in the wet laboratory. Required coursework in biology typically includes genetics, biochemistry, molecular and cell biology, with lab work associated with each of these. Coursework in quantitative modeling might include probability, statistics, information theory, numerical optimization, artificial intelligence and machine learning, graph and network theory, and nonlinear dynamics. Of the biological coursework, genetics is particularly important, because the logic of genetics is, to a large degree, the logic of systems biology. Of the coursework in quantitative modeling, graph theory and machine-learning techniques are of particular interest, because systems approaches often reduce cellular function to a search on a network of biological components and interactions. A course of study integrating life and quantitative sciences helps students to appreciate the practical constraints imposed by experimental biology and to effectively tailor research to the needs of the laboratory biologist. At the same time, knowledge of the major algorithmic techniques for analysis of biological systems will be crucial for making sense of the data.

An alternative to pursuing a cross-disciplinary program is to tackle one field initially and then learn another in graduate school. Examples would include choosing an undergraduate major in bioengineering and then obtaining a PhD in molecular biology, or starting within biochemistry then pursuing graduate coursework in bioengineering and systems biology. This leads to a common question: when contemplating a transition, is it better to switch from quantitative sciences to biology or vice versa? Although some believe that it is easier to move from engineering into biology, the honest answer is that either trajectory can work. Some practical advice is that if coming from biology, it is best to start by becoming familiar with Unix, Perl, and Java before diving into more complex computational methodologies. If coming from the quantitative sciences, an effective strategy is to jump into a wet laboratory as soon as possible.

With the formation of myriad new academic departments and centers, the academic job market is booming. On the other hand, biotechnology firms and ‘big pharma’ have been more cautious about getting involved. However, most agree that in the long term systems approaches promise to influence drug development in several areas: (a) target identification, in which drugs are developed to target a specific molecule or molecular interaction within a pathway; (b) prediction of drug mechanism of action (MOA), in which a compound has known therapeutic effects but the molecular mechanisms by which it achieves these effects are unclear; and (c) prediction of drug toxicity and properties related to absorption, distribution, metabolism, and excretion (ADME/Tox). In all of these cases, the key contribution of systems biology would be a comprehensive blueprint of cellular pathways used for identifying proteins at key pathway control points, or proteins for which the predicted perturbation phenotypes most closely resemble those observed experimentally with a pharmacologic or toxic agent.

Looking toward higher levels of living systems behavioral hierarchy, students preparing for research careers in integrative systems physiology should build a strong foundation in core life sciences, mathematics and engineering. It is particularly useful to be immersed in life sciences courses which present biological principles in the context of mathematical models and engineering methodologies. An example of such a course is the year-long course entitled Physiological Foundations of Biomedical Engineering offered in the Biomedical Engineering department at the Johns Hopkins University. Foundation courses in mathematics could include ordinary and partial differential equation theory as well as probability theory and stochastic processes. While not commonly available, introductory course work in nonlinear dynamical systems theory would be valuable. Students may also opt to build a strong foundation in a core engineering discipline such as mechanical, chemical or electrical engineering.

Students pursuing any aspect of computational or systems biology at the graduate level face the hard fact that they *must* be as deeply educated in relevant areas of the life sciences as their biological colleagues, and they *must* be as strong in appropriate areas of engineering and mathematics as their colleagues in traditional areas of engineering and mathematics. Students will only be successful in this endeavor if they have a true love for both their chosen areas of biology and math/engineering. The broad discipline of quantitative modeling of biological systems is one that is developing rapidly and is seeing increased representation in bio- and biomedical engineering, life sciences and traditional engineering departments. Students may therefore undertake combined experimental and modeling research or modeling research conducted in collaboration with experimental investigators with reasonable confidence that they will be able to find an academic department which appreciates and supports the particular balance they have chosen between modeling and experimentation.

Comparative Programs

A comparison between U.S. and other countries in education is difficult, largely because the programs that do exist are of such recent origin. One striking difference is the major role that engineering has played in the development of systems biology and systems biology training in the U.S. and its relative absence in the EU and the U.K. (This was not as true in Japan and clear exceptions can be found in Europe, such as the Max Planck Institute for Complex Technical Systems in Magdeburg.) In general, the backgrounds of the major practitioners outside the U.S. show a heavier representation in physics and mathematics (the latter particularly in the U.K.) and very little in engineering. This is stated neither as a positive or negative, but it does reflect organizational and cultural differences that will be inevitably be reflected in the approaches to training.

However, at this point in time the most identifiable characteristic of training in systems biology is its absence. The most common response to queries about educational programs was that they didn't exist. Since the research efforts are themselves of recent origin perhaps this isn't surprising. However, it is a significant barrier to the development of the field. It is therefore worth examining some of the programs that the WTEC panel was able to see.

Japan

<u>Department of Computational Biology, Graduate School of Frontier Sciences, University of Tokyo</u>

This is the first official department of computational biology in Japan, although others are emerging. It began in 2003 and has six faculty and about 30 graduate students. Although its focus is on bioinformatics and "omics" rather than systems biology, it is clearly moving in the direction of modeling and analysis of dynamics in biological systems.

<u>Bioinformatics Center, Institute for Chemical Research, Kyoto University</u>
<u>Human Genome Center, Institute of Medical Science, University of Tokyo</u>

This is a joint effort that has both teaching faculties and has generated a curriculum in bioinformatics that is widely used. Although most of the courses are focused on genomics, there are also courses on network analysis and pathway reconstruction, and modeling and simulation.

<u>Institute for Advanced Biosciences, Keio University</u>

This is both a graduate and undergraduate program with a wide array of courses from genomics to genetic networks to software engineering. A striking innovation is a student laboratory where bioinformatics students (presumably those without much of a background in biology) are trained in experimental techniques for up to a year.

Europe

<u>Humboldt University, Berlin</u>

Humboldt University has an Institute for Theoretical Biology, a Department of Theoretical Biophysics, and a graduate school in the Dynamics of Cellular processes. Courses in modeling began as long ago as 1993, making it an early adopter in areas related to systems biology. Students come from all of these programs to take a well-developed curriculum in bioinformatics, theoretical biology, and biophysics including formal courses on systems biology and mathematical modeling. They also run two-week workshops in the area. International graduate collaboration programs have been established with Bioinformatics at Boston University, the Kyoto Genomics and Bioinformatics Center, and with the BioCentrum in Amsterdam. These programs include joint workshops, PhD student exchanges and post-doctoral fellowships.

<u>Free University of Brussels</u>

The Free University is initiating a MS degree in Bioinformatics and Modeling. This will have three orientations, one of which would be chosen by the students during the second year of the MS: "classical" bioinformatics; computational structural biology; and modeling of dynamic biological phenomena. This will not begin until 2007.

<u>Free University of Amsterdam</u>

The Free University has a new MS program in Biomolecular Integration/Systems Biology. The aim is to provide both expertise in advanced conceptual and modeling methodologies as well as insight into important biological/biomedical issues. It is a two-year program, and involves a detailed research project where the student spends half the time in Amsterdam and half the time with a partner group in a different country, both locations being involved in the advising. Teaching efforts at the PhD level were still evolving at the time of our visit, and involved a joint graduate school with Humboldt University.

<u>Centre for Mathematics and Physics in the Life Sciences, University College, London</u>

This provides a PhD program that trains students who arrive with a background in biology in mathematics, while students with a background in physics and math take courses in biology. All take a course in modeling and bioinformatics and one in physical techniques in the life sciences. This is followed by a set of case studies in interdisciplinary research plus seminars and special courses. The didactic part of the program is one year, followed by three years of research.

University of Warwick, Mathematics Institute

Interdisciplinary Program in Cellular Regulation
Molecular Organization and Assembly in Cells

This is a four-year program which currently aims to train eight postdocs with backgrounds in mathematics so that they are equipped to study biological problems. The program is focused on theoretical analysis, and the funding does not support experimental data generation. Postdocs are trained in biology through seminars, journal clubs, and attendance in group meetings of biology labs and single-afternoon symposia designed for the program. They take courses for MOAC (Molecular Organization and Assembly in Cells) students (see below). They are also paired with biologists for particular projects. The PhD program in MOAC integrates areas of mathematics, biology, and chemistry. This program is funded by a seven-year Doctoral Training Center grant which started in 2003. The program starts with a six-month course for the three areas, followed by a lab rotation in each of the three areas before students choose labs for their PhD projects. Most students entering the program usually have backgrounds in physical chemistry.

Conclusions

The general impression is that most of the formal teaching programs, in the U.S. and abroad, are in bioinformatics rather than systems biology. Relatively few examples of training in modeling are focused on biological systems, and where they do exist they tend to be isolated courses rather than fully integrated programs in systems biology. Most of the programs offer somewhat *ad hoc* “menu selection” curricula. The difficulty of training quantitative students in biology and *vice versa* is clearly well understood and no real solution has yet been provided, although a number of experiments are underway. In addition to the examples given here, there is also a program in the U.K. to allow senior faculty to train in other disciplines (“discipline hopping” see below). It is much too early to tell which, if any, of these are successful in producing qualified researchers in systems biology. Given the importance of this issue and its embryonic state, some mechanisms for exchanging information internationally and locally on best practices are essential.

Selected Programs in Systems Biology

a. Graduate Programs with Systems Biology Courses

Europe and Great Britain

Flanders and Ghent University
Department of Plant Systems Biology
http://www.psb.ugent.be/

Max Planck Institutes
Institute of Molecular Genetics
Institute of Dynamics of Complex Systems
http://lectures.molgen.mpg.de/
http://www.mpi-magdeburg.mpg.de/

University of Rostock
Systems Biology & Bioinformatics Program
http://www.sbi.uni-rostock.de

University of Stuttgart
Systems Biology Group
http://www.sysbio.de/

Humboldt University Berlin
Institute for Theoretical Biology
http://itb.biologie.hu-berlin.de/

Department of Theoretical Biophysics
http://www.biologie.hu-berlin.de/~theorybp/

Free University of Amsterdam
BioMolecular Integration/Systems Biology
http://www.systembiology.net/topmaster/topmasterbmisbam.htm

University College, London
Centre for Mathematics and Physics in the Life Sciences
http://www.ucl.ac.uk/CoMPLEX/

University of Warwick
Interdisciplinary Program in Cellular Regulation
Molecular Organization and Assembly in Cells
http://www.maths.warwick.ac.uk/ipcr/

University of Oxford
Centre for Mathematical Biology
http://www.maths.ox.ac.uk/cmb

Asia

*A*Star Bioinformatics Institute, Singapore*
http://www.bii.a-star.edu.sg/

University of Tokyo
Graduate School of Information Science and Technology
http://www.i.u-tokyo.ac.jp/index-e.htm

Department of Computational Biology, Graduate School of Frontier Sciences
http://www.k.u-tokyo.ac.jp/renewal-e/course_jyoho/senkou-e.html

Kyoto University and University of Tokyo
Education and Research Organization for Genome Information Science
http://www.bic.kyoto-u.ac.jp/egis/

Keio University
Institute for Advanced Biosciences
http://www.iab.keio.ac.jp/

North America

Cornell, Sloan-Kettering, and Rockefeller Universities
Physiology, Biophysics & Systems Biology
Program in Comp. Biology and Medicine
http://www.cs.cornell.edu/grad/cbm/
http://biomedsci.cornell.edu

Massachusetts Institute of Technology
Computational and Systems Biology Initiative (CSBi), Biological Engineering Division
http://csbi.mit.edu/

Princeton University
Lewis-Sigler Institute for Integrative Genomics
http://www.genomics.princeton.edu

Stanford University
Medical Informatics (SMI) and BioX
http://smi-web.stanford.edu/

University of California Berkeley
Graduate Group in Computational and Genomic Biology
http://cb.berkeley.edu/

University of California San Diego
Department of Bioengineering
http://www-bioeng.ucsd.edu/

University of Toronto
Program in Proteomics and Bioinformatics
http://www.utoronto.ca/medicalgenetics/

University of Washington
Department of Bioengineering, Deptartment of Genome Sciences
http://www.gs.washington.edu/

Virginia Tech
Program in Genetics, Bioinformatics and Computational Biology
http://www.grads.vt.edu/gbcb/phd_gbcb.htm

Washington University
Computational Biology Program
http://www.ccb.wustl.edu/

b. Short Courses

Humboldt University
Berlin Graduate Program
Dynamics and Evolution of Cellular and Macromolecular Processes
http://www.biologie.hu-berlin.de/

Biocentrum Amsterdam
Molecular Systems Biology Course
http://www.science.uva.nl/biocentrum/

Cold Spring Harbor Laboratory
Course in Computational Genomics
http://meetings.cshl.org/

Institute of Systems Biology
Introduction to Systems Biology and Proteomics Informatics courses
http://www.systemsbiology.org

University of Oxford
Genomics, Proteomics and Beyond
http://www.conted.ox.ac.uk/cpd/biosciences/courses/short_courses/Genome_Analysis.asp

c. Emerging Initiatives

German Systems Biology Research Program
http://www.systembiologie.de/

Harvard University
Department of Systems Biology
http://sysbio.med.harvard.edu/

Manchester Interdisciplinary Biocentre (MIB)
http://www.mib.umist.ac.uk/

University of Texas Southwestern
Program in Molecular, Computational and Systems Biology
Integrative Biology Graduate Program
http://www.utsouthwestern.edu/utsw/home/education/integrativebiology/

NATIONAL PROGRAMS FOR SUPPORT OF SYSTEMS BIOLOGY

Systems biology is a relatively new discipline that involves the integration of engineering, physics, and biology. Its future depends on new sources of funding, since most existing funding programs have difficulty crossing disciplinary lines. Also, given that the discipline is at a relatively early stage in its development, the scientific activities represent basic research and will be primarily funded by the primary supporters of basic research, i.e. governmental entities. The following section provides an overview of some of the national programs supporting systems biology.

National Programs in the U.S.

The first efforts to directly support systems biology were a training program initiated in 1996 by a private foundation, the Burroughs-Welcome Fund, and, in 1998, a research support program begun by the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH). In more recent years, virtually every federal agency involved in supporting science has generated programs for the support of systems biology. (More details can be found in Cassman, 2005.)

Since the NIH is by far the largest supporter of research in the biological sciences in the U.S., it is not surprising that it contains the largest and most diverse array of programs supporting systems biology. These include programs for support of individual and institutional training, individual research grants, centers, and targeted disease-oriented studies, e.g. Integrative Cancer Biology (National Institute of General Medical Sciences, 2005; National Cancer Institute, 2005). In addition, new trans-institute programs have been established to develop “National Technology Centers for Networks and Pathways” and centers in “Metabolomics Technology Development” (National Institutes of Health, 2005). There are also significant programs at the National Science Foundation (NSF), such as “Quantitative Systems Biotechnology” and “Frontiers in Integrative Biological Research” (National Science Foundation, 2005); at the Department of Energy (DoE), through multi-institutional consortia focused on the analysis of microbial systems (Department of Energy, 2005); and at the Defense Advanced Research Projects Agency (DARPA), which has a large program to develop computational models and tools for *in silico* analysis (Defense Advanced Research Projects Agency, 2005).

These are programs that identify themselves as supporting quantitative approaches to biological networks. They leave out the much larger array of support mechanisms for proteomics and genomics, some of which include within them activities that are indistinguishable from the programs cited. Furthermore, there is even more support for research identified as investigator-initiated, i.e. not specifically promoted through identifiable programs.

National Programs Outside the U.S.

U.K.

The Engineering and Physical Sciences Research Council (EPSRC) Life Sciences Interphase (LSI) Programme “aims to fund high-quality research at the boundary between engineering and the physical sciences and the life sciences” (Engineering and Physical Sciences Research Council, 2005). These include networks “which are expected to lead to new collaborative multidisciplinary research proposals,” and which funded for a total of £60,000 each for three years. There are also postdoctoral mobility awards and LSI doctoral training centers that encourage multidisciplinary training.

The Medical Research Council (MRC) has a program called “Discipline Hopping” run jointly with ESPRC and the Biotechnology and Biological Sciences Research Council (BBSRC) (Medical Research Council, 2005). As defined in the notice, “this scheme allows researchers who have a track record in their own field in the physical sciences to apply for funding to investigate and develop ideas, skills and collaborations in the areas of biological, clinical, health services and public health research. Alternatively, life science researchers can apply for funding to develop ideas, skills and collaborations with physical scientists.” The awards are for three months to one year and are for no more than £60,000.

The BBSRC has also initiated a program for Centres for Integrative Systems Biology which will “integrate traditionally separate disciplines such as biology, chemistry, computer science, engineering, mathematics and physics in a programme of international quality research in quantitative and predictive systems biology” (Biotechnology and Biological Sciences Research Council, 2005). A number of new Centers will be awarded for up to £5 million each.

Germany

The Federal Ministry of Education and Research (BMBF) began discussions in 2001 to determine new funding strategies in bioscience, with a focus on cross-disciplinary activities. Their conclusions resulted in the promotion of systems biology in Germany through a network of centers of excellence. Individual research projects would be developed as collaborative projects between science and industry, and focus on the hepatocyte. It was planned to provide up to €50 million over five years. The initial funding began in January, 2004. Twenty-five groups are now supported with funding of €14 million over the next three years (Federal Ministry of Education and Research, 2005).

The support for systems biology also has contributions from independent research organizations. For example, the Helmholtz Association supports a number of research centers in biology and medicine, among which is the Heidelberg Cancer Research Institute (DKFZ), a key organization in the BMBF program. Additionally, it maintains a Network for Bioinformatics which provides coordinated access to bioinformatics resources. The Max Planck Institutes (MPI) are another key element in supporting systems biology. The MPI for Plant Physiology in Berlin; the MPI for Molecular Genetics in Potsdam; and the MPI for Dynamics of Complex Technical Systems in Magdeburg all support systems biology. Finally, there is significant activity in universities, such as the Theoretical Biology Department at Humboldt University, and through support from states within the federal republic. This is seen most notably in the state of Baden-Wuerttemberg life science centers, one of which is the Center for Biosystems Analysis in Freiburg.

Switzerland

The ETH Zürich, the University of Zürich, and the University of Basel have generated a collaborative project entitled "Systems*X*" intended to serve as a focus for systems biology in Switzerland (Systems*X*, 2005). The structure will accommodate collaborative efforts across disciplines and locations. To ensure integration between the several sites involved, the project management will be at the highest levels, comprising the president of the ETH Zürich, the rector of the University of Basel, the rector of the University of Zürich, the vice-president of research at the ETH Zürich, the vice-rector of research at the University of Basel, the pro-rector for research at the University of Zürich, research representatives from Novartis and Roche, plus the spokesperson for Systems*X*. Components of Systems*X* will include the Functional Genomics Centre Zürich, the Glycomics Initiative at ETH Zürich and the University of Zürich, the Oncology Cell Transfer Project at the University of Zürich, the Basel Bioinformatics Initiative, and the Life Sciences Training Facility at the University of Basel.

The Netherlands

Dutch initiatives in systems biology include the NWO (the Dutch equivalent of a hybrid between NSF and NIH) that is funding projects in bioinformatics from molecule to cell, and in computational biology, at a level of $6 million each. There are also initiatives in National Genomic Centers (funded at $200 million), including two that address some aspects of systems biology (Center for Medical Systems Biology at Leiden, Vrije Univ., Institute of Environmental and Energy Technology (TNO), and the Kluyver Center). Future efforts were described in the form of a set of focused program proposals that focus on organisms (*L. lactis* and *S. cerevisiae*) as well as tools (the Silicon Cell). Finally, there is an active center conducting systems biology in the Netherlands, at the Biozentrum Amsterdam.

Other European Activities

Active centers exist in Brussels (Free University), Sweden (University of Lund), and through a program that stimulates collaborations in science, including systems biology, between Sweden and Denmark. Additionally, a number of trans-European programs attempt to support activities across national boundaries. Many of these are supported through the European Commission, including an effort focused on computational biology. A recently funded program relevant to systems biology is a network of organizations involved in *in silico* simulation of biological systems with the goal of aiding in drug design. Coordinated by the Technical University of Denmark, the network will be comprised of 25 universities, a number of national medical agencies, and the pharmaceutical company Novo Nordisk. The European Union will provide €10.7 million over five years.

Japan

The primary mode of governmental support for the sciences is through MEXT (the Ministry of Education, Culture, Sports, Science, and Technology). The programmatic activities are then conducted by the Japan Science and Technology (JST) agency, whose mission is basic research and support of infrastructure. Another branch of MEXT is the Japan Society for the Promotion of Science (JSPS), which supports research, fellowships, and, particularly, international collaborations. In addition, several other agencies support science. For example, the Ministry of Economics, Trade, and Industry (METI) funds the Japan Biological Informatics Consortium (JBIC).

The Japanese government has initiated a number of very large-scale projects primarily in the areas of genomics and proteomics. This includes the Millenium Project, which incorporates national efforts in the rice genome, human genome diversity, and bioinformatics. Bioinformatics includes, among other things, structural and functional genomics, a number of databases, and the development of bioinformatics technologies. Other large-scale efforts include mouse and human full-length complementary deoxyribonucleic acid (cDNA) annotation programs and a high-throughput structural genomics effort.

Until recently there has been no national program directly targeted to systems biology. However, MEXT has initiated the Genome Network Program in 2004 that will include an investigation of the human genome network, most of which will be carried out by RIKEN, the Institute for Physical and Chemical Research (MEXT, 2005). Additionally there will be components which will integrate the information gained into a broader database of genomic and proteomic information; development of new genome analysis technologies; and spin-off applications to specific biological projects.

Although targeted efforts to systems biology are recent, other support has been available. For example, the JST database lists 16 funded applications under the rubric of systems biology (JST, 2005). Perhaps most striking is the number of well-appointed institutes that are doing systems biology in some form. The largest of these is RIKEN. It has five campuses in Japan plus several abroad. The site in Yokohama includes a Genome Science Center and a Plant Science Center, both of which have activities related to systems biology. Institutes have also been set up through funding of local prefectures. Examples are the Kazusa DNA Research Institute, largely supported by Chiba Prefecture, and the Institute for Advanced Biosciences, supported by Yamagata Prefecture and Tsuruoka City. Finally, industry is significantly involved in the academic and national/regional institutes, through a number of mechanisms. These include gift funding, collaborations, and commercial start-ups.

Conclusions

The U.S. remains one of the few countries that has a significant targeted investment in systems biology. A clear exception is Germany, which has developed a new initiative in the systems biology of hepatocytes, beginning in January 2004. National programs have also been initiated in Switzerland and the U.K. in the last few years, while the European Commission has acted as a catalyst for multi-national programs. Additionally, activities in systems biology are underway in many locations, as part of ongoing “traditional” governmental support programs. This is perhaps particularly noticeable in Japan. However, it is hard to avoid the conclusion that both the breadth and the scale of systems biology support from governmental entities are significantly greater in the U.S. than elsewhere in the world.

A possible caveat to this conclusion depends on the definition of systems biology. As noted in the Introduction, there is a distinction between “systems biology” and “systematic biology.” Systematic biology, the high-throughput collection of targeted data sets, is a booming business everywhere, fuelled by the success of the genome project. Systems biology, the computational analysis of biological networks, is much more sparsely represented. Although this is also true in the U.S., encouragement of these activities through federal funding programs is significant and growing. It was slightly discouraging to see how frequently systems and systematic biology were conflated. Although data collection is clearly critical, it was not often the case that there was a connection between the data collected and its potential use in modeling and simulation of biological systems. In general, the future of systems biology worldwide depends on the support of programs which consider experimental and data-driven approaches together with the computational methods needed to model specific biological problems. Relatively few funding programs explicitly focus on this.

INFRASTRUCTURE

The term “infrastructure” is almost as ambiguous as “systems biology.” It can mean anything from a new building to a simple laboratory spectrometer. The benefits to systems biology of buildings that house investigators with common interests may be of more significance than in other disciplines, given the requirements of interdisciplinary research, but it does not necessarily require any special attention in this report. Everyone wants new space, and the benefits may be real, but the arguments are too diverse and too tied to local circumstances for us to get involved. Similarly, laboratory instrumentation is an absolute requirement for any discipline that has an experimental base, but needs are too varied to provide arguments for any specific tool.

The infrastructure to be discussed in this section is limited to large-scale resources, specifically databases, software repositories, and centers. The term “large scale” is used to mean resources such as the Entrez databases which not only serve to centralize and index knowledge but provide a common core of data for many investigators and research areas. For example, the centralization, standardization and dissemination of sequence data at the National Center for Biotechnology Information (NCBI), the European Molecular Biology Laboratory (EMBL) and elsewhere have allowed the growth and improvement of algorithms for phylogeny, homology and other functional assignment which make up the core to nearly all molecular biology. This “network” effect has also been partly realized in the area of structural biology, but other data ranging from microbial phenotype to genetic manipulation data to functional genomics have not yet been similarly controlled, standardized and centralized to achieve the same degree of synergy (and thus quality control).

Many of these issues have been discussed at greater length in other chapters in this volume. The discussion here will focus on some specific concerns.

Software Repositories

The WTEC team’s visits abroad as well as the team’s knowledge of activities in the U.S. confirmed the existence of extensive activities in the development of software, which, especially in modeling and simulation, is a critical tool for systems biology. The WTEC panel saw little reason to believe that the state-of-the-art in this area is significantly more advanced in one country or area of the world rather than another. Indeed, it is frequently directed to similar goals. (One group at RIKEN even called their software YAGNS, “Yet Another Gene Network Simulator.”) The reasons for this cottage industry in software are many, including the need to accommodate data derived locally; the requirement for visualization to accommodate specific requirements of collaborators; and a lack of knowledge of what is already available. In general, however, it is a terrible waste of time, money, and effort. At the moment, locally created software is practically inaccessible, even when the developers are willing to release it, since documentation is often so scanty that the barriers to use are prohibitive. (A more detailed discussion can be found in Chapter 4 in this volume.)

A reasonable set of expectations for software is that it should be interoperable, transparent to the user, and sufficiently well documented so that it can be modified and adjusted to circumstances. In systems biology there is the additional complication that the data sets used are frequently very diverse and often inconsistent with each other. For the developers and skilled users these problems may occasionally be overcome. However, the benefits of systems biology will only become manifest when working biologists, who are not themselves sufficiently trained to use such software, can manipulate and use these techniques. Admittedly, the translation of systems biology to a broadly based approach is complicated by the innumeracy of most biologists. Some modicum of mathematical training, which is now lacking, will be required, (see section on education). However, there is an immediate need to provide an opportunity for potential users to access and effectively use bioinformatics, modeling, and simulation software. One possible approach is to create a central organization that would serve as a repository for systems biology software as well as serve as a mechanism for validating and documenting their utility and for standardizing the developers’ interfaces and data input/output formats. Like central data repositories, having a central software repository with software engineering standards in place should create a network effect wherein synergy is created by the combined use and reconfiguration of tools for more sophisticated analysis.

Given that agencies worldwide are engaged in promoting the development of systems biology software (see items above describing support by Defense Advanced Research Projects Agency and the National Institutes of Health in the U.S.; the EC; and Japan), it would seem reasonable to create a structure that will preserve and enhance the benefits from these programs.

Databases

The problems with the diversity of software noted above are paralleled by the diversity in the way data used for modeling is collected, annotated, and stored. (These questions are discussed in more detail in Chapter 2) These issues are even more complex than for sequencing since systems biology is highly context dependent.

In order for these data to be useful outside of the laboratory in which they were generated they must be standardized, presented using a uniform ontology, and annotated sufficiently so that the specific cell type, conditions of the medium, etc., are clearly reproducible. Systems biology often requires the use of multiple forms of data, e.g. metabolite and mRNA profiling, kinetic and thermodynamic measurements, etc. It is important to insure that the data is presented and annotated in a form that allows for all these data types to be effectively correlated. Additionally, one of the important functions of easy access to data and software is peer-review. In order to evaluate, in an ongoing way, the increasingly complex data, both raw and processed, the increasingly sophisticated analysis tools, and the increasingly less complete papers (that cannot include all information because of the very complexity of the experiments and tools), it is vitally important that reviewers and community have continuous access to the results and tools used to produce the literature. Dealing with this very complex issue will require a focused effort by the researchers involved as well as the funding agencies. It must be done soon.

Centers

The benefits of focused centers containing large numbers of scientists are almost offset by their shortcomings. Although there are synergies to be had from complementary groups of investigators, there is also the tendency of such structures to become sclerotic over time. They must be approached with caution. Nevertheless, there are examples of such centers being of great value. Perhaps the most clear-cut examples are those where economy of scale yields results not otherwise possible, such as the high-throughput sequencing projects. These can be found at a number of sites in the U.S. supported by the NIH for both DNA sequencing and protein structure determination; at RIKEN Yokohama, JBIRC, and Kazusa DNA Research Institute in Japan; at the Max Planck Institute of Molecular Plant Physiology and EMBL in Germany; and at the European Bioinformatics Institute (EBI) in the U.K.

A different kind of center is one created around specific biological problems. It is possible to identify several examples of these related to systems biology. In the U.S. there are five centers established by the National Institute of General Medical Sciences of the NIH as well as the private Institute for Systems Biology in Seattle. Additionally, a number of centers are being planned through the NIH trans-agency "Roadmap" programs. One recent award, part of the National Centers for Biomedical Computing, was to Stanford University for "a simulation toolkit that enables scientists worldwide to model and simulate biological systems from single atoms to entire organisms." The Department of Energy has created centers focusing on bacterial systems. Another interesting approach is that of the Alliance for Cell Signaling which plans a comprehensive analysis of the signaling molecules and pathways in eukaryotic cells. In contrast to other centers, which are co-located, this group involves seven laboratories in five different locations. This was supported by a mechanism called "glue grants," offered by the National Institute of General Medical Sciences of the NIH, and was specifically designed to coordinate focused research efforts across multiple institutions.

In Germany, the Max Planck Institutes provide an institutional framework that makes it possible to create centers focused on specific problem areas. An example is the Institute for Complex Technical Systems in Magdeburg, Germany, headed by Prof. Ernst Dieter Gilles. This is an institute with a number of related focus areas in engineering, one of which is in systems theory. This group has 17 faculty members, all of whom work in some aspect of systems biology. Although they have some capability in experimental areas, their strength is in the depth and breadth of theory, and they develop this through a wide array of external collaborations in experimental programs. Another example is the Max Planck Institute for Plant Physiology in Golm, Germany, headed by Professors Lothar Willmitzer and Mark Stitt. The departments in this Institute are effectively merged, with the goal of an integrated research approach to solve basic questions in plant metabolism, combining methods from genetics, molecular biology, and chemistry. The strengths of such an integrated approach were particularly visible in the linkage of the bioinformatics capability of the Institute to all of its laboratories, and in the existence of an overall strategy for addressing phenotyping coupled to the disruption of critical genes, all linked to a comprehensive sequencing and bioinformatics infrastructure. Most recently, a distributed program focused on the systems biology of the hepatocyte has been established. This is most comparable to the Alliance for Cell Signaling in the U.S.

In the Netherlands, Prof. Hans Westerhoff at the Free University in Amsterdam has a group combining experimental and theoretical approaches to systems biology, much of it oriented around metabolic control analysis and hierarchical control analysis. The existence of a group of investigators working in related areas has a clear synergistic influence, not only in research but in training as well.

In Japan the government, both centrally and at local levels, has created a number of institutes targeting specific problem areas. Most of these are in the area of large-scale data generation, organized collections of full-length cDNA clones, and in collection of mutant lines. However, several are more closely linked to systems biology. As examples, it is worth mentioning two that are part of Keio University and are largely the creation of strong leaders. Prof. Hiraoki Kitano has developed the Symbiotic Systems Project in Tokyo, while Prof. Masaru Tomita has initiated the Institute for Advanced Biosciences in Tsuruoka. Both of these investigators have a long history of research in systems biology and have created programs which link theory and experiment. In addition, the National Institute for Advanced Science and Technology (AIST) has created a novel research structure, the Computational Biology Research Center (CBRC) in Tokyo. The CBRC is organized to conduct research in bioinformatics only in terms of information theory completely independent from experimental biology projects, although they engage in many collaborations. Areas of emphasis include genome informatics, molecular modeling and design, and cellular informatics.

A variety of centers are being created in the U.S., largely to provide technology development. An argument could be made for the creation of centers targeted to specific research problems, and specific experimental systems, in systems biology in the U.S. The need for consistent and reproducible data and the need for close collaboration between theorists and experimentalists are both arguments for co-located groups that can interact easily and often. It is also far easier to enforce standards at such centers. There are a number of examples of such organizational structures outside the U.S. In general, they are characterized by strong leaders with clear programmatic goals. These kinds of centers are relatively rare in the U.S., at least in part because the scientific culture is oriented to smaller groups with more distributed authority. This has in fact been a major strength of the U.S. system for many years and in most research fields. However, it should not blind us to the possibility that other approaches can supplement this model. At this point in time, systems biology can benefit from stronger centralized approaches that will allow the testing of model systems in an optimum environment.

Conclusions

The key issues of infrastructure in systems biology—databases, software, and centers—are common across all the countries involved. In particular, the availability of a common structure for the use of data and software is lacking and requires immediate attention. International collaborations will be needed to accomplish this.

REFERENCES

Biotechnology and Biomedical Sciences Research Council, http://bbsrc.mondosearch.com/ (Accessed October 1, 2005)

Cassman, M., (2004). Systems Biology in the U.S. *IEE Sys Biol* 1: 204–205.

Defense Advanced Research Project Agency, http://www.darpa.mil/ipto/programs/biocomp/index.htm (Accessed October 1, 2005)

Department of Energy, http://doegenomestolife.org/ (Accessed October 1, 2005)

Engineering and Physical Sciences Research Council, http://www.epsrc.ac.uk/ResearchFunding/Programmes/LifeSciencesInterface/default.htm (Accessed October 1, 2005)

Federal Ministry of Education and Research, http://www.bmbf.de/en/1140.php (Accessed October 1, 2005)

JST, http://j-east.tokyo.jst.go.jp/cgi-bin/DbSearchList.cgi, (Accessed October 1, 2005)

Medical Research Council, http://www.mrc.ac.uk/index/funding/funding-specific_schemes/funding-current_grant_schemes/funding-discipline_hoppers.htm (Accessed October 1, 2005)

MEXT, http://www.mext-life.jp/genome/english/ (Accessed October 1, 2005)

National Cancer Institute, http://dcb.nci.nih.gov/newsdetail.cfm?ID=11 (Accessed October 1, 2005)

National Institute of General Medical Sciences, http://www.nigms.nih.gov/funding/complex_systems.html (Accessed October 1, 2005)

National Institutes of Health, http://nihroadmap.nih.gov/ (Accessed October 1, 2005)

National Science Foundation, http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=6188&org=EF (Accessed October 1, 2005)

Systems*X*, http://www.bsse.ethz.ch/ (Accessed October 1, 2005)